Nathan Jacobson

Algebra di base I

Seconda edizione

Edizione italiana di lavoro

Traduzione italiana da *Basic Algebra I*, seconda edizione
W. H. Freeman and Company, 1985

A Florie

Indice

Prefazione alla seconda edizione **6**

Prefazione alla prima edizione **7**

Introduzione Nozioni di teoria degli insiemi. Gli interi **11**

0.1 L'insieme delle parti di un insieme 12
0.2 Il prodotto cartesiano. Applicazioni 14
0.3 Relazioni di equivalenza. Fattorizzazione di un'applicazione attraverso una relazione di equivalenza 19
0.4 I numeri naturali 22
0.5 Il sistema $\mathbb{Z}$ dei numeri interi 25
0.6 Alcuni fatti aritmetici fondamentali su $\mathbb{Z}$ 27
0.7 Un cenno ai numeri cardinali 29

Capitolo 1 Monoidi e gruppi **30**

1.1 Monoidi di trasformazioni e monoidi astratti 31
1.2 Gruppi di trasformazioni e gruppi astratti 33
1.3 Isomorfismo. Il teorema di Cayley 37
1.4 Associatività generalizzata. Commutatività 39
1.5 Sottomonoidi e sottogruppi generati da un sottoinsieme. Gruppi ciclici 41
1.6 Decomposizione delle permutazioni in cicli 45
1.7 Orbite. Classi laterali di un sottogruppo 47
1.8 Congruenze. Monoidi e gruppi quoziente 49
1.9 Omomorfismi 52
1.10 Sottogruppi di un'immagine omomorfa. Due teoremi fondamentali di isomorfismo 56
1.11 Oggetti liberi. Generatori e relazioni 58
1.12 Gruppi che agiscono su insiemi 61
1.13 I teoremi di Sylow 67

Capitolo 2 Anelli **72**

2.1 Definizione e proprietà elementari 72
2.2 Tipi di anelli 75
2.3 Anelli di matrici 77
2.4 Quaternioni 81
2.5 Ideali e anelli quoziente 83
2.6 Ideali e anelli quoziente di $\mathbb{Z}$ 84
2.7 Omomorfismi di anelli. Teoremi fondamentali 86
2.8 Anti-isomorfismi 89
2.9 Campo delle frazioni di un dominio commutativo 92

2.10 Anelli di polinomi 95
2.11 Alcune proprietà degli anelli di polinomi e loro applicazioni 99
2.12 Funzioni polinomiali 104
2.13 Polinomi simmetrici 106
2.14 Monoidi e anelli fattoriali 107
2.15 Domini a ideali principali e domini euclidei 112
2.16 Estensioni polinomiali dei domini fattoriali 115
2.17 «Rng»: anelli senza unità 117

Capitolo 3 Moduli su un dominio a ideali principali 119
3.1 L'anello degli endomorfismi di un gruppo abeliano 120
3.2 Moduli sinistri e destri 123
3.3 Nozioni e risultati fondamentali 125
3.4 Moduli liberi e matrici 128
3.5 Somme dirette di moduli 131
3.6 Moduli finitamente generati su un dominio a ideali principali. Risultati preliminari 133
3.7 Equivalenza delle matrici a coefficienti in un dominio a ideali principali 135
3.8 Teorema di struttura dei moduli finitamente generati su un dominio a ideali principali 139
3.9 Moduli di torsione e componenti primarie. Teorema di invarianza 141
3.10 Applicazioni ai gruppi abeliani e alle trasformazioni lineari 144
3.11 L'anello degli endomorfismi di un modulo finitamente generato su un dominio a ideali principali 151

Capitolo 4 Teoria di Galois delle equazioni 156
4.1 Risultati preliminari, alcuni già noti, altri nuovi 157
4.2 Costruzioni con riga e compasso 160
4.3 Campo di spezzamento di un polinomio 165
4.4 Radici multiple 168
4.5 Il gruppo di Galois. La corrispondenza fondamentale di Galois 171
4.6 Alcuni risultati sui gruppi finiti 178
4.7 Il criterio di Galois per la risolubilità per radicali 183
4.8 Il gruppo di Galois come gruppo di permutazioni delle radici 186
4.9 L'equazione generale di grado n 191
4.10 Equazioni a coefficienti razionali aventi per gruppo di Galois il gruppo simmetrico 194
4.11 Poligoni regolari costruibili. Campi ciclotomici su $\mathbb{Q}$ 197
4.12 Trascendenza di e e di π. Il teorema di Lindemann–Weierstrass 200
4.13 Campi finiti 207
4.14 Basi speciali per le estensioni di grado finito 208
4.15 Tracce e norme 212
4.16 Riduzione modulo p 216

Capitolo 5 Equazioni e disuguaglianze polinomiali reali **220**
5.1 Campi ordinati. Campi reali chiusi 220
5.2 Il teorema di Sturm 223
5.3 Formalizzazione dell'algoritmo euclideo e teorema di Sturm 226
5.4 Procedimenti di eliminazione. Risultanti 230
5.5 Un metodo di decisione per una curva algebrica 233
5.6 Il teorema di Tarski 238

Capitolo 6 Spazi vettoriali metrici e gruppi classici **245**
6.1 Funzionali lineari e forme bilineari 246
6.2 Forme alternanti 250
6.3 Forme quadratiche e forme bilineari simmetriche 253
6.4 Nozioni fondamentali di geometria ortogonale 259
6.5 Il teorema di cancellazione di Witt 262
6.6 Il teorema di Cartan–Dieudonné 266
6.7 Struttura del gruppo lineare generale $\mathrm{GL}_n(F)$ 268
6.8 Struttura dei gruppi ortogonali 273
6.9 Geometria simplettica. Il gruppo simplettico 279
6.10 Ordini dei gruppi ortogonali e simplettici su un campo finito 283
6.11 Postilla sulle forme hermitiane e sulla geometria unitaria 286

Capitolo 7 Algebre su un campo **289**
7.1 Definizione ed esempi di algebre associative 290
7.2 Algebre esterne. Applicazione ai determinanti 293
7.3 Rappresentazioni matriciali regolari delle algebre associative. Norme e tracce . . . 299
7.4 Cambiamento del campo di base. Transitività della traccia e della norma 302
7.5 Algebre non associative. Algebre di Lie e di Jordan 305
7.6 Il problema di Hurwitz. Algebre di composizione 311
7.7 I teoremi di Frobenius e di Wedderburn sulle algebre associative di divisione . . . 320

Capitolo 8 Reticoli e algebre di Boole **323**
8.1 Insiemi parzialmente ordinati e reticoli 323
8.2 Distributività e modularità 327
8.3 Il teorema di Jordan–Hölder–Dedekind 330
8.4 Il reticolo dei sottospazi di uno spazio vettoriale. Il teorema fondamentale della geometria proiettiva 331
8.5 Algebre di Boole 335
8.6 La funzione di Möbius di un insieme parzialmente ordinato 339

Appendice **345**
Alcuni argomenti per lo studio individuale 345

Indice analitico **348**

Nota alla traduzione. Nel testo e nell'indice analitico i rinvii a numeri di pagina seguono la seconda edizione inglese. Il sommario riporta la paginazione italiana. Le formule conservano la numerazione del rispettivo capitolo.

Prefazione alla seconda edizione

Dalla pubblicazione di *Basic Algebra I* nel 1974, diversi docenti e studenti che hanno utilizzato il testo hanno comunicato all'autore correzioni e proposte di miglioramento, nonché nuovi esercizi. Molti di questi contributi sono stati accolti nella presente edizione. Particolarmente degne di nota sono state le osservazioni di Huah Chu della National Taiwan University, di Marvin J. Greenberg dell'Università della California a Santa Cruz, di J. D. Reid della Wesleyan University, di Tsuneo Tamagawa della Yale University e di F. D. Veldkamp dell'Università di Utrecht. Siamo grati a loro e a quanti ci hanno incoraggiato a credere di essere sulla strada giusta nell'adottare il punto di vista di *Basic Algebra I*.

Due modifiche importanti riguardano il capitolo 4, sulla teoria di Galois. La prima è la completa riscrittura del paragrafo sui campi finiti (§ 4.13). Nella nuova versione i risultati principali sono formulati esplicitamente come teoremi. Nella prima edizione erano sepolti nell'esposizione, un vero esercizio di concisione. Abbiamo inoltre inserito nel testo la dimostrazione della formula di Gauss per il numero $N(n, q)$ dei polinomi monici irriducibili di grado n su un campo finito di q elementi. Nella prima edizione la formula compariva in un esercizio (esercizio 20, p. 145). L'esercizio è stato ora modificato e richiede soltanto di determinare $N(2, q)$ e $N(3, q)$. La seconda modifica importante è l'aggiunta del § 4.16, «Riduzione modulo p». Vi si presenta una dimostrazione, dovuta a John Tate, di un teorema di Dedekind: dalla fattorizzazione modulo un primo p di un polinomio monico irriducibile $f(x)$ a coefficienti interi si può dedurre l'esistenza di certi cicli nel gruppo di Galois, considerato come gruppo di permutazioni delle radici. Gli esercizi alla fine del paragrafo presentano varie interessanti applicazioni del teorema.

Nel capitolo 5 abbiamo dato una nuova dimostrazione del teorema fondamentale di eliminazione (teorema 5.6). La dimostrazione è del tutto elementare e indipendente dai metodi formali sviluppati nello stesso capitolo per dimostrare il teorema di Tarski sull'eliminazione dei quantificatori per i campi reali chiusi. Il motivo di questa scelta è che il teorema 5.6 costituisce il passo principale nella dimostrazione del teorema degli zeri di Hilbert (*Nullstellensatz*) data alle pp. 424–426 di *Basic Algebra II*. La modifica è pensata per i lettori che desiderano seguire quella dimostrazione senza doversi familiarizzare con i metodi formali del capitolo 5.

Al termine del libro abbiamo aggiunto un'appendice intitolata «Alcuni argomenti per lo studio individuale», che propone dieci temi. Di ciascuno si danno una breve descrizione e alcuni riferimenti bibliografici. Alcuni avrebbero potuto trovare posto nel corpo del testo, ma riteniamo che gli studenti ne trarranno maggior beneficio affrontandoli per conto proprio.

Le novità elencate occupano circa dieci pagine aggiuntive. Le altre quindici pagine circa che si sono aggiunte in questa edizione si devono a miglioramenti apportati in vari punti dell'esposizione e a nuovi esercizi.

Il testo della seconda edizione è stato interamente ricomposto, il che ha comportato la correzione delle bozze di un lungo manoscritto. Questo gravoso compito è stato assunto in gran parte da Huah Chu, già ricordato, da Jone-Wen Cohn della Shanghai Normal University, da Florence D. Jacobson («Florie», alla quale il libro è dedicato) e da James D. Reid, anch'egli già menzionato. Siamo loro profondamente grati per l'aiuto ricevuto.

Nathan Jacobson
Hamden, Connecticut, 1 novembre 1984

Prefazione alla prima edizione

Sono passati più di vent'anni da quando l'autore intraprese la stesura dei tre volumi delle *Lectures in Abstract Algebra*. I primi due apparvero rispettivamente nel 1951 e nel 1953, il terzo nel 1964. Dal momento in cui l'opera fu concepita, intorno al 1950, l'algebra ha compiuto progressi considerevoli, anche al livello di quei testi. Anzitutto sono state introdotte alcune idee fondamentali. Fra gli esempi più notevoli vi sono lo sviluppo della teoria delle categorie, che offre un quadro utile a gran parte della matematica, l'algebra omologica e le applicazioni della teoria dei modelli all'algebra. Forse ancor più sorprendente della comparsa di queste idee è stata l'accoglienza riservata al metodo assiomatico e concettuale dell'algebra astratta, la cui influenza si è estesa a tutta la matematica. Oggi si dà per acquisito che i metodi algebrici siano uno strumento essenziale della matematica. D'altra parte, nella ricerca recente si osserva un rinnovato interesse per la sfida posta da problemi piuttosto concreti, la cui soluzione richiede spesso strumenti di notevole complessità tecnica.

Un altro cambiamento profondo di questi vent'anni, soprattutto da quando l'Unione Sovietica stupì il mondo mettendo in orbita i suoi Sputnik, è stato l'innalzamento del livello dell'insegnamento della matematica nelle scuole elementari e secondarie. (Di recente si è avuto qualche passo indietro, ma c'è da sperare che si tratti di una deviazione soltanto temporanea.) Ne è seguito un corrispondente innalzamento del livello della matematica universitaria. Un esempio significativo è lo studio precoce dell'algebra lineare, volto a preparare adeguatamente sia all'analisi in più variabili sia alle applicazioni in altri campi. Inoltre, ai corsi di algebra lineare seguono spesso, senza intervallo, corsi di algebra «astratta»: così argomenti che vent'anni fa si insegnavano nei corsi successivi alla laurea vengono oggi presentati a studenti con un'esperienza matematica relativamente modesta.

Questo libro, *Basic Algebra I*, e il futuro *Basic Algebra II* erano stati concepiti inizialmente come nuove edizioni delle nostre *Lectures*. Riflettendo sul compito, e in particolare sulla necessità di tener conto dei nuovi programmi dei primi cicli universitari e degli studi superiori, decidemmo di disporre la materia in maniera assai diversa. La ripartizione segue due livelli di astrazione: il primo, trattato in questo volume, comprende le parti dell'algebra che uno studente agli inizi può apprezzare più facilmente. Gran parte della materia ha un sapore classico. Ci auguriamo che ciò favorisca l'apprezzamento dei grandi contributi del passato, in particolare della matematica dell'Ottocento. Nell'esposizione abbiamo cercato di servirci degli strumenti moderni più efficaci. È stato perciò necessario sviluppare un corpo considerevole di nozioni fondamentali del genere ormai consueto nei testi di algebra astratta. Abbiamo però cercato di tenere sempre in primo piano obiettivi ben definiti, capaci, crediamo, di attrarre anche uno studente con poca esperienza algebrica. D'altra parte, gli argomenti sono approfonditi spesso ben oltre l'uso corrente, e ciò richiederà a tratti notevoli doti e concentrazione. Nel secondo volume contiamo di seguire una via più tradizionale nell'esporre una materia di natura più astratta e sofisticata. Ci auguriamo che lo studio del primo abbia dato allo studente la maturità necessaria per affrontare con naturalezza il livello di astrazione del secondo.

Diamo ora qualche indicazione sul contenuto e sull'organizzazione di *Basic Algebra I*. L'introduzione, dedicata alla teoria degli insiemi e al sistema dei numeri interi, contiene argomenti familiari alla maggior parte dei lettori: l'algebra degli insiemi, la definizione di applicazione e l'induzione matematica. Meno familiari, e di importanza capitale per gli sviluppi successivi, sono le relazioni di equivalenza e gli insiemi quoziente che esse definiscono. Introduciamo anche i diagrammi commutativi e la fattorizzazione di un'applicazione attraverso una relazione di equivalenza. Dimostriamo il teorema fondamentale dell'aritmetica e il teorema di ricorsione (o di definizione per induzione).

Il capitolo 1 tratta i monoidi e i gruppi. Partiamo dai monoidi di trasformazioni e dai gruppi di trasformazioni, seguendo in questo lo sviluppo storico della materia. La nozione di omomorfismo compare relativamente tardi, quando il lettore ha avuto il tempo di assimilare alcune idee più semplici e intuitive. Una volta introdotta, tuttavia, ne sviluppiamo con un certo dettaglio le conseguenze principali: i teoremi fondamentali di isomorfismo e la corrispondenza fra i sottogruppi di un'immagine omomorfa e i sottogruppi che contengono il nucleo. Introduciamo, illustrandolo con numerosi esempi, il concetto di gruppo che opera su un insieme, oggi tanto importante in geometria. Ne ricaviamo un metodo di conteggio per i gruppi finiti, del quale l'equazione delle classi contiene un caso particolare. Questi risultati permettono di dedurre i teoremi di Sylow, ultimo argomento del capitolo.

La prima parte del capitolo 2 riprende per gli anelli molte idee sviluppate nel primo capitolo. Si considerano quindi diverse costruzioni che, da anelli dati, ne producono di nuovi: anelli di matrici, campi delle frazioni di domini commutativi, anelli di polinomi. L'ultima parte è dedicata alla teoria elementare della fattorizzazione nei monoidi commutativi con legge di cancellazione e nei domini commutativi.

L'obiettivo principale del capitolo 3 è la teoria della struttura dei moduli finitamente generati su un dominio a ideali principali, con le sue applicazioni ai gruppi abeliani e alle forme canoniche delle matrici. Prima di arrivarvi occorre naturalmente introdurre le definizioni e le nozioni consuete sui moduli. Si insiste sull'analogia con un gruppo che opera su un insieme e sul fatto che il modulo generalizza in modo naturale la nozione familiare di spazio vettoriale. Il capitolo termina con alcuni teoremi sull'anello degli endomorfismi di un modulo finitamente generato su un dominio a ideali principali. Essi generalizzano i risultati classici di Frobenius sull'anello delle matrici che commutano con una matrice assegnata.

Il capitolo 4 è dedicato quasi per intero agli sviluppi di due problemi classici: la risolubilità delle equazioni per radicali e le costruzioni con riga e compasso. Il primo è di gran lunga il più difficile. Lo strumento forgiato da Galois per affrontarlo, la corrispondenza fra i sottocampi del campo di spezzamento di un polinomio separabile e i sottogruppi del gruppo degli automorfismi, ha assunto un'importanza centrale nell'algebra e nella teoria dei numeri. Riteniamo tuttavia più efficace, a questo livello, concentrarsi sui problemi che diedero impulso alla teoria di Galois e trattarli a fondo. Si amplia la teoria dei gruppi finiti avviata nel primo capitolo con i risultati necessari a stabilire il criterio di Galois per la risolubilità di un'equazione per radicali. Vi è anche una dimostrazione della trascendenza di π, necessaria per provare l'impossibilità della quadratura del cerchio con riga e compasso. (Poiché richiede pochissimo lavoro in più, dimostriamo in realtà il teorema più generale di Lindemann e Weierstrass sull'indipendenza algebrica degli esponenziali.) Alla fine del capitolo completiamo il quadro applicando la teoria di Galois ai principali risultati sui campi finiti, ai teoremi sull'elemento primitivo e sulle basi normali, e ai teoremi fondamentali su norme e tracce.

Nel capitolo 5 prosegue lo studio delle equazioni polinomiali. Lavoriamo ora in un campo reale chiuso, generalizzazione algebrica del campo dei numeri reali. Dimostriamo una generalizzazione del «teorema fondamentale dell'algebra»: per ogni campo reale chiuso R, la chiusura algebrica di R è $R(\sqrt{-1})$. Deduciamo poi il teorema di Sturm, che fornisce un metodo costruttivo per determinare il numero delle radici in R di un'equazione polinomiale in un'incognita a coefficienti in R. L'ultima parte del capitolo riguarda i sistemi di equazioni e disequazioni polinomiali in più incognite. Trattiamo dapprima il problema puramente algebrico dell'eliminazione delle incognite in un tale sistema, quindi stabiliamo un'ampia generalizzazione del teorema di Sturm dovuta a Tarski. In tutto il capitolo l'accento cade sui metodi costruttivi.

La prima parte del capitolo 6 espone i fondamenti della teoria delle forme quadratiche e delle forme alternanti su un campo arbitrario. Vi figurano il teorema di Sylvester sull'indice

d'inerzia e la sua generalizzazione che segue dal teorema di cancellazione di Witt. Dimostriamo l'importante teorema di Cartan–Dieudonné sulla generazione del gruppo ortogonale mediante simmetrie. La seconda parte studia la struttura dei cosiddetti gruppi classici: il gruppo lineare generale, il gruppo ortogonale e il gruppo simplettico. Per i tre tipi di gruppi adottiamo un metodo uniforme, introdotto da Iwasawa per il gruppo lineare generale ed esteso da Tamagawa ai gruppi ortogonali. I risultati forniscono importanti classi di gruppi semplici dei quali, nel caso dei campi finiti, è facile calcolare l'ordine.

Il capitolo 7 introduce la teoria delle algebre, sia associative sia non associative. Fra gli argomenti principali della teoria associativa consideriamo l'algebra esterna di uno spazio vettoriale. Essa ha un ruolo importante in geometria e qui viene applicata per dedurre i principali teoremi sui determinanti. Definiamo inoltre la rappresentazione regolare, la traccia e la norma di un'algebra associativa, e dimostriamo un teorema generale sulla transitività di queste funzioni. Per le algebre non associative diamo definizioni ed esempi delle classi più importanti. Segue una dimostrazione interamente elementare del bel teorema di Hurwitz sulla composizione delle forme quadratiche. Concludono il capitolo le dimostrazioni del teorema di Frobenius sulle algebre con divisione sul campo reale e del teorema di Wedderburn sulle algebre finite con divisione.

Il capitolo 8 dà una breve introduzione ai reticoli e alle algebre di Boole. I temi principali sono il teorema di Jordan–Hölder sui reticoli semimodulari, il cosiddetto «teorema fondamentale della geometria proiettiva», il teorema di Stone sull'equivalenza fra algebre di Boole e anelli booleani (gli anelli in cui ogni elemento è idempotente), e infine la funzione di Möbius di un insieme parzialmente ordinato.

Basic Algebra I è pensato per un primo corso di algebra successivo all'algebra lineare. Contiene molto più di quanto si possa svolgere in un anno. In base alla nostra recente esperienza con versioni precedenti del testo proponiamo qualche indicazione per un corso annuale diviso in due semestri o in tre trimestri. In un semestre siamo riusciti a coprire l'introduzione, trattandola rapidamente, e quasi tutti i capitoli 1–3. Abbiamo dovuto omettere la dimostrazione del teorema di ricorsione nell'introduzione, il paragrafo sui gruppi liberi nel capitolo 1, l'ultimo paragrafo del capitolo 2 (sui cosiddetti *rng*, anelli non necessariamente unitari) e l'ultimo paragrafo del capitolo 3. Il capitolo 4, sulla teoria di Galois, è un ottimo punto di partenza per il secondo semestre. Data la ricchezza dell'argomento, non rimarrà molto tempo per altro. Omettendo qualche parte, per esempio la dimostrazione del teorema di Lindemann–Weierstrass, alla fine resteranno probabilmente alcune settimane. Si può allora scegliere fra varie possibilità. Una è passare dalle equazioni in un'incognita ai sistemi di equazioni polinomiali in più incognite: un aspetto di questo problema è trattato nel capitolo 5, che in parte si accorda assai bene con il quarto. D'altra parte, anche un deciso cambio di tema ha i suoi vantaggi. Si potrebbe affrontare il capitolo sulle algebre, oppure studiare una parte del capitolo sulle forme quadratiche e sui gruppi classici, o ancora l'ultimo capitolo, sui reticoli e sulle algebre di Boole.

Un corso in tre trimestri potrebbe articolarsi così: introduzione e capitoli 1 e 2 nel primo trimestre; capitolo 3 e una parte considerevole del capitolo 6 nel secondo. In tal caso occorrerà integrare la teoria dei campi con quanto serve del capitolo 4 per affrontare il sesto. Si potrebbe concludere dedicando il terzo trimestre al capitolo 4, sulla teoria di Galois.

Ci auguriamo che lo studente integri i corsi basati sul testo con la lettura individuale delle parti omesse. Riteniamo inoltre che parecchi argomenti si prestino a programmi di studio autonomo sotto la guida di un docente.

Siamo molto grati agli amici e colleghi che hanno letto parti della penultima stesura e ci hanno dato preziosi suggerimenti, dei quali abbiamo tenuto conto nella versione definitiva. Walter Feit e Richard Lyons hanno suggerito vari esercizi di teoria dei gruppi. Abraham Robinson, Tsuneo Tamagawa e Neil White hanno letto le parti del libro di cui sono specialisti (rispettivamente i

capitoli 5, 6 e 8) e individuato difetti che ci erano sfuggiti. George Seligman ha letto l'intero manoscritto e proposto alcuni miglioramenti sostanziali. S. Robert Gordon, James Hurley, Florence Jacobson e David Rush hanno utilizzato parti della prima stesura in corsi della durata di un periodo accademico o più, segnalandoci numerosi punti in cui l'esposizione poteva essere migliorata.

Diverse persone hanno avuto una parte importante nella preparazione del volume. Ricordiamo in particolare Florence Jacobson e Jerome Katz, che ci sono stati di grande aiuto nel faticoso compito della correzione delle bozze. Una parola speciale va infine a Mary Scheller, che ha battuto a macchina di buon grado l'intero manoscritto e la versione preliminare, quasi altrettanto lunga.

A tutte le persone che abbiamo ricordato, e alle altre che ci hanno aiutato, siamo profondamente debitori e desideriamo esprimere qui il nostro sincero apprezzamento e la nostra gratitudine.

Nathan Jacobson
Hamden, Connecticut

Introduzione
Nozioni di teoria degli insiemi. Gli interi

Lo scopo principale di questo volume è introdurre le strutture fondamentali dell'algebra: gruppi, anelli, campi, moduli, algebre e reticoli. Questi concetti offrono un ambiente naturale a una vasta parte dell'algebra, compresa l'algebra classica. È degno di nota che molti siano nati per risolvere problemi concreti di geometria, di teoria dei numeri o di teoria delle equazioni algebriche, oppure per comprendere più a fondo soluzioni già note di tali problemi. Un buon esempio dell'interazione fra teoria astratta e problemi concreti è la teoria di Galois. Galois la creò per rispondere a una domanda precisa: «Quali equazioni polinomiali in un'incognita hanno soluzioni esprimibili a partire dai coefficienti dati mediante operazioni razionali ed estrazioni di radici?». Per rispondere occorre anzitutto formulare il problema con precisione, e ciò richiede i concetti di campo, di estensione di campi e di campo di spezzamento di un polinomio. Per comprendere la soluzione di Galois del problema delle equazioni algebriche occorrono la nozione di gruppo e le proprietà dei gruppi risolubili. Nella teoria di Galois i risultati erano enunciati in termini di gruppi di permutazioni delle radici. In seguito si giunse a una comprensione assai più profonda passando dalle permutazioni delle radici alla nozione più astratta di gruppo degli automorfismi di un'estensione di campi. Tutto ciò verrà discusso ampiamente nel capitolo 4.

Naturalmente, una volta messi a punto gli strumenti per affrontare una classe di problemi, è probabile che si rivelino utili anche in altre circostanze. Essi generano inoltre problemi nuovi, che appaiono interessanti di per sé.

Nel corso dell'esposizione cercheremo di mettere in evidenza il ruolo della teoria generale nella soluzione di problemi interessanti, in particolare di origine classica. Per raggiungere alcuni dei teoremi di maggior interesse sarà necessario spingere la teoria oltre i fondamenti. Talvolta troveremo conveniente sviluppare alcune applicazioni negli esercizi. Per questa e per altre ragioni, svolgere un buon numero di esercizi è indispensabile per comprendere a fondo la materia.

Gli ingredienti fondamentali delle strutture che studieremo sono gli insiemi e le applicazioni. Probabilmente il lettore possiede già le cognizioni di teoria degli insiemi necessarie. Tuttavia, per fissare notazioni e terminologia e mettere in risalto gli aspetti che saranno fondamentali per noi, sembra opportuno richiamare brevemente alcuni elementi della teoria.[1] In vista di ciò che segue, conviene insistere sulle relazioni di equivalenza e sulla fattorizzazione di un'applicazione attraverso una relazione di equivalenza. Ritroveremo queste idee in moltissime forme nel corso del nostro studio. Nella seconda parte dell'introduzione tratteremo brevemente il sistema $\mathbb{Z}$ dei numeri interi e il sistema più elementare $\mathbb{N}$ dei numeri naturali, ossia i numeri con cui contiamo: $0, 1, 2, \ldots$, punto di partenza per lo sviluppo costruttivo dell'algebra. Data l'importanza oggi attribuita allo sviluppo dei sistemi numerici nelle scuole primarie e secondarie, sembra superfluo trattare $\mathbb{N}$ e $\mathbb{Z}$ nei particolari. Ci limiteremo quindi a delineare i passi principali di uno dei modi di introdurli e a dare dimostrazioni accurate di due risultati necessari per lo studio dei gruppi nel capitolo 1. Si tratta dell'esistenza dei massimi comuni divisori di numeri interi e del «teorema fondamentale dell'aritmetica», che stabilisce l'unicità della fattorizzazione di ogni numero naturale diverso da 0 e da 1 in un prodotto di numeri primi. Più avanti, nel capitolo 2, ritroveremo questi risultati come casi particolari dell'aritmetica dei domini a ideali principali.

[1] Come riferimento generale di teoria degli insiemi adeguato ai nostri scopi segnaliamo al lettore il delizioso volumetto di Paul R. Halmos, *Naive Set Theory*, Van Nostrand Reinhold, 1960.

0.1 L'insieme delle parti di un insieme

Iniziamo con una breve rassegna di alcune nozioni insiemistiche che avranno un ruolo essenziale in questo libro.

Sia S un insieme (o collezione) arbitrario di elementi che indichiamo con a, b, c e così via. La loro natura è irrilevante. Per indicare che un elemento a appartiene all'insieme S scriviamo $a \in S$ (talvolta $S \ni a$). La negazione di $a \in S$ si scrive $a \notin S$. Se S è un insieme finito di elementi a_i, $1 \leq i \leq n$, scriviamo $S = \{a_1, a_2, \ldots, a_n\}$. Ogni insieme S dà luogo a un altro insieme, $\mathcal{P}(S)$, costituito dai suoi sottoinsiemi. Fra questi figurano S stesso e il sottoinsieme vuoto, o insieme vuoto, che indichiamo con $\varnothing$. Per esempio, se $S = \{a_1, a_2, \ldots, a_n\}$ ha n elementi, allora $\mathcal{P}(S)$ comprende $\varnothing$, gli n insiemi $\{a_i\}$ di un solo elemento, gli $n(n-1)/2$ insiemi $\{a_i, a_j\}$ con $i \neq j$ di due elementi, e in generale $\binom{n}{i} = \dfrac{n!}{i!(n-i)!} = \dfrac{n(n-1)\cdots(n-i+1)}{1 \cdot 2 \cdots i}$ sottoinsiemi di i elementi. La cardinalità di $\mathcal{P}(S)$, ossia il numero dei suoi elementi, è dunque

$$1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = (1+1)^n = 2^n.$$

Chiameremo $\mathcal{P}(S)$ *insieme delle parti* di S.[2] Spesso specificheremo un sottoinsieme di S mediante una proprietà o un insieme di proprietà. A tal fine si scrive abitualmente

$$A = \{x \in S \mid \ldots\}$$

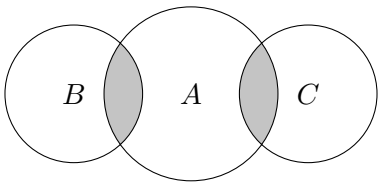
(o, se S risulta chiaro, $A = \{x \mid \ldots\}$), dove al posto dei puntini si elencano le proprietà che caratterizzano A. Per esempio, se $\mathbb{Z}$ indica l'insieme dei numeri interi, allora $\mathbb{N} = \{x \in \mathbb{Z} \mid x \geq 0\}$ definisce il sottoinsieme degli interi non negativi, o numeri naturali.

Se $A, B \in \mathcal{P}(S)$, cioè se A e B sono sottoinsiemi di S, diciamo che A è contenuto in B o che è un sottoinsieme di B (oppure che B contiene A), e scriviamo $A \subset B$ (oppure $B \supset A$), quando ogni elemento a di A appartiene anche a B. In simboli possiamo scrivere $a \in A \Rightarrow a \in B$, dove $\Rightarrow$ si legge «implica». L'affermazione $A = B$ equivale alle due affermazioni $A \supset B$ e $B \supset A$; in simboli, $A = B \Leftrightarrow (A \supset B \text{ e } B \supset A)$, dove $\Leftrightarrow$ si legge «se e solo se». Se $A \subset B$ e $A \neq B$, scriviamo $A \subsetneq B$ e diciamo che A è un sottoinsieme proprio di B. In alternativa possiamo scrivere $B \supsetneq A$.

Se A e B sono sottoinsiemi di S, il sottoinsieme di S formato dagli elementi c tali che $c \in A$ e $c \in B$ si chiama *intersezione* di A e B. Lo indichiamo con $A \cap B$. Se nessun elemento di S appartiene sia ad A sia a B, ossia se $A \cap B = \varnothing$, gli insiemi A e B si dicono *disgiunti*. L'*unione* (o somma logica) $A \cup B$ è il sottoinsieme degli elementi d tali che $d \in A$ oppure $d \in B$. Un'importante proprietà che lega $\cap$ e $\cup$ è la legge distributiva

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C). \tag{1}$$

La regione ombreggiata nella figura rappresenta i due membri della (1).



[2] Lo si chiama spesso anche *booleano di* S, $\mathcal{B}(S)$, dal nome di George Boole, che ne intraprese lo studio sistematico. La ragione del termine inglese *power set* è indicata nella nota 3 (p. 5 dell'originale).

Per dimostrare (1), sia $x \in A \cap (B \cup C)$. Poiché $x \in B \cup C$, si ha $x \in B$ oppure $x \in C$. Poiché $x \in A$, segue $x \in A \cap B$ oppure $x \in A \cap C$. Dunque $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Sia ora $y \in (A \cap B) \cup (A \cap C)$. Allora $y \in A \cap B$ oppure $y \in A \cap C$. In entrambi i casi $y \in A$, e inoltre $y \in B$ oppure $y \in C$. Dunque $y \in A \cap (B \cup C)$, da cui $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$. Abbiamo così entrambe le inclusioni, e quindi la (1).

Vale anche un'altra legge distributiva, duale della (1) nel senso che si ottiene da essa scambiando $\cup$ e $\cap$:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{2}$$

Lasciamo al lettore il compito di tracciare un diagramma per questa legge e di svolgerne la dimostrazione. Meglio ancora, il lettore potrà mostrare che la (2) è conseguenza della (1) e che, per simmetria, la (1) è conseguenza della (2).

Le intersezioni e le unioni si possono definire per un insieme arbitrario di sottoinsiemi di S. Sia Γ un tale insieme di sottoinsiemi (ossia un sottoinsieme di $\mathcal{P}(S)$). Definiamo

$$\bigcap_{A \in \Gamma} A = \{x \in S \mid x \in A \text{ per ogni } A \in \Gamma\},$$
$$\bigcup_{A \in \Gamma} A = \{x \in S \mid x \in A \text{ per qualche } A \in \Gamma\}.$$

Se Γ è finito, per esempio $\Gamma = \{A_1, A_2, \ldots, A_n\}$, scriviamo per l'intersezione anche $\bigcap_{i=1}^{n} A_i$ o $A_1 \cap A_2 \cap \cdots \cap A_n$, e usiamo una notazione simile per l'unione. È facile vedere che le leggi distributive si estendono a intersezioni e unioni arbitrarie:

$$B \cap \bigcup_{A \in \Gamma} A = \bigcup_{A \in \Gamma} (B \cap A), \qquad B \cup \bigcap_{A \in \Gamma} A = \bigcap_{A \in \Gamma} (B \cup A).$$

0.2 Il prodotto cartesiano. Applicazioni

Il lettore conosce certamente il ruolo centrale del concetto di funzione nella matematica e nelle sue applicazioni. Nei primi corsi di analisi si considera un sottoinsieme della retta reale $\mathbb{R}$, di solito un intervallo aperto o chiuso oppure l'intera retta, e una regola che associa a ogni elemento x del sottoinsieme un unico numero reale $f(x)$. Alla funzione così «definita» è associato il suo grafico nel piano numerico $\mathbb{R}^{(2)}$, costituito dai punti $(x, f(x))$. Ci si accorge presto che il grafico determina f e che lo caratterizza la seguente proprietà: ogni retta parallela all'asse y passante per un punto x del dominio (sull'asse x) incontra il grafico in uno e un solo punto. In particolare, se (x, y) e (x, y') appartengono al grafico, allora $y = y'$. Un grafico con questa proprietà è un oggetto ben definito, mentre l'idea intuitiva di una funzione data da una «regola» non ha la stessa precisione. Siamo così condotti a sostituire alla definizione originaria una definizione in termini del grafico.

Procederemo in questa direzione e sostituiremo alla parola «funzione» il termine geometrico «applicazione», oggi più comune nei contesti di cui ci occuperemo. Desideriamo inoltre passare dalle funzioni reali di una variabile reale alle applicazioni arbitrarie. Definiamo anzitutto il *prodotto cartesiano* $S \times T$ di due insiemi arbitrari S e T: è l'insieme delle coppie (s, t) con $s \in S$ e $t \in T$. Gli insiemi S e T possono coincidere. In $S \times T$ le coppie (s, t) e (s', t') sono uguali se e solo se $s = s'$ e $t = t'$. Se dunque S è formato da m elementi $s_1, \ldots, s_m$ e T da n elementi $t_1, \ldots, t_n$, il prodotto $S \times T$ è formato dalle mn coppie (s_i, t_j).

Possiamo ora definire un'*applicazione dell'insieme S nell'insieme T*. Essa consiste nell'insieme S, detto *dominio*, nell'insieme T, detto *codominio*, e in un sottoinsieme α di $S \times T$ (il *grafico*) con le due proprietà seguenti:

1. Per ogni $s \in S$ esiste un $t \in T$ tale che $(s, t) \in \alpha$.
2. Se (s, t) e (s, t') appartengono ad α, allora $t = t'$.

La seconda proprietà è detta *univocità*. Quando si dà una definizione, si dice spesso che «la funzione è ben definita» una volta accertato che vale la condizione 2. Insieme, le due condizioni affermano che per ogni $s \in S$ esiste un unico $t \in T$ tale che $(s, t) \in \alpha$. La notazione classica per questo t è $\alpha(s)$. Lo si chiama *immagine di s mediante α*. In molti libri di algebra, compresi i nostri precedenti, per $\alpha(s)$ si trovano le notazioni s^α e $s\alpha$. Esse offrono qualche vantaggio nella composizione delle applicazioni. Tuttavia, poiché la notazione classica $\alpha(s)$ è di gran lunga la più diffusa, la adotteremo in questo libro.

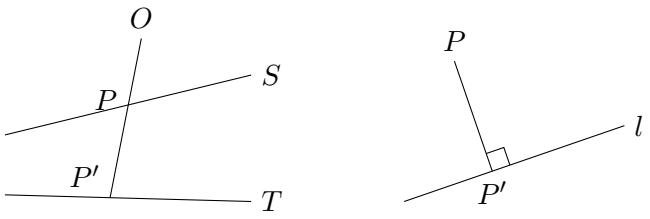
Due applicazioni sono uguali se e solo se hanno lo stesso dominio, lo stesso codominio e lo stesso grafico. Indicheremo con T^S l'insieme delle applicazioni «da S a T», cioè con dominio S e codominio T.[3]

Se A è un sottoinsieme di S, scriviamo $\alpha(A) = \{\alpha(a) \mid a \in A\}$ e lo chiamiamo *immagine di A mediante α*. In particolare $\alpha(S)$ si chiama *immagine dell'applicazione*; la indicheremo anche con $\operatorname{im}\alpha$. Quando il dominio e il codominio sono chiari, parleremo dell'«applicazione α» (o della «funzione α»), benché, a rigore, α sia soltanto una componente dell'applicazione.

Se S_1 è un sottoinsieme di S e α un'applicazione di S in T, restringendo il dominio a S_1 si ottiene un'applicazione di S_1 in T. Il suo grafico è il sottoinsieme di $S_1 \times T$ formato dalle coppie $(s_1, \alpha(s_1))$, con $s_1 \in S_1$. L'applicazione si chiama *restrizione di α a S_1* e si indica con $\alpha|S_1$. Inversamente, diremo che $\alpha : S \to T$ è un *prolungamento* di $\beta : S_1 \to T$ se $\beta = \alpha|S_1$.

[3] Se T ha due elementi, 0 e 1, possiamo scrivere $T = 2$ e considerare l'insieme 2^S delle applicazioni di S in $\{0, 1\}$. Una tale applicazione è caratterizzata dall'insieme $A = \{a \in S \mid \alpha(a) = 1\}$. Viceversa, a ogni sottoinsieme A di S si associa la sua funzione caratteristica χ_A, definita da $\chi_A(a) = 1$ se $a \in A$ e $\chi_A(a) = 0$ se $a \notin A$. Così 2^S si identifica con l'insieme dei sottoinsiemi di S, cioè con $\mathcal{P}(S)$. Di qui il nome inglese *power set*.

Come abbiamo ricordato, i termini inglesi *map* e *mapping* provengono dalla geometria. Consideriamo due esempi geometrici. Il primo è descritto dalla figura a sinistra.



Le rette S e T sono rispettivamente il dominio e il codominio. Si fissa un punto O esterno a entrambe e si manda il punto P di S nel punto P' in cui la retta OP incontra T. Queste applicazioni, dette *prospettività*, hanno un ruolo importante nella geometria proiettiva. Dal nostro punto di vista l'applicazione consiste negli insiemi S e T e nel sottoinsieme di $S \times T$ costituito dalle coppie (P, P'). Il secondo esempio, tratto dalla geometria euclidea, è la proiezione ortogonale su una retta (figura a destra). Il dominio è il piano, il codominio è la retta, e ogni punto P del piano viene mandato nel piede della perpendicolare condotta da P alla retta data. Si intende che se P appartiene a l, allora $P' = P$.

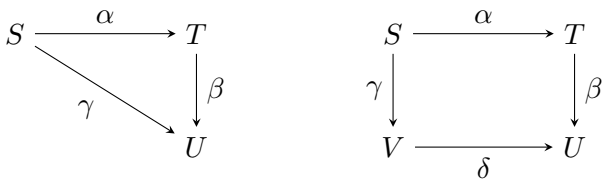
Come in questi esempi, nel trattare le applicazioni conviene tenere presente l'immagine intuitiva e riservare la definizione più precisa alle situazioni che richiedono maggiore rigore. La geometria suggerisce anche di indicare un'applicazione di S in T con $\alpha : S \to T$ oppure con $S \xrightarrow{\alpha} T$, e di specificarne l'azione scrivendo $x \mapsto y$, dove y è l'immagine di x: per esempio $P \mapsto P'$ nell'esempio precedente. Quando dominio e codominio coincidono, un'applicazione di S in sé si chiama spesso *trasformazione* di S.

Un'applicazione $S \xrightarrow{\alpha} T$ si dice *suriettiva* se $\operatorname{im}\alpha = T$, ossia se l'immagine coincide con il codominio. Si dice *iniettiva* se elementi distinti di S hanno immagini distinte in T, cioè se $s_1 \neq s_2 \Rightarrow \alpha(s_1) \neq \alpha(s_2)$. Se è sia iniettiva sia suriettiva si dice *biiettiva*, oppure si dice che α è una *corrispondenza biunivoca* fra S e T. Per esempio, la prospettività descritta sopra è biiettiva.

Siano $S \xrightarrow{\alpha} T$ e $T \xrightarrow{\beta} U$. Definiamo $S \xrightarrow{\beta\alpha} U$ come l'applicazione di dominio S, codominio U e grafico costituito dalle coppie $(s, \beta(\alpha(s)))$ di $S \times U$, per $s \in S$. Dunque, per definizione,

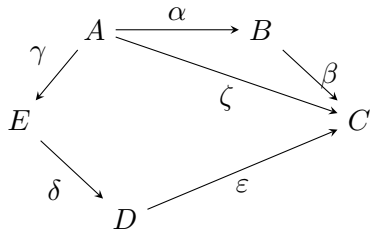
$$(\beta\alpha)(s) = \beta(\alpha(s)).$$

La chiamiamo *composta* (o *prodotto*, talvolta *risultante*) di α e β, con β applicata dopo α.[4] È spesso utile esprimere la relazione $\gamma = \beta\alpha$ dicendo che il triangolo a sinistra è *commutativo*. Analogamente, per $\alpha : S \to T$, $\beta : T \to U$, $\gamma : S \to V$ e $\delta : V \to U$, la relazione $\beta\alpha = \delta\gamma$ significa che il rettangolo a destra è commutativo.



In generale, quando ha senso parlarne, la commutatività di un diagramma significa che, fissati un punto iniziale e un punto finale, tutti i percorsi indicati fra quei punti danno la stessa applicazione. Per esempio, la commutatività di

[4] La composizione si scrive nell'ordine inverso a quello in cui si eseguono le operazioni: $\beta\alpha$ è α seguita da β. Per ricordare l'ordine conviene leggere $\beta\alpha$ come «β dopo α».

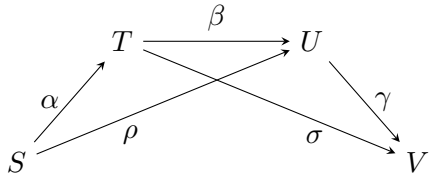


significa che $\beta\alpha = \zeta = \varepsilon(\delta\gamma)$.

La composizione delle applicazioni soddisfa la legge associativa: se $\alpha : S \to T$, $\beta : T \to U$ e $\gamma : U \to V$, allora $\gamma(\beta\alpha) = (\gamma\beta)\alpha$. Osserviamo anzitutto che le due applicazioni hanno lo stesso dominio S e lo stesso codominio V. Inoltre, per ogni $s \in S$,

$$(\gamma(\beta\alpha))(s) = \gamma((\beta\alpha)(s)) = \gamma(\beta(\alpha(s))),$$
$$((\gamma\beta)\alpha)(s) = (\gamma\beta)(\alpha(s)) = \gamma(\beta(\alpha(s))).$$

Le due applicazioni sono quindi identiche. La figura seguente illustra il risultato.



La legge associativa equivale ad affermare che se i triangoli STU e TUV sono commutativi, allora è commutativo l'intero diagramma.

Per ogni insieme S si definisce l'*applicazione identica* 1_S (o 1 quando S è chiaro) come $S \xrightarrow{1_S} S$, dove il grafico 1_S è il sottoinsieme delle coppie (s, s) di $S \times S$. Esso si chiama *diagonale* di $S \times S$. Se $\alpha : S \to T$, si verifica subito che $1_T\alpha = \alpha = \alpha 1_S$. Enunciamo ora un risultato importante.

Un'applicazione $\alpha : S \to T$ è biiettiva se e solo se esiste un'applicazione $\beta : T \to S$ tale che $\beta\alpha = 1_S$ e $\alpha\beta = 1_T$.

Dimostrazione. Supponiamo α biiettiva. Consideriamo il sottoinsieme β di $T \times S$ costituito dalle coppie $(\alpha(s), s)$. Dato $t \in T$, per la suriettività di α esiste $s \in S$ tale che $\alpha(s) = t$. Pertanto β soddisfa la condizione 1 nella definizione di applicazione da T a S. La condizione 2 segue dall'iniettività: se (t, s_1) e (t, s_2) appartengono a β, allora $\alpha(s_1) = t = \alpha(s_2)$, quindi $s_1 = s_2$. Abbiamo così ottenuto un'applicazione $\beta : T \to S$. Per $s \in S$, da $(s, \alpha(s)) \in \alpha$ e $(\alpha(s), s) \in \beta$ segue $\beta(\alpha(s)) = s$, dunque $\beta\alpha = 1_S$. Per $t \in T$ si ha $t = \alpha(s)$ per un $s \in S$, e $(t, s) \in \beta$, sicché $\beta(t) = s$. Ne segue $\alpha(\beta(t)) = \alpha(s) = t$, cioè $\alpha\beta = 1_T$. Viceversa, siano $\alpha : S \to T$ e $\beta : T \to S$ tali che $\beta\alpha = 1_S$ e $\alpha\beta = 1_T$. Dato $t \in T$, poniamo $s = \beta(t)$: allora $\alpha(s) = \alpha(\beta(t)) = t$, dunque α è suriettiva. Se poi $\alpha(s_1) = \alpha(s_2)$, si ha $s_1 = \beta(\alpha(s_1)) = \beta(\alpha(s_2)) = s_2$, e α è iniettiva. □

L'applicazione β con $\beta\alpha = 1_S$ e $\alpha\beta = 1_T$ è unica. Se infatti $\beta' : T \to S$ soddisfa le stesse condizioni, allora

$$\beta' = 1_S\beta' = (\beta\alpha)\beta' = \beta(\alpha\beta') = \beta 1_T = \beta.$$

Indicheremo β con α^{-1} e la chiameremo *inversa* dell'applicazione biiettiva α. Il risultato precedente mostra che α^{-1} è biiettiva e che $(\alpha^{-1})^{-1} = \alpha$.

Come prima applicazione del criterio di biiettività, dimostriamo formalmente un fatto di per sé abbastanza evidente: il prodotto di due applicazioni biiettive è biiettivo. Siano infatti

$\alpha : S \to T$ e $\beta : T \to U$ biiettive. Esistono le inverse $\alpha^{-1} : T \to S$ e $\beta^{-1} : U \to T$, e la composta $\alpha^{-1}\beta^{-1} : U \to S$. Inoltre,

$$(\beta\alpha)(\alpha^{-1}\beta^{-1}) = ((\beta\alpha)\alpha^{-1})\beta^{-1} = (\beta(\alpha\alpha^{-1}))\beta^{-1} = \beta\beta^{-1} = 1_U,$$
$$(\alpha^{-1}\beta^{-1})(\beta\alpha) = \alpha^{-1}(\beta^{-1}(\beta\alpha)) = \alpha^{-1}((\beta^{-1}\beta)\alpha) = \alpha^{-1}\alpha = 1_S.$$

Dunque $\alpha^{-1}\beta^{-1}$ è inversa di $\beta\alpha$, ossia

$$(\beta\alpha)^{-1} = \alpha^{-1}\beta^{-1}. \tag{3}$$

Questa formula importante è stata chiamata «principio del vestirsi e dello svestirsi»: gli indumenti si tolgono nell'ordine inverso rispetto a quello in cui si indossano. Si pensi alle calze e alle scarpe.

È importante estendere la nozione di prodotto cartesiano di due insiemi al prodotto di un numero finito qualsiasi di insiemi.[5] Dati gli insiemi $S_1, S_2, \ldots, S_r$, si definisce $\prod S_i$, o $S_1 \times S_2 \times \cdots \times S_r$, come l'insieme delle r-uple $(s_1, s_2, \ldots, s_r)$ con componente i-esima $s_i \in S_i$. Si pone $(s_1, \ldots, s_r) = (s'_1, \ldots, s'_r)$ se $s_i = s'_i$ per ogni i. Se tutti gli S_i coincidono con S, scriviamo $S^{(r)}$ per $\prod S_i$. Il prodotto cartesiano permette di definire le funzioni di due o più variabili: per esempio, una funzione di due variabili in S a valori in T è un'applicazione di $S \times S$ in T. Le applicazioni di $S^{(r)}$ in S si chiamano *composizioni r-arie* (o *prodotti r-ari*) sull'insieme S. Le strutture che considereremo nei primi due capitoli (monoidi, gruppi e anelli) sono definite mediante particolari composizioni binarie, cioè 2-arie, su S. Per ora ci limitiamo a registrare la definizione e a osservare che abbiamo già incontrato vari esempi di prodotti binari. Per esempio, in $\mathcal{P}(S)$ abbiamo i prodotti binari $A \cup B$ e $A \cap B$, cioè $(A, B) \mapsto A \cup B$ e $(A, B) \mapsto A \cap B$.

[5] E anche ai prodotti infiniti. Poiché in questo volume non ci serviranno, non ne discuteremo qui.

Esercizi

1. Si considerino le applicazioni $f : X \to Y$ e $g : Y \to Z$. Dimostrare che (a) se f e g sono iniettive, gf è iniettiva; (b) se gf è iniettiva, f è iniettiva; (c) se f e g sono suriettive, gf è suriettiva; (d) se gf è suriettiva, g è suriettiva. (e) Dare esempi di un insieme X e di un'applicazione $f : X \to X$ iniettiva ma non suriettiva, e viceversa. (f) Se gf è biiettiva, che cosa si può dire rispettivamente di f e di g quanto a iniettività e suriettività?

2. Dimostrare che $\alpha : S \to T$ è iniettiva se e solo se esiste $\beta : T \to S$ tale che $\beta\alpha = 1_S$, ed è suriettiva se e solo se esiste $\beta : T \to S$ tale che $\alpha\beta = 1_T$. In entrambi i casi esaminare l'affermazione: se β è unica, allora α è biiettiva.

3. Dimostrare che $\alpha : S \to T$ è suriettiva se e solo se non esistono due applicazioni β_1, β_2 di T in un insieme U tali che $\beta_1 \neq \beta_2$ e $\beta_1\alpha = \beta_2\alpha$. Dimostrare che α è iniettiva se e solo se non esistono due applicazioni γ_1, γ_2 di un insieme U in S tali che $\gamma_1 \neq \gamma_2$ e $\alpha\gamma_1 = \alpha\gamma_2$.

4. Sia $\alpha : S \to T$ e siano A, B sottoinsiemi di S. Dimostrare che $\alpha(A \cup B) = \alpha(A) \cup \alpha(B)$ e che $\alpha(A \cap B) \subset \alpha(A) \cap \alpha(B)$. Dare un esempio in cui $\alpha(A \cap B)$ non coincide con $\alpha(A) \cap \alpha(B)$.

5. Sia $\alpha : S \to T$ e sia A un sottoinsieme di S. Indichiamo con $\widetilde{A}$ il complementare di A in S, ossia l'insieme degli elementi di S che non appartengono ad A. Mostrare che, in generale, $\alpha(\widetilde{A}) \not\subset \widetilde{\alpha(A)}$. Che cosa accade se α è iniettiva? E se è suriettiva?

0.3 Relazioni di equivalenza. Fattorizzazione di un'applicazione attraverso una relazione di equivalenza

Diciamo che su un insieme S è definita una relazione (binaria) se, data una coppia ordinata (a, b) di elementi di S, possiamo stabilire se a è o no nella relazione data con b. Un esempio è la relazione d'ordine «$>$» nell'insieme dei numeri reali: dati due numeri reali a, b, si suppone di poter stabilire se $a > b$. Un altro esempio di relazione d'ordine è l'ordine lessicografico delle parole, dal quale dipende il loro posto in un dizionario. Fra le persone si ha, per esempio, la relazione di parentela fra cugini primi (a e b hanno un nonno in comune). Astraendo ciò che è essenziale in queste e in analoghe situazioni, siamo condotti a definire formalmente una *relazione (binaria) R su S* come un qualsiasi sottoinsieme del prodotto cartesiano $S \times S$. Se $(a, b) \in R$, diciamo che «a è nella relazione R con b» e scriviamo aRb. Particolarmente importanti per quanto segue sono le relazioni di equivalenza, che ora definiamo.

Una relazione E su S si chiama *relazione di equivalenza* se per ogni $a, b, c \in S$ valgono le seguenti condizioni:

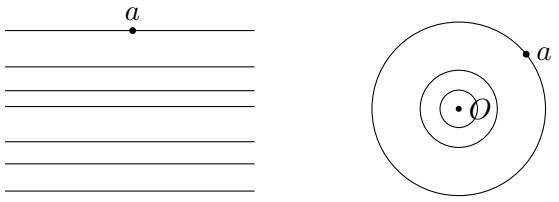
1. aEa (riflessività);
2. $aEb \Rightarrow bEa$ (simmetria);
3. aEb e $bEc \Rightarrow aEc$ (transitività).

Si ottiene un esempio prendendo come S l'insieme dei punti del piano e ponendo aEb se a e b giacciono sulla stessa retta orizzontale. Sul medesimo S si ottiene un'altra relazione di equivalenza E' ponendo $aE'b$ quando a e b sono equidistanti da uno stesso punto fissato, per esempio dall'origine O.

Mostreremo ora che il concetto di relazione di equivalenza equivale a quello di partizione di un insieme. Per *partizione* $\pi(S)$ di un insieme S intendiamo un insieme di sottoinsiemi non vuoti di S (ossia $\pi(S)$ è un sottoinsieme di $\mathcal{P}(S)$ che non contiene $\varnothing$) tale che l'unione dei suoi elementi sia tutto S e che due elementi distinti di $\pi(S)$ siano disgiunti. I sottoinsiemi che costituiscono $\pi(S)$ si chiamano *blocchi* della partizione. A ogni relazione di equivalenza E su S si può associare una partizione $\pi_E(S)$, e a ogni partizione π si può associare una relazione di equivalenza E_π. Le due corrispondenze sono reciproche nel senso che $\pi_{E_\pi} = \pi$ ed $E_{\pi_E} = E$. Supponiamo dapprima assegnata E. Per $a \in S$ poniamo

$$\bar{a}_E \text{ (o, più semplicemente, } \bar{a}) = \{b \in S \mid bEa\}.$$

Chiamiamo $\bar{a}_E$ la *classe di equivalenza determinata da a* relativamente a E, o *E-classe* di a. Nel primo esempio del capoverso precedente $\bar{a}_E$ è la retta orizzontale passante per a; nel secondo è la circonferenza di centro O passante per a.



In entrambi gli esempi le classi di equivalenza formano evidentemente una partizione del piano. Il fatto è generale. Sia $\{\bar{a} \mid a \in S\}$ l'insieme delle classi di equivalenza determinate da E. Poiché aEa, si ha $a \in \bar{a}$; ogni elemento di S appartiene perciò a una classe e $\bigcup_{a \in S} \bar{a} = S$. Osserviamo poi che $\bar{a} = \bar{b}$ se e solo se aEb. Supponiamo infatti aEb e sia $c \in \bar{a}$. Allora cEa e, per la condizione

3, cEb. Pertanto $c \in \bar{b}$, da cui $\bar{a} \subset \bar{b}$. Per la condizione 2 si ha anche bEa, sicché $\bar{b} \subset \bar{a}$. Dunque $\bar{a} = \bar{b}$. Viceversa, se $\bar{a} = \bar{b}$, poiché $a \in \bar{a} = \bar{b}$, la definizione di $\bar{b}$ dà aEb. Supponiamo ora che $\bar{a}$ e $\bar{b}$ abbiano un elemento c in comune. Allora cEa e cEb, quindi $\bar{a} = \bar{c} = \bar{b}$. Due classi di equivalenza distinte sono pertanto disgiunte. Segue che $\{\bar{a} \mid a \in S\}$ è una partizione di S. La indichiamo con π_E.

Viceversa, sia π una partizione qualsiasi di S. Ogni $a \in S$ appartiene a uno e un solo $A \in \pi$. Definiamo una relazione E_π ponendo $aE_\pi b$ se e solo se a e b appartengono allo stesso $A \in \pi$. Questa relazione è evidentemente riflessiva, simmetrica e transitiva, dunque è una relazione di equivalenza. È chiaro inoltre che la classe $\bar{a}$ di a relativamente a E_π è proprio il sottoinsieme della partizione π che contiene a. La partizione π_{E_π} associata a E_π è dunque la π assegnata. È altrettanto chiaro che, data una relazione di equivalenza E e posta $\pi_E = \{\bar{a} \mid a \in S\}$, la relazione E_{π_E}, per la quale due elementi sono equivalenti se e solo se appartengono alla medesima $\bar{a}$, coincide con E.

Se E è una relazione di equivalenza, la partizione associata $\pi = \{\bar{a} \mid a \in S\}$ si chiama *insieme quoziente di S rispetto alla relazione E*. Generalmente indicheremo π con S/E. Ricordiamo che gli elementi di S/E sono sottoinsiemi di S: il quoziente è dunque un sottoinsieme di $\mathcal{P}(S)$. Consideriamo l'applicazione ν di S in S/E definita da

$$\nu : a \longmapsto \bar{a}.$$

La chiamiamo *applicazione naturale* di S sull'insieme quoziente S/E. Chiaramente ν è suriettiva.

Esaminiamo ora alcuni importanti legami fra applicazioni e relazioni di equivalenza. Sia $\alpha : S \to T$. Si può definire una relazione E_α su S ponendo $aE_\alpha b$ se e solo se $\alpha(a) = \alpha(b)$. È chiaro che si tratta di una relazione di equivalenza. Per $c \in T$ poniamo

$$\alpha^{-1}(c) = \{a \in S \mid \alpha(a) = c\} \tag{4}$$

e chiamiamo questo insieme *immagine inversa*, o *controimmagine*, di c. Più in generale, se C è un sottoinsieme di T, definiamo

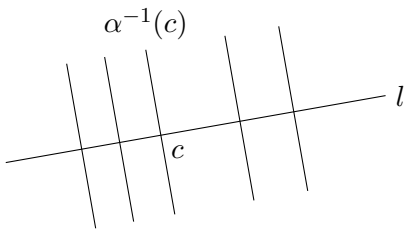
$$\alpha^{-1}(C) = \{a \in S \mid \alpha(a) \in C\}. \tag{5}$$

Chiaramente $\alpha^{-1}(C) = \bigcup_{c \in C} \alpha^{-1}(c)$, e $\alpha^{-1}(c) = \varnothing$ se $c \notin \operatorname{im} \alpha$. D'altra parte, se $c = \alpha(a)$ per qualche $a \in S$, allora

$$\alpha^{-1}(c) = \alpha^{-1}(\alpha(a)) = \{b \in S \mid \alpha(b) = \alpha(a)\},$$

che è appunto la classe di equivalenza $\bar{a}_{E_\alpha}$ determinata da a in S. Chiameremo questo sottoinsieme di S anche *fibra sopra l'elemento* $c \in \operatorname{im} \alpha$. L'insieme di tali fibre costituisce la partizione di S determinata da E_α: le fibre sono gli elementi del quoziente S/E_α.

Sia, per esempio, α la proiezione ortogonale del piano su una sua retta l, come nel § 0.2. Se c è un punto della retta, la fibra $\alpha^{-1}(c)$ è l'insieme dei punti della perpendicolare a l passante per c.



Si ottiene una biiezione dell'insieme di queste fibre su l associando alla fibra $\alpha^{-1}(c)$ il punto c, sua intersezione con l.

In generale α definisce un'applicazione $\bar{\alpha}$ di S/E_α in T. Scrivendo $\bar{a}$ per $\bar{a}_{E_\alpha} = \alpha^{-1}(\alpha(a))$, basta porre

$$\bar{\alpha}(\bar{a}) = \alpha(a). \tag{6}$$

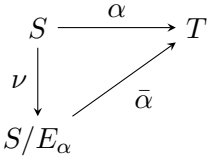
Poiché $\bar{a} = \bar{b}$ se e solo se $\alpha(a) = \alpha(b)$, il secondo membro è indipendente dalla scelta di a nella classe $\bar{a}$, e abbiamo effettivamente un'applicazione. Chiamiamo $\bar{\alpha}$ l'applicazione di S/E_α *indotta* da α. È iniettiva: da $\bar{\alpha}(\bar{a}) = \bar{\alpha}(\bar{b})$ segue $\alpha(a) = \alpha(b)$, e quindi $\bar{a} = \bar{b}$ per la definizione di E_α. Naturalmente, se α è già iniettiva, allora $aE_\alpha b$, ossia $\alpha(a) = \alpha(b)$, implica $a = b$. In questo caso S/E_α si può identificare con S e $\bar{\alpha}$ con α.

Osserviamo ora che $\bar{\alpha}(\nu(a)) = \bar{\alpha}(\bar{a}) = \alpha(a)$. Abbiamo così la fattorizzazione

$$\alpha = \bar{\alpha}\nu \tag{7}$$

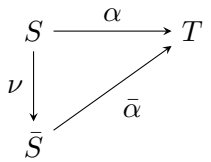
dell'applicazione data nel prodotto dell'applicazione naturale ν di S in S/E_α e dell'applicazione indotta $\bar{\alpha}$ di S/E_α in T. L'applicazione $\bar{\alpha}$ è iniettiva e ν è suriettiva. La (7) equivale alla commutatività del diagramma

(8)



Poiché ν è suriettiva, $\operatorname{im}\alpha = \operatorname{im}\bar{\alpha}$. Pertanto $\bar{\alpha}$ è biiettiva se e solo se α è suriettiva. Osserviamo infine che $\bar{\alpha}$ è la sola applicazione di S/E_α in T che rende commutativo il diagramma (8). Infatti, se $\beta : S/E_\alpha \to T$ soddisfa $\beta\nu = \alpha$, allora $\beta(\bar{a}) = \beta(\nu(a)) = \alpha(a)$, e la definizione (6) dà $\beta = \bar{\alpha}$.

Queste semplici considerazioni ammettono un'utile generalizzazione. Siano date un'applicazione $\alpha : S \to T$ e una relazione di equivalenza E su S. Diremo che α è *compatibile* con E se, per $a, b \in S$, da aEb segue $\alpha(a) = \alpha(b)$. In tal caso possiamo definire un'applicazione $\bar{\alpha}$ di $\bar{S} = S/E$ in T ponendo $\bar{a} = \bar{a}_E \mapsto \alpha(a)$. L'applicazione è ben definita. Se ν è la suriezione naturale $a \mapsto \bar{a}$, allora $\alpha = \bar{\alpha}\nu$, ossia è commutativo il diagramma



In questo caso l'applicazione indotta $\bar{\alpha}$ può non essere iniettiva. Precisamente, è iniettiva se e solo se $E = E_\alpha$.

I risultati di questo paragrafo, che a questo punto potrebbero sembrare una tediosa raccolta di ovvietà, avranno un ruolo fondamentale nel seguito.

Esercizi

1. Sia $\mathbb{N} = \{0, 1, 2, \ldots\}$. Mostrare che le seguenti sono partizioni di $\mathbb{N}$:

   (i) $\{0, 2, 4, \ldots, 2k, \ldots\}$, $\{1, 3, 5, \ldots, 2k+1, \ldots\}$, con $k \in \mathbb{N}$;

   (ii) $\{0, 3, 6, \ldots, 3k, \ldots\}$, $\{1, 4, 7, \ldots, 3k+1, \ldots\}$, $\{2, 5, 8, \ldots, 3k+2, \ldots\}$.

2. Sia $\mathbb{N}$ come nell'esercizio 1 e sia $\mathbb{N}^{(2)} = \mathbb{N}\times\mathbb{N}$. In $\mathbb{N}^{(2)}$ si ponga $(a,b) \sim (c,d)$ se $a+d = b+c$. Verificare che $\sim$ è una relazione di equivalenza.

3. Sia S l'insieme dei segmenti orientati PQ (di punto iniziale P e punto finale Q) del piano euclideo. Per quale relazione di equivalenza su S l'insieme quoziente è l'usuale insieme dei vettori del piano?

4. Dati gli insiemi S e T, chiamiamo *corrispondenza* da S a T un sottoinsieme di $S \times T$. (La definizione comprende sia le applicazioni sia le relazioni.) Se C è una corrispondenza da S a T, si definisce C^{-1} come la corrispondenza da T a S formata dalle coppie (t,s) tali che $(s,t) \in C$. Se C va da S a T e D da T a U, si definisce DC da S a U come l'insieme delle coppie $(s,u) \in S \times U$ per le quali esiste $t \in T$ tale che $(s,t) \in C$ e $(t,u) \in D$. Verificare la legge associativa $(ED)C = E(DC)$ e la legge dell'identità $C1_S = C = 1_T C$.

5. Dimostrare che una relazione E su S è di equivalenza se e solo se (i) $E \supset 1_S$, (ii) $E = E^{-1}$, (iii) $E \supset EE$.

6. Sia C una relazione binaria su S. Per $r = 1, 2, 3, \ldots$ si definisca C^r come l'insieme delle coppie (s,t) per le quali esistono $s_1, \ldots, s_{r-1} \in S$ con $sCs_1,\ s_1Cs_2, \ldots, s_{r-1}Ct$. Sia

$$E = 1_S \cup (C \cup C^{-1}) \cup (C \cup C^{-1})^2 \cup (C \cup C^{-1})^3 \cup \cdots .$$

   Mostrare che E è una relazione di equivalenza e che ogni relazione di equivalenza su S che contiene C contiene anche E. Si dice che E è la relazione di equivalenza *generata* da C.

7. Quante relazioni binarie distinte vi sono su un insieme S di due elementi? Di tre elementi? Di n elementi? Quante di esse sono relazioni di equivalenza?

8. Siano $S \xrightarrow{\alpha} T \xrightarrow{\beta} U$. Mostrare che per ogni $U_1 \subset U$ si ha $(\beta\alpha)^{-1}(U_1) = \alpha^{-1}(\beta^{-1}(U_1))$.

9. Sia $\alpha : S \to T$ e siano C, D sottoinsiemi di T. Mostrare che $\alpha^{-1}(C \cup D) = \alpha^{-1}(C) \cup \alpha^{-1}(D)$ e che $\alpha^{-1}(C \cap D) = \alpha^{-1}(C) \cap \alpha^{-1}(D)$ (si confronti l'esercizio 4 del § 0.2).

10. Siano $\mathbb{C}$ l'insieme dei numeri complessi e $\mathbb{R}^+$ l'insieme dei reali non negativi. Sia $f : \mathbb{C} \to \mathbb{R}^+$ l'applicazione $z \mapsto |z|$ (modulo di z). Qual è la relazione di equivalenza che f definisce su $\mathbb{C}$?

11. Sia $\mathbb{C}^*$ l'insieme dei numeri complessi diversi da zero e sia g l'applicazione $z \mapsto |z|^{-1}z$. Qual è la relazione di equivalenza che g definisce su $\mathbb{C}^*$?

0.4 I numeri naturali

Il sistema dei numeri naturali, ossia dei numeri $0, 1, 2, 3, \ldots$ con cui contiamo, è fondamentale in algebra sotto due aspetti. Anzitutto serve come punto di partenza per la costruzione di sistemi più elaborati: i numeri interi, i razionali e infine i reali, l'anello delle classi di resti modulo un intero e così via. In secondo luogo, nello studio di alcune strutture algebriche hanno un ruolo importante certe applicazioni dell'insieme dei naturali nella struttura considerata. Per esempio, in una struttura S dotata di composizione binaria associativa e di unità, ogni elemento $a \in S$ definisce un'applicazione $n \mapsto a^n$, dove $a^0 = 1$, $a^1 = a$ e $a^k = a^{k-1}a$. Tali applicazioni servono a studiare S.

Un punto di partenza comodo e tradizionale per lo studio del sistema $\mathbb{N}$ è l'assiomatizzazione dovuta a Peano. Da questo punto di vista partiamo da un insieme non vuoto $\mathbb{N}$, da un particolare elemento di $\mathbb{N}$, designato con 0, e da un'applicazione $a \mapsto a^+$ di $\mathbb{N}$ in sé, detta *applicazione successore*. Gli assiomi di Peano sono:

1. $0 \neq a^+$ per ogni a (cioè 0 non appartiene all'immagine di $\mathbb{N}$ mediante $a \mapsto a^+$).

2. L'applicazione $a \mapsto a^+$ è iniettiva.

3. *Assioma di induzione.* Ogni sottoinsieme di $\mathbb{N}$ che contiene 0 e che, con ogni suo elemento, ne contiene anche il successore coincide con $\mathbb{N}$.

L'assioma 3 è alla base delle dimostrazioni mediante il *primo principio di induzione*. Ecco il suo enunciato. A ciascun numero naturale n sia associata una proposizione $E(n)$, per esempio $0+1+2+\cdots+n = n(n+1)/2$. Supponiamo vera $E(0)$ e supponiamo che $E(r^+)$ sia vera ogni volta che è vera $E(r)$ (questa seconda parte è detta *passo induttivo*). Allora $E(n)$ è vera per ogni $n \in \mathbb{N}$. Il risultato segue immediatamente dall'assioma 3. Sia S il sottoinsieme degli $s \in \mathbb{N}$ per cui $E(s)$ è vera. Allora $0 \in S$, e se $r \in S$ anche $r^+ \in S$. Per l'assioma 3, $S = \mathbb{N}$, dunque $E(n)$ vale per ogni numero naturale.

Le dimostrazioni per induzione sono molto comuni in matematica e certamente familiari al lettore. Abbastanza spesso, senza esserne consapevoli, si incontrano anche *definizioni* per induzione. Un esempio è la definizione di a^n già ricordata: $a^0 = 1$, $a^{r+1} = a^r a$. Una definizione per induzione è meno ovvia di quanto possa sembrare a prima vista. Il suo significato preciso è dato dal seguente risultato.

Teorema di ricorsione. *Siano S un insieme, φ un'applicazione di S in sé e a un elemento di S. Allora esiste una e una sola applicazione f di $\mathbb{N}$ in S tale che*[6]

$$1. \quad f(0) = a, \qquad 2. \quad f(n^+) = \varphi(f(n)), \quad n \in \mathbb{N}.$$

Dimostrazione. Consideriamo il prodotto $\mathbb{N} \times S$. Sia Γ l'insieme dei sottoinsiemi U di $\mathbb{N} \times S$ con le due proprietà: (i) $(0, a) \in U$; (ii) se $(n, b) \in U$ allora $(n^+, \varphi(b)) \in U$. Poiché $\mathbb{N} \times S$ le possiede, $\Gamma \neq \varnothing$. Sia f l'intersezione di tutti gli $U \in \Gamma$. Mostriamo che f è la funzione cercata. In primo luogo, per induzione, per ogni $n \in \mathbb{N}$ esiste $b \in S$ tale che $(n, b) \in f$. Per mostrare che f è un'applicazione resta da provare che, se (n, b) e (n, b') appartengono a f, allora $b = b'$. È lo stesso che provare che il sottoinsieme T degli n per i quali vale questa implicazione è tutto $\mathbb{N}$. Procediamo per induzione.

Anzitutto $0 \in T$. Altrimenti in f si avrebbero $(0, a)$ e $(0, a')$ con $a' \neq a$. Sia f' il sottoinsieme ottenuto eliminando $(0, a')$ da f. È immediato che f' soddisfa le condizioni (i) e (ii) che definiscono gli elementi di Γ. Quindi $f' \supset f$. Ma $f' \subsetneq f$, essendosi eliminato $(0, a')$: questa contraddizione prova che $0 \in T$. Supponiamo ora che per un naturale r si abbia $r \in T$ e $r^+ \notin T$. Sia $(r, b) \in f$. Allora $(r^+, \varphi(b)) \in f$ e, poiché $r^+ \notin T$, esiste un $c \neq \varphi(b)$ con $(r^+, c) \in f$. Consideriamo il sottoinsieme f' di f ottenuto eliminando (r^+, c). Poiché $r^+ \neq 0$ e $(0, a) \in f$, si ha $(0, a) \in f'$. Lo stesso argomento, usando l'iniettività del successore, mostra che, se $n \neq r$ e $(n, d) \in f'$, allora $(n^+, \varphi(d)) \in f'$. Se invece $(r, b') \in f'$, si ha $b' = b$ e $(r^+, \varphi(b)) \in f'$, perché nel passare a f' non abbiamo eliminato questa coppia. Ne segue che $f' \in \Gamma$, e di nuovo otteniamo la contraddizione $f' \supset f$, $f' \subsetneq f$. Abbiamo così provato che $r \in T$ implica $r^+ \in T$. Per induzione $T = \mathbb{N}$, ed è quindi dimostrata l'esistenza della funzione f che soddisfa le condizioni assegnate. Quanto all'unicità, se g è un'altra applicazione che le soddisfa, $g \in \Gamma$, dunque $g \supset f$. Per due applicazioni g, f dello stesso dominio ciò implica $g = f$, per la definizione di applicazione. Pertanto f è unica. □

[6] Si sarebbe tentati di dire che le condizioni 1 e 2 *definiscono* induttivamente f. Tuttavia questo non ha senso di per sé: per parlare di una funzione su $\mathbb{N}$ occorre una definizione a priori di $f(n)$ per ogni $n \in \mathbb{N}$. Per dimostrare l'esistenza di f si devono usare tutti gli assiomi di Peano; lo illustra l'esercizio 4 a fine paragrafo. Per una trattazione più ampia si veda Leon Henkin, «On mathematical induction», *American Mathematical Monthly*, 67 (1960), pp. 323–338. Henkin dimostra il teorema mediante funzioni «parziali» su $\mathbb{N}$. La dimostrazione che daremo si deve, indipendentemente, a P. Lorenzen e a D. Hilbert e P. Bernays congiuntamente.

Il teorema di ricorsione permette di definire l'addizione e la moltiplicazione dei numeri naturali. Per definire l'addizione di m a n si prende $a = m$ e come φ l'applicazione successore $n \mapsto n^+$. Si ottengono le due formule

(a) $$0 + m = m,$$
(b) $$n^+ + m = (n + m)^+.$$

Per la moltiplicazione per m si prende $a = 0$ e come φ l'applicazione $n \mapsto n + m$. Si hanno così

(a) $$0m = 0,$$
(b) $$n^+m = nm + m.$$

Si possono dimostrare le leggi associative, commutative e di cancellazione per l'addizione e la moltiplicazione:[7]

A1	$(x + y) + z = x + (y + z)$	(associatività),
A2	$x + y = y + x$	(commutatività),
A3	$x + z = y + z \;\Rightarrow\; x = y$	(cancellazione),
M1	$(xy)z = x(yz)$,	
M2	$xy = yx$,	
M3	$xz = yz, \quad z \neq 0 \;\Rightarrow\; x = y.$	

Vale inoltre la regola fondamentale che lega addizione e moltiplicazione:

D $$z(x + y) = zx + zy \qquad \text{(legge distributiva)}.$$

Per il sistema $\mathbb{N}$ è fondamentale la relazione d'ordine. Si dice che il naturale a è maggiore o uguale al naturale b (in simboli $a \geq b$ o $b \leq a$) se l'equazione $a = b + x$ ha una soluzione $x \in \mathbb{N}$. La relazione gode delle seguenti proprietà:

O1	$x \geq y$ e $y \geq x \iff x = y$;
O2	$x \geq y$ e $y \geq z \implies x \geq z$;
O3	per ogni $(x, y) \in \mathbb{N} \times \mathbb{N}, \quad x \geq y$ oppure $y \geq x$.

L'insieme dei naturali gode inoltre della seguente proprietà di *buon ordinamento*.

O4 *Ogni sottoinsieme non vuoto S di $\mathbb{N}$ ha un minimo: esiste $l \in S$ tale che $l \leq s$ per ogni $s \in S$.*

Dimostrazione. Sia M l'insieme dei naturali m tali che $m \leq s$ per ogni $s \in S$. Si ha $0 \in M$ e, se $s \in S$, allora $s^+ \notin M$. Dunque $M \neq \mathbb{N}$ e per l'assioma di induzione esiste $l \in M$ con $l^+ \notin M$. Il numero cercato è l: infatti $l \leq s$ per ogni $s \in S$. Inoltre $l \in S$, perché altrimenti $l < s$ per ogni $s \in S$ e quindi $l^+ \leq s$ per ogni $s \in S$, contro $l^+ \notin M$. □

La proprietà di buon ordinamento è alla base del seguente *secondo principio di induzione*. Per ogni $n \in \mathbb{N}$ sia data una proposizione $E(n)$. Supponiamo di poter dimostrare, per ciascun r, che $E(r)$ è vera se $E(s)$ è vera per ogni $s < r$. (In particolare, questo presuppone che si sappia dimostrare $E(0)$.) Allora $E(n)$ è vera per ogni n. Per provarlo occorre mostrare che il sottoinsieme F degli r per i quali $E(r)$ è falsa è vuoto. Se F non fosse vuoto, per O4 avrebbe un

[7] Dimostrazioni dettagliate si trovano in E. Landau, *Foundations of Analysis*, 2ª ed., Chelsea Publishing Co., New York, 1960. Una traccia delle dimostrazioni si trova in Paul R. Halmos, *Naive Set Theory*, Van Nostrand Reinhold, New York, 1960.

minimo t. Allora $E(t)$ sarebbe falsa e $E(s)$ vera per ogni $s < t$, in contraddizione con l'ipotesi. Dunque $F = \varnothing$.

I legami principali fra ordine e addizione e fra ordine e moltiplicazione sono i seguenti:

OA $\quad a \geq b \;\Rightarrow\; a + c \geq b + c,$

OM $\quad a \geq b \;\Rightarrow\; ac \geq bc.$

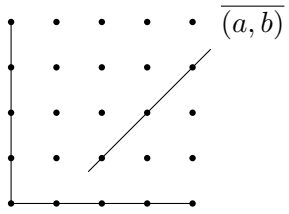
Esercizi

1. Dimostrare che, se $a \geq b$ e $c \geq d$, allora $a + c \geq b + d$ e $ac \geq bd$.
2. Dimostrare la seguente estensione del primo principio di induzione. Sia $s \in \mathbb{N}$ e sia assegnata una proposizione $E(n)$ per ogni $n \geq s$. Supponiamo che valga $E(s)$ e che, se vale $E(r)$ per qualche $r \geq s$, valga anche $E(r^+)$. Allora $E(n)$ è vera per ogni $n \geq s$. Enunciare e dimostrare l'analoga estensione del secondo principio.
3. Dimostrare per induzione che se c è reale, $c \geq -1$ e $n \in \mathbb{N}$, allora $(1 + c)^n \geq 1 + nc$.
4. (Henkin.) Sia $N = \{0, 1\}$ e si ponga $0^+ = 1$, $1^+ = 1$. Mostrare che N soddisfa gli assiomi di Peano 1 e 3, ma non il 2. Sia $\varphi : N \to N$ definita da $\varphi(0) = 1$ e $\varphi(1) = 0$. Mostrare che il teorema di ricorsione viene meno per N e questa φ: non esiste alcuna applicazione $f : N \to N$ tale che $f(0) = 0$, $f(n^+) = \varphi(f(n))$.
5. Dimostrare A1 e M2.

0.5 Il sistema $\mathbb{Z}$ dei numeri interi

Costruiremo il sistema degli interi in un modo che ci sembra più naturale e intuitivo dell'usuale procedimento di aggiungere a $\mathbb{N}$ gli opposti dei suoi elementi. Il metodo è inoltre analogo a quello consueto con cui si costruisce il sistema $\mathbb{Q}$ dei numeri razionali a partire da $\mathbb{Z}$.

Partiamo dal prodotto cartesiano $\mathbb{N} \times \mathbb{N}$. In esso introduciamo la relazione $(a, b) \sim (c, d)$ se $a + d = b + c$. È facile verificare che è una relazione di equivalenza. Nel dare questa definizione intendiamo far svolgere alla classe $\overline{(a, b)}$ determinata da (a, b) il ruolo della differenza fra a e b. Se rappresentiamo come al solito la coppia (a, b) mediante il punto di ascissa a e ordinata b, allora $\overline{(a, b)}$ è l'insieme dei punti a coordinate naturali sulla retta di pendenza 1 passante per (a, b). Chiamiamo *interi* le classi $\overline{(a, b)}$ e ne indichiamo la totalità con $\mathbb{Z}$.



Prima di definire l'addizione osserviamo che, se $(a, b) \sim (a', b')$ e $(c, d) \sim (c', d')$, allora

$$(a + c, b + d) \sim (a' + c', b' + d').$$

Infatti per ipotesi $a + b' = a' + b$ e $c + d' = c' + d$. Pertanto $a + c + b' + d' = a' + c' + b + d$, che è proprio la relazione voluta. Dunque l'intero $\overline{(a + c, b + d)}$ è determinato univocamente da $\overline{(a, b)}$ e $\overline{(c, d)}$. Lo definiamo come loro *somma*:

$$\overline{(a, b)} + \overline{(c, d)} = \overline{(a + c, b + d)}.$$

È facile verificare le regole A1, A2 e A3. Inoltre $(a,a) \sim (b,b)$ e, se poniamo $0 = \overline{(a,a)}$ (da distinguere per ora dallo 0 di $\mathbb{N}$), allora

A4 $$0 + x = x \quad \text{per ogni } x \in \mathbb{Z}.$$

Infine ogni intero ha un opposto: se $x = \overline{(a,b)}$, indichiamo con $-x$ la classe $\overline{(b,a)}$, indipendente dal rappresentante (a,b) di x. Si ha

A5 $$x + (-x) = 0.$$

Osserviamo ora che, se $(a,b) \sim (a',b')$ e $(c,d) \sim (c',d')$, allora $a + b' = a' + b$ e $c + d' = c' + d$. Ne segue

$$\begin{aligned} &c(a+b') + d(a'+b) + a'(c+d') + b'(c'+d) \\ &\qquad = c(a'+b) + d(a+b') + a'(c'+d) + b'(c+d'). \end{aligned}$$

Sviluppando,

$$\begin{aligned} &ac + b'c + a'd + bd + a'c + a'd' + b'c' + b'd \\ &\qquad = a'c + bc + ad + b'd + a'c' + a'd + b'c + b'd'. \end{aligned}$$

Per la legge di cancellazione,

$$ac + bd + a'd' + b'c' = bc + ad + a'c' + b'd',$$

il che prova che $(ac + bd, ad + bc) \sim (a'c' + b'd', a'd' + b'c')$. Ponendo dunque

$$\overline{(a,b)}\,\overline{(c,d)} = \overline{(ac + bd, ad + bc)}$$

otteniamo un prodotto univocamente determinato. Si verifica che esso è associativo e commutativo, e distributivo rispetto all'addizione. Vale la legge di cancellazione purché il fattore z da cancellare non sia zero.

Poniamo $\overline{(a,b)} \geq \overline{(c,d)}$ se $a + d \geq b + c$. La relazione è ben definita, cioè è indipendente dai rappresentanti scelti nelle classi. Si verifica facilmente che valgono O1, O2, O3 e OA. La proprietà OM va modificata come segue:

OM′ $$\text{Se } x \geq y \text{ e } z \geq 0, \text{ allora } xz \geq yz.$$

Consideriamo l'insieme $\mathbb{N}'$ degli interi non negativi. Per definizione è il sottoinsieme di $\mathbb{Z}$ degli $x \geq 0$, ossia degli elementi della forma $\overline{(b+u,b)}$. È immediato che $(b+u,b) \sim (c+u,c)$. Per ogni numero naturale u (elemento di $\mathbb{N}$) poniamo $u' = \overline{(b+u,b)}$. Le nostre osservazioni mostrano che $u \mapsto u'$ definisce un'applicazione di $\mathbb{N}$ in $\mathbb{Z}$ con immagine $\mathbb{N}'$. Inoltre da $(b+u,b) \sim (c+v,c)$ segue $b+u+c = b+c+v$, quindi $u = v$. L'applicazione è dunque iniettiva. Lasciamo al lettore la verifica delle proprietà

$$(u+v)' = u' + v', \qquad (uv)' = u'v', \qquad u \geq v \iff u' \geq v'.$$

Queste proprietà e il fatto che $u \mapsto u'$ sia una biiezione di $\mathbb{N}$ su $\mathbb{N}'$ mostrano che, quanto alle operazioni fondamentali e all'ordine, i due sistemi sono indistinguibili. Possiamo quindi sostituire al sistema originario dei numeri naturali l'insieme degli interi non negativi, sottoinsieme di $\mathbb{Z}$, e usare per esso le notazioni che appartenevano a $\mathbb{N}$. D'ora in poi lo indicheremo dunque con $\mathbb{N}$ e i suoi elementi con $0, 1, 2, \ldots$. Si vede facilmente che gli altri elementi di $\mathbb{Z}$ si possono elencare come $-1, -2, \ldots$.

Esercizi

1. Mostrare che $x \geq y \iff -x \leq -y$.

2. Dimostrare che ogni insieme non vuoto S di interi limitato inferiormente (superiormente), nel senso che esiste un intero b (rispettivamente B) con $b \leq s$ (rispettivamente $B \geq s$) per ogni $s \in S$, possiede un minimo (rispettivamente un massimo).

3. Si definisca $|x| = x$ per $x \geq 0$ e $|x| = -x$ per $x < 0$. Dimostrare che $|xy| = |x||y|$ e che $|x + y| \leq |x| + |y|$.

0.6 Alcuni fatti aritmetici fondamentali su $\mathbb{Z}$

Diremo che l'intero b è un *fattore* o *divisore* dell'intero a se esiste $c \in \mathbb{Z}$ tale che $a = bc$. Si dice anche che a è un *multiplo* di b e si scrive $b \mid a$. La relazione è chiaramente transitiva. Se $b \mid a$ e $a \mid b$, abbiamo $a = bc$ e $b = ad$, dunque $a = adc$. Se $a \neq 0$, per cancellazione $dc = 1$. Allora $|d||c| = 1$ e $d = \pm 1$, $c = \pm 1$. Pertanto, se $b \mid a$ e $a \mid b$ con $a \neq 0$, si ha $b = \pm a$. Un intero p si dice *primo* (o *irriducibile*) se $p \neq 0, \pm 1$ e i suoi soli divisori sono $\pm p$ e ± 1. Se p è primo, anche $-p$ lo è.

Punto di partenza della teoria dei numeri è il fatto che ogni intero positivo diverso da 1 si scrive in uno e un solo modo come prodotto di primi positivi: $a = p_1 p_2 \cdots p_s$, con p_i primi e $s \geq 1$. Unicità significa qui «unicità a meno dell'ordine dei fattori». È questo il *teorema fondamentale dell'aritmetica*. Ne daremo una dimostrazione dovuta a E. Zermelo e fondata sull'induzione matematica.

Sia $n > 1$ un intero. O n è primo oppure $n = n_1 n_2$, con n_1, n_2 maggiori di 1 e dunque minori di n. Supponendo che ogni intero maggiore di 1 e minore di n sia un prodotto di primi positivi, ciò vale per n_1 e n_2, quindi per $n = n_1 n_2$. Per il secondo principio di induzione ogni intero maggiore di 1 è dunque un prodotto di primi positivi. Resta da provare l'unicità. Sia

$$n = p_1 p_2 \cdots p_s = q_1 q_2 \cdots q_t,$$

dove i p_i e i q_j sono primi positivi. Supponiamo dapprima $p_1 = q_1$. Cancellando questo fattore otteniamo $m = p_2 \cdots p_s = q_2 \cdots q_t < n$. Se $m = 1$ abbiamo finito; altrimenti per l'ipotesi induttiva $p_2, \ldots, p_s$ coincidono, a meno dell'ordine, con $q_2, \ldots, q_t$. Lo stesso vale quindi per le due liste iniziali. L'unicità segue così per n. Supponiamo ora $p_1 \neq q_1$ e, per fissare le idee, $p_1 < q_1$. Necessariamente $t > 1$ e $0 < p_1 q_2 \cdots q_t < n = q_1 q_2 \cdots q_t$. Sottraendo $p_1 q_2 \cdots q_t$ da n otteniamo

$$m = p_1(p_2 \cdots p_s - q_2 \cdots q_t) = (q_1 - p_1) q_2 \cdots q_t < n.$$

Poiché $t > 1$, si ha $m > 1$. Scomponendo in primi positivi $p_2 \cdots p_s - q_2 \cdots q_t$ e $q_1 - p_1$, otteniamo due fattorizzazioni di m. Nella prima compare p_1; nella seconda compaiono $q_2, \ldots, q_t$ e i primi che dividono $q_1 - p_1$. Per l'unicità già ammessa per m, o p_1 coincide con uno dei primi $q_2, \ldots, q_t$, oppure divide $q_1 - p_1$. Quest'ultima possibilità è esclusa, perché comporterebbe $p_1 \mid q_1$ e quindi $p_1 = q_1$. Dunque $p_1 = q_j$ per un $j \geq 2$. Ponendo quel q_j al primo posto si ritorna al caso precedente.[8]

Il teorema fondamentale dell'aritmetica si può enunciare anche nella forma seguente:

Ogni intero diverso da 0 *e da* ± 1 *è un prodotto di primi. A meno dell'ordine e dei segni dei fattori, tale fattorizzazione è unica.*

[8] Una dimostrazione diversa e alcune generalizzazioni di questo risultato verranno date nel capitolo 2.

Il risultato si può esprimere anche nel sistema $\mathbb{Q}$ dei razionali.[9] Ogni razionale diverso da $0, \pm 1$ si può scrivere nella forma $p_1^{\varepsilon_1} p_2^{\varepsilon_2} \cdots p_s^{\varepsilon_s}$, dove i p_i sono interi primi e $\varepsilon_i = \pm 1$. Nella forma ridotta l'espressione è unica a meno dei segni e dell'ordine.

Per $n \in \mathbb{Z}$ possiamo scrivere $n = \pm p_1 p_2 \cdots p_s$ con p_i primi positivi (supponendo sempre $n \neq 0, \pm 1$). Riordinando i fattori e cambiando notazione otteniamo $n = \pm p_1^{k_1} p_2^{k_2} \cdots p_t^{k_t}$, con p_i primi positivi *distinti*. Dal teorema fondamentale segue che ogni divisore m di n ha la forma $\pm p_1^{l_1} p_2^{l_2} \cdots p_t^{l_t}$ con $0 \leq l_i \leq k_i$. Due interi non nulli m, n si possono scrivere servendosi degli stessi primi, purché si consentano esponenti nulli (ricordiamo che $a^0 = 1$ per $a \neq 0$). Dunque possiamo porre

$$m = \pm p_1^{e_1} \cdots p_s^{e_s}, \qquad n = \pm p_1^{f_1} \cdots p_s^{f_s},$$

dove i p_i sono primi positivi distinti ed $e_i, f_i \geq 0$. Poniamo $g_i = \min(e_i, f_i)$, $h_i = \max(e_i, f_i)$ e consideriamo i due interi

$$(m, n) \equiv p_1^{g_1} p_2^{g_2} \cdots p_s^{g_s}, \qquad [m, n] \equiv p_1^{h_1} p_2^{h_2} \cdots p_s^{h_s}. \tag{9}$$

Si vede subito che (m, n) è un *massimo comune divisore* (MCD) di m, n, nel senso che $(m, n) \mid m$, $(m, n) \mid n$ e ogni intero d tale che $d \mid m$ e $d \mid n$ divide (m, n). Analogamente $[m, n]$ è un *minimo comune multiplo* (mcm) di m, n: $m \mid [m, n]$, $n \mid [m, n]$ e, se $m \mid e$ e $n \mid e$, allora $[m, n] \mid e$. Dalla (9) risulta che, se m e n sono positivi,

$$mn = (m, n)[m, n]. \tag{10}$$

Vi è un altro modo per dimostrare l'esistenza del MCD di due interi. Esso non richiede la fattorizzazione in primi e dà l'informazione ulteriore che il MCD si può scrivere nella forma $mu + nv$, con $u, v \in \mathbb{Z}$. Si basa sul risultato seguente.

Algoritmo della divisione in $\mathbb{Z}$. *Se a, b sono interi e $b \neq 0$, esistono interi q, r, con $0 \leq r < |b|$, tali che $a = bq + r$.*

Dimostrazione. Consideriamo l'insieme M dei multipli interi $x|b|$ di $|b|$ tali che $x|b| \leq a$. Questo insieme non è vuoto, poiché $-|a||b| \leq -|a| \leq a$. Quindi M ha un massimo $h|b|$ (esercizio 2 del § 0.5). Si ha $h|b| \leq a$ e dunque $a = h|b| + r$ con $r \geq 0$. D'altra parte $(h+1)|b| = h|b| + |b| > h|b|$, perciò $(h + 1)|b| > a$ e $h|b| + |b| > h|b| + r$. Dunque $r < |b|$. Poniamo $q = h$ se $b > 0$, $q = -h$ se $b < 0$. Allora $h|b| = qb$ e $a = qb + r$, come richiesto. □

Siano ora $m, n \in \mathbb{Z}$, con $n \neq 0$, e sia $I = \{mx + ny \mid x, y \in \mathbb{Z}\}$. L'insieme contiene $|n| > 0$. Esiste dunque un minimo intero positivo $d = mu + nv$ in I. Affermiamo che d è un MCD di m e n. Per l'algoritmo della divisione scriviamo $m = dq + r$ con $0 \leq r < d$. Allora

$$r = m - dq = m - (mu + nv)q = m(1 - uq) - nvq \in I.$$

Poiché d è il minimo intero positivo di I, deve essere $r = 0$. Dunque $d \mid m$; analogamente $d \mid n$. Se, d'altra parte, $e \mid m$ ed $e \mid n$, allora $e \mid mu$ ed $e \mid nv$, quindi $e \mid mu + nv$, cioè $e \mid d$.[10]

Se d' e d sono entrambi MCD di m, n, la seconda condizione della definizione dà $d \mid d'$ e $d' \mid d$. Quindi $d' = \pm d$. Se $n \neq 0$, anche $d \neq 0$ e possiamo scegliere $d > 0$. È questa la determinazione che si ottiene mediante le fattorizzazioni in primi e che indichiamo con (m, n).

[9] Supponiamo nota al lettore la costruzione di $\mathbb{Q}$ a partire da $\mathbb{Z}$. Una situazione più generale, che comprende questa costruzione, sarà trattata nel § 2.9.

[10] Esiste un terzo procedimento, di natura meccanica, per determinare il MCD di due interi, chiamato *algoritmo di Euclide*. È descritto nell'esercizio 11 a p. 150 dell'originale.

Esercizi

1. Dimostrare che, se p è primo e $p \mid ab$, allora $p \mid a$ oppure $p \mid b$.
2. Definire il MCD e il mcm di più di due interi e dimostrarne l'esistenza.
3. Dimostrare che, se k e m sono interi positivi e $m \neq n^k$ per ogni $n \in \mathbb{Z}$, allora $m^{1/k}$ è irrazionale.

0.7 Un cenno ai numeri cardinali

In questo libro avremo spesso occasione di servirci del concetto di numero cardinale di un insieme. Conviene elencare fin d'ora i principali fatti che ci occorreranno. Ometteremo le dimostrazioni: si possono trovare in vari testi, in particolare in *Naive Set Theory* di Halmos.

Cominciamo col dire che due insiemi hanno lo *stesso numero cardinale* o la *stessa cardinalità* (sono *equipotenti* o semplicemente *equivalenti*) se esiste fra essi una corrispondenza biunivoca. Per esempio, gli insiemi $\mathbb{N}$, $\mathbb{Z}$ e l'insieme $\mathbb{Q}$ dei razionali hanno lo stesso numero cardinale. L'insieme $\mathbb{R}$ dei reali ha invece cardinalità maggiore di $\mathbb{Q}$. Come rappresentante della classe degli insiemi di uguale cardinalità si prende un particolare numero ordinale di quella classe e lo si chiama *il numero cardinale* di ciascun insieme della classe. Non daremo qui una definizione dei numeri ordinali, salvo per quelli finiti. Per $n \in \mathbb{N}$ definiamo l'*ordinale* n come il sottoinsieme di $\mathbb{N}$ costituito dai naturali minori di n. Un insieme si dice *finito* se si può mettere in corrispondenza biunivoca con un ordinale finito, ossia con l'insieme dei numeri naturali minori di un numero assegnato. Altrimenti esso è *infinito*.

Indichiamo in generale con $|S|$ il numero cardinale di S e scriviamo $|S| < \infty$ oppure $|S| = \infty$ secondo che S sia finito o infinito. Occorre sapere che, se m e n sono due naturali distinti, non esiste una biiezione fra i corrispondenti ordinali. Supposto $m < n$, ciò si dimostra facilmente per induzione su n. Un altro modo di esprimere questo fatto è il seguente: se S, T sono insiemi finiti con $|S| > |T|$ (in particolare, se T è un sottoinsieme proprio di S), per ogni applicazione suriettiva α di S su T esistono $s_1, s_2 \in S$ distinti tali che $\alpha(s_1) = \alpha(s_2)$. Questo fatto semplice e familiare è il *principio dei cassetti*: se le lettere sono più numerose degli scomparti del casellario, almeno uno scomparto deve contenere più di una lettera. Il principio ha molte applicazioni importanti in matematica ed è caratteristico degli insiemi finiti. Per ogni insieme infinito esistono infatti biiezioni su suoi sottoinsiemi propri. Se S e T sono finiti, allora $|S \times T| = |S|\,|T|$ e $|S^T| = |S|^{|T|}$, dove S^T indica l'insieme delle applicazioni di T in S.

Un risultato importante sui numeri cardinali degli insiemi infiniti è il *teorema di Schröder–Bernstein*: se esistono applicazioni iniettive di S in T e di T in S, allora $|S| = |T|$.

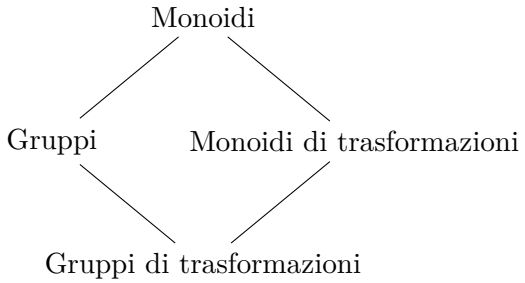
Capitolo 1
Monoidi e gruppi

La teoria dei gruppi è uno dei rami più antichi e più ricchi dell'algebra. I gruppi di trasformazioni hanno un ruolo importante in geometria e, come vedremo nel capitolo 4, i gruppi finiti sono alla base delle scoperte di Galois nella teoria delle equazioni. Da questi due campi venne l'impulso iniziale allo sviluppo della teoria dei gruppi, il cui studio sistematico risale ai primi decenni dell'Ottocento.

Più generale del concetto di gruppo è quello di monoide. Si tratta semplicemente di un insieme dotato di una composizione binaria associativa e di un elemento neutro; i gruppi sono i monoidi in cui ogni elemento ha un inverso rispetto all'elemento neutro. Sebbene la teoria dei monoidi sia assai meno ricca di quella dei gruppi, in tempi recenti se ne sono trovate importanti applicazioni «esterne», particolarmente nella teoria degli automi. Cominceremo con la nozione più semplice e generale di monoide, anche se il nostro obiettivo principale è la teoria dei gruppi. Speriamo che lo studio preliminare dei monoidi chiarisca, collocandoli in una prospettiva più ampia, alcuni risultati sui gruppi. I risultati sui monoidi ci serviranno inoltre nello studio degli anelli, che si possono considerare come coppie di monoidi sullo stesso insieme soggiacente, soggette ad alcune condizioni ulteriori, come le leggi distributive.

Gran parte del capitolo è dedicata ai fondamenti. Il lettore incontrerà molti concetti nuovi e potrà aver bisogno di tempo per assimilarli tutti. A chi vi si accosta per la prima volta il punto di vista potrà sembrare alquanto astratto. Abbiamo cercato di superare questa difficoltà proponendo molti esempi ed esercizi che diano concretezza alla teoria. Il metodo assiomatico che impiegheremo in tutto il libro, e in particolare in questo capitolo, è probabilmente familiare al lettore: si pensi allo sviluppo assiomatico della geometria euclidea e del sistema dei numeri reali. Vi è tuttavia una differenza notevole. In quelle teorie i sistemi di assiomi sono categorici, nel senso che esiste essenzialmente un unico sistema che li soddisfa. Nelle situazioni che considereremo le cose stanno assai diversamente. Le nostre assiomatizzazioni intendono applicarsi simultaneamente a un gran numero di modelli e quasi mai conosciamo tutta la loro portata. In generale gioverà comunque tener presenti alcuni esempi.

I principali sistemi che considereremo sono i monoidi, i monoidi di trasformazioni, i gruppi e i gruppi di trasformazioni. Le relazioni fra questi quattro concetti sono indicate dal diagramma



Esso indica che la classe dei gruppi e quella dei monoidi di trasformazioni sono contenute nella classe dei monoidi e che la loro intersezione è la classe dei gruppi di trasformazioni. A queste nozioni si aggiunge il concetto fondamentale di omomorfismo, che individua il tipo di applicazioni naturale per i sistemi in esame. Introdurremo per prima la nozione più intuitiva di isomorfismo.

Alla fine del capitolo, con la dimostrazione dei teoremi di Sylow per i gruppi finiti, andremo oltre i fondamenti. Altri risultati sui gruppi finiti saranno dati nel capitolo 4, quando ne avremo bisogno per la teoria delle equazioni. Più avanti, nel capitolo 6, studieremo la struttura di alcuni gruppi geometrici classici, per esempio i gruppi delle rotazioni.

1.1 Monoidi di trasformazioni e monoidi astratti

Nel § 0.2 abbiamo visto che la composizione delle applicazioni fra insiemi soddisfa la legge associativa. Se $S \xrightarrow{\alpha} T$, $T \xrightarrow{\beta} U$ e $U \xrightarrow{\gamma} V$, e se $\beta\alpha : S \to U$ è definita da $(\beta\alpha)(s) = \beta(\alpha(s))$, allora $\gamma(\beta\alpha) = (\gamma\beta)\alpha$. Ricordiamo anche che se 1_T è l'applicazione identica $t \mapsto t$ di T in sé, allora $1_T\alpha = \alpha$ e $\beta 1_T = \beta$ per ogni $\alpha : S \to T$ e $\beta : T \to U$. Consideriamo ora il caso particolare dell'insieme $M(S)$ delle trasformazioni, ossia delle applicazioni, di S in sé. Per esempio, sia $S = \{1, 2\}$. In questo caso $M(S)$ è costituito dalle quattro trasformazioni

$$1_S = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \quad \alpha = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \quad \gamma = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix};$$

sotto ciascun elemento della prima riga ne abbiamo indicato l'immagine. Si verifica facilmente che i prodotti in $M(S)$ sono dati dalla tavola

(1)

	1	α	β	γ
1	1	α	β	γ
α	α	1	γ	β
β	β	β	β	β
γ	γ	γ	γ	γ

All'intersezione della riga ρ con la colonna σ figura il prodotto $\rho\sigma$, con $\rho, \sigma = 1, \alpha, \beta, \gamma$. Più in generale, se $S = \{1, 2, \ldots, n\}$, allora $M(S)$ consta di n^n trasformazioni; fissato n, si può scrivere per $M(S)$ una tavola di moltiplicazione come la (1). Per ogni S non vuoto, $M(S)$ è un esempio di monoide, cioè di insieme non vuoto dotato di una composizione binaria associativa e di un elemento neutro: un elemento 1 il cui prodotto, in entrambi gli ordini, con un elemento qualsiasi restituisce quest'ultimo. Più formalmente diamo la seguente definizione.

Definizione 1.1. *Un monoide è una terna $(M, p, 1)$ dove M è un insieme non vuoto, p una composizione binaria associativa (o prodotto) in M e 1 un elemento di M tale che $p(1, a) = a = p(a, 1)$ per ogni $a \in M$.*

Se si sopprime l'ipotesi di associatività di p, si ottiene un sistema talvolta chiamato *monade*. Se invece si sopprime l'ipotesi su 1, rimane un insieme dotato di una composizione binaria associativa: un *semigruppo* (M, p). Abbrevieremo $p(a, b)$, prodotto di a e b rispetto a p, nella forma consueta ab (o $a \cdot b$). Un elemento 1 di (M, p) tale che $a1 = a = 1a$ per ogni $a \in M$ è detto *elemento neutro* di (M, p). Se $1'$ è un altro elemento con questa proprietà, allora $1'1 = 1$ e $1'1 = 1'$, dunque $1' = 1$. L'elemento neutro, quando esiste, è perciò unico, sicché possiamo parlare dell'elemento neutro di (M, p). È chiaro che un monoide si può anche definire come un semigruppo dotato di elemento neutro. Preferiamo tuttavia attenerci alla prima definizione. Una volta introdotto il monoide $(M, p, 1)$, se non vi sono ambiguità parleremo più brevemente del «monoide M», benché a rigore M sia l'insieme soggiacente e costituisca soltanto uno degli ingredienti di $(M, p, 1)$.

Nella matematica già familiare al lettore abbondano gli esempi di monoidi. Ne elenchiamo alcuni.

Esempi

1. $(\mathbb{N}, +, 0)$: $\mathbb{N}$ è l'insieme dei numeri naturali, $+$ la loro addizione usuale e 0 il primo elemento di $\mathbb{N}$.

2. $(\mathbb{N}, \cdot, 1)$, dove $\cdot$ è il prodotto usuale e 1 è il naturale uno.

3. $(P, \cdot, 1)$, con P insieme degli interi positivi, $\cdot$ e 1 come nell'esempio 2.

4. $(\mathbb{Z}, +, 0)$, con $\mathbb{Z}$ insieme degli interi, $+$ e 0 nel significato usuale.

5. $(\mathbb{Z}, \cdot, 1)$, con $\cdot$ e 1 usuali.

6. Sia S un insieme non vuoto e sia $\mathcal{P}(S)$ l'insieme dei suoi sottoinsiemi. Si hanno i due monoidi $(\mathcal{P}(S), \cup, \varnothing)$ e $(\mathcal{P}(S), \cap, S)$.

7. Sia σ una trasformazione di S. Definiamo per induzione $\sigma^0 = 1$, $\sigma^r = \sigma^{r-1}\sigma$ per $r > 0$. Allora $\sigma^r\sigma^s = \sigma^{r+s}$, come si vede facilmente e come dimostreremo nel § 1.4. L'insieme $\langle\sigma\rangle = \{\sigma^k \mid k \in \mathbb{N}\}$, con l'usuale composizione delle trasformazioni e con $\sigma^0 = 1$, costituisce un monoide.

Se M è un monoide, un suo sottoinsieme N si dice *sottomonoide* di M quando contiene 1 ed è chiuso rispetto al prodotto in M, ossia $n_1 n_2 \in N$ per ogni $n_1, n_2 \in N$. Per esempio, il monoide 2, $(\mathbb{N}, \cdot, 1)$, è un sottomonoide di $(\mathbb{Z}, \cdot, 1)$; il monoide 3, $(P, \cdot, 1)$, è un sottomonoide di $(\mathbb{N}, \cdot, 1)$. Il sottoinsieme $\{0\}$ di $\mathbb{N}$ è chiuso rispetto alla moltiplicazione, ma non è un sottomonoide del monoide 2, poiché non contiene 1. Se N è un sottomonoide di M, esso costituisce un monoide con il prodotto di M ristretto a N e con lo stesso elemento neutro. È chiaro che un sottomonoide di un sottomonoide di M è un sottomonoide di M. Chiameremo *monoide di trasformazioni* (di S) un sottomonoide del monoide $M(S)$ di tutte le trasformazioni di S. Per definizione, un sottoinsieme N di $M(S)$ è un monoide di trasformazioni se e solo se contiene l'identità e, insieme a due applicazioni qualsiasi, contiene anche la loro composta.

Un monoide si dice *finito* quando ha un numero finito di elementi. Chiameremo ordinariamente *ordine* di un monoide la sua cardinalità, denotata con $|M|$. Per lo studio di un monoide finito è utile disporre di una tavola di moltiplicazione. Come nel caso particolare già considerato, se $M = \{a_1 = 1, a_2, \ldots, a_m\}$ essa ha la forma

(2)

	1	a_2	$\cdots$	a_j	$\cdots$	a_m
1						
$\vdots$				$\vdots$		
a_i		$\cdots$		$a_i a_j$	$\cdots$	
$\vdots$				$\vdots$		
a_m						

dove $a_i a_j$ occupa l'intersezione della riga intestata ad a_i con la colonna intestata ad a_j.

Esercizi

1. Sia S un insieme. Si definisca in S il prodotto $ab = b$. Dimostrare che S è un semigruppo. Sotto quale condizione S contiene un elemento neutro?

2. Sia $M = \mathbb{Z} \times \mathbb{Z}$, l'insieme delle coppie di interi (x_1, x_2). Si ponga $(x_1, x_2)(y_1, y_2) = (x_1y_1 + 2x_2y_2, x_1y_2 + x_2y_1)$ e $1 = (1, 0)$. Dimostrare che si ottiene un monoide. (Osservare che la moltiplicazione è commutativa.) Dimostrare che, se $(x_1, x_2) \neq (0, 0)$, vale la legge di cancellazione per (x_1, x_2), cioè

$$(x_1, x_2)(y_1, y_2) = (x_1, x_2)(z_1, z_2) \implies (y_1, y_2) = (z_1, z_2).$$

3. Una macchina accetta parole di otto lettere (una qualsiasi sequenza di otto lettere dell'alfabeto, anche priva di significato) e stampa una parola di otto lettere composta dalle prime cinque della prima parola e dalle ultime tre della seconda. Dimostrare che, rispetto a questa composizione, le parole di otto lettere formano un semigruppo. Che cosa accade se la macchina stampa le ultime quattro lettere della prima parola seguite dalle prime quattro della seconda? Uno dei due sistemi è un monoide?

4. Sia $(M, p, 1)$ un monoide e sia $m \in M$. Definire in M un nuovo prodotto p_m ponendo $p_m(a, b) = amb$. Dimostrare che si ottiene un semigruppo. Per quali m esiste un elemento neutro rispetto a p_m?

5. Sia S un semigruppo e sia u un elemento che non appartiene a S. Formare $M = S \cup \{u\}$ ed estendere il prodotto di S a un prodotto binario su M ponendo $ua = a = au$ per ogni $a \in M$. Dimostrare che M è un monoide.

1.2 Gruppi di trasformazioni e gruppi astratti

Un elemento u di un monoide M si dice *invertibile* (o *unità*[1]) quando esiste $v \in M$ tale che

$$uv = 1 = vu. \tag{3}$$

Se anche v' soddisfa $uv' = 1 = v'u$, allora $v' = (vu)v' = v(uv') = v$. L'elemento v che soddisfa (3) è dunque unico: lo chiamiamo *inverso* di u e scriviamo $v = u^{-1}$. È anche chiaro che u^{-1} è invertibile e $(u^{-1})^{-1} = u$. Possiamo ora dare la seguente definizione.

Definizione 1.2. *Un gruppo G (o $(G, p, 1)$) è un monoide i cui elementi sono tutti invertibili.*

Un sottomonoide di un monoide M (in particolare, di un gruppo) si chiama *sottogruppo* se, considerato come monoide, è un gruppo. Poiché l'elemento neutro di un sottomonoide coincide con quello di M, un sottoinsieme G di M è un sottogruppo se e solo se soddisfa le seguenti condizioni di chiusura: $1 \in G$, $g_1 g_2 \in G$ per ogni $g_1, g_2 \in G$, e ogni $g \in G$ è invertibile con $g^{-1} \in G$.

Indichiamo con $U(M)$ l'insieme degli elementi invertibili del monoide M e siano $u_1, u_2 \in U(M)$. Allora

$$(u_1 u_2)(u_2^{-1} u_1^{-1}) = u_1(u_2 u_2^{-1})u_1^{-1} = 1;$$

analogamente $(u_2^{-1} u_1^{-1})(u_1 u_2) = 1$. Quindi $u_1 u_2 \in U(M)$. Abbiamo visto che $u \in U(M)$ implica $u^{-1} \in U(M)$ e da $1 \cdot 1 = 1$ segue naturalmente $1 \in U(M)$. Dunque $U(M)$ è un sottogruppo di M. Lo chiameremo *gruppo delle unità*, o degli elementi invertibili, di M. Per esempio, se $M = (\mathbb{Z}, \cdot, 1)$, allora $U(M) = \{1, -1\}$, mentre per $M = (\mathbb{N}, \cdot, 1)$ si ha $U(M) = \{1\}$.

Consideriamo il monoide $M(S)$ delle trasformazioni di un insieme non vuoto S. Qual è il corrispondente gruppo delle unità $U(M(S))$? Abbiamo visto (p. 8 dell'originale) che una trasformazione è invertibile se e solo se è biiettiva. Il nostro gruppo è dunque l'insieme delle trasformazioni biiettive di S, con la composizione delle applicazioni come prodotto e l'identità come elemento neutro. Chiameremo $U(M(S))$ *gruppo simmetrico* dell'insieme S e lo indicheremo con $\operatorname{Sym} S$. Quando $S = \{1, 2, \ldots, n\}$, scriveremo S_n per $\operatorname{Sym} S$ e parleremo del *gruppo simmetrico su n lettere*. Gli elementi di S_n si chiamano ordinariamente *permutazioni* di $\{1, 2, \ldots, n\}$. È facile elencarli tutti e determinare l'ordine di S_n. Nella notazione introdotta per $n = 2$, una trasformazione di $\{1, 2, \ldots, n\}$ si rappresenta con il simbolo

$$\alpha = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1' & 2' & \cdots & n' \end{pmatrix}, \tag{4}$$

[1] Questa espressione è abbastanza comune. Purtroppo il termine «unità» viene usato anche per l'elemento neutro 1. Il contesto chiarirà di solito il significato.

che indica la trasformazione $i \mapsto i'$ per $1 \leq i \leq n$. Affinché α sia iniettiva, la seconda riga $1', \ldots, n'$ non deve contenere ripetizioni: nessun i può comparirvi due volte. Ciò garantisce anche la biiettività, poiché non si può applicare iniettivamente $\{1, 2, \ldots, n\}$ in un suo sottoinsieme proprio. Per contare gli elementi di S_n, osserviamo che per $1'$ nel simbolo (4) possiamo prendere uno qualsiasi degli n numeri $1, 2, \ldots, n$. Abbiamo così n scelte per $1'$. Effettuata la prima scelta, per evitare ripetizioni si deve prendere $2'$ fra gli $n-1$ numeri diversi da $1'$: vi sono quindi $n-1$ scelte. Scelte le immagini di 1 e 2, per $3'$ rimangono $n-2$ scelte, e così via. Si hanno dunque $n!$ simboli (4) che rappresentano gli elementi di S_n. Abbiamo così dimostrato il seguente risultato.

Teorema 1.1. *L'ordine di S_n è $n!$.*

Si confronti tale ordine con l'ordine n^n del monoide delle trasformazioni di $S = \{1, 2, \ldots, n\}$.

Abbiamo chiamato monoide di trasformazioni un sottomonoide del monoide delle trasformazioni di un insieme. Analogamente, un sottogruppo del gruppo simmetrico di S si chiama *gruppo di trasformazioni*. Se S è finito, per un gruppo di trasformazioni di S si adopera generalmente il termine *gruppo di permutazioni*. Un insieme G di trasformazioni di S è un gruppo di trasformazioni se e solo se è costituito da biiezioni e soddisfa le condizioni di chiusura $1 = 1_S \in G$, $\alpha\beta \in G$ per $\alpha, \beta \in G$ e $\alpha^{-1} \in G$ per $\alpha \in G$.

Esempi

1. $(\mathbb{Z}, +, 0)$, il gruppo additivo degli interi.[2] L'inverso di a è $-a$.
2. $(\mathbb{Q}, +, 0)$, dove $\mathbb{Q}$ è l'insieme dei numeri razionali; la composizione è l'addizione e l'inverso di a è $-a$.
3. $(\mathbb{R}, +, 0)$: i numeri reali con l'addizione e lo zero usuali.
4. $(\mathbb{C}, +, 0)$: i numeri complessi con l'addizione e lo zero usuali.
5. $(\mathbb{Q}^*, \cdot, 1)$, dove $\mathbb{Q}^*$ è l'insieme dei razionali non nulli, la composizione è la moltiplicazione, 1 è l'uno abituale e a^{-1} è il reciproco.
6. $(\mathbb{R}^*, \cdot, 1)$: i reali non nulli con moltiplicazione, unità e inversi usuali.
7. $(\mathbb{C}^*, \cdot, 1)$: i complessi non nulli con moltiplicazione, unità e inversi usuali.
8. $(\mathbb{R}^{(3)}, +, 0)$: le terne di reali (x, y, z), con la somma

$$(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2)$$

   e $0 = (0, 0, 0)$. L'inverso di (x, y, z) è $(-x, -y, -z)$. Si può anche descrivere questo esempio come il gruppo dei vettori dello spazio euclideo tridimensionale, con l'usuale costruzione geometrica della somma.

9. L'insieme delle rotazioni del piano intorno a un punto O, con la composizione usuale. Preso O come origine, la rotazione di un angolo θ è rappresentata analiticamente dall'applicazione $(x, y) \mapsto (x', y')$ con

$$x' = x\cos\theta - y\sin\theta, \qquad y' = x\sin\theta + y\cos\theta.$$

   Per $\theta = 0$ si ottiene l'identità. L'inversa della rotazione di angolo θ è quella di angolo $-\theta$.

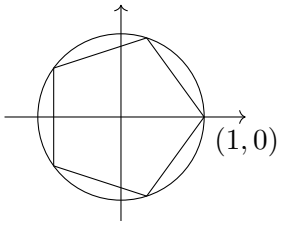
[2] In tutto il libro usiamo le notazioni ormai consuete $\mathbb{N}$ per i naturali $0, 1, 2, \ldots$, $\mathbb{Z}$ per gli interi, $\mathbb{Q}$ per i razionali, $\mathbb{R}$ per i reali e $\mathbb{C}$ per i complessi.

10. L'insieme delle rotazioni e delle riflessioni rispetto alle rette passanti per O. Le riflessioni si esprimono analiticamente mediante $(x, y) \mapsto (x', y')$, dove

$$x' = x\cos\theta + y\sin\theta, \qquad y' = x\sin\theta - y\cos\theta.$$

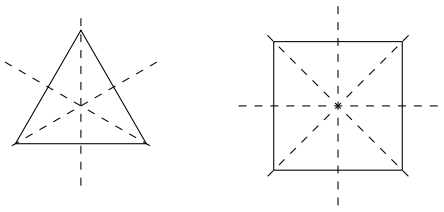
Il prodotto di due riflessioni è una rotazione; il prodotto, in qualsiasi ordine, di una riflessione e di una rotazione è una riflessione.

11. Consideriamo un n-agono regolare (cioè un poligono regolare di n lati), inscritto nella circonferenza unitaria del piano, con un vertice in $(1, 0)$, per esempio il pentagono regolare della figura.



Le semirette uscenti dall'origine e passanti per i vertici formano con il semiasse positivo delle x angoli di $0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$ radianti. Le rotazioni che mandano la figura in sé sono le n rotazioni di questi angoli; formano un sottogruppo R_n del gruppo delle rotazioni.

12. Consideriamo l'insieme D_n delle rotazioni e riflessioni che mandano in sé l'n-agono regolare dell'esempio 11. Esse formano un sottogruppo del gruppo dell'esempio 10. Chiameremo gli elementi di questo gruppo *simmetrie* dell'n-agono regolare. La riflessione nell'asse x è una simmetria. Componendola con le n simmetrie rotatorie si ottengono n simmetrie di riflessione distinte. Le abbiamo così ottenute tutte: detta S la riflessione nell'asse x e T una qualunque simmetria di riflessione, ST è una delle n simmetrie rotatorie $R_1, \ldots, R_n$, per esempio R_j. Da $S^2 = 1$ e $ST = R_j$ segue $T = SR_j$, una delle riflessioni che abbiamo contato. Dunque D_n consiste di n rotazioni e n riflessioni e ha ordine $2n$. È detto *gruppo diedrale*. Per $n = 3$ e $n = 4$ gli assi delle riflessioni che sono simmetrie del poligono sono tratteggiati nelle figure.



13. Indichiamo con U_n l'insieme delle radici n-esime dell'unità nei complessi, cioè i numeri z tali che $z^n = 1$. È facile determinarli mediante la rappresentazione polare $z = re^{i\theta} = r(\cos\theta + i\sin\theta)$, dove $r = |z|$ e θ è l'argomento (l'angolo) di z. Se $z_1 = r_1 e^{i\theta_1}$ e $z_2 = r_2 e^{i\theta_2}$, allora $z_1 z_2 = r_1 r_2 e^{i(\theta_1+\theta_2)}$. Pertanto se $z^n = 1$, allora $r = |z| = 1$ e si può scegliere θ fra $0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$. Poiché $1^n = 1$, e da $z_1^n = z_2^n = 1$ segue $(z_1 z_2)^n = z_1^n z_2^n = 1$ e $(z_1^{-1})^n = (z_1^n)^{-1} = 1$, vediamo che U_n è un sottogruppo del gruppo moltiplicativo $\mathbb{C}^*$ dell'esempio 7.

14. Il gruppo delle rotazioni dello spazio euclideo tridimensionale. È l'insieme delle rotazioni intorno all'origine O nello spazio numerico $\mathbb{R}^{(3)}$ delle terne (x, y, z) di reali. Dalla geometria analitica sappiamo che queste applicazioni hanno forma $(x, y, z) \mapsto (x', y', z')$, dove

$$x' = \lambda_1 x + \mu_1 y + \nu_1 z, \qquad y' = \lambda_2 x + \mu_2 y + \nu_2 z, \qquad z' = \lambda_3 x + \mu_3 y + \nu_3 z$$

e i coefficienti λ_i, μ_i, ν_i sono reali qualsiasi tali che

$$\lambda_i^2 + \mu_i^2 + \nu_i^2 = 1, \quad \lambda_i\lambda_j + \mu_i\mu_j + \nu_i\nu_j = 0 \quad (i \neq j), \qquad \begin{vmatrix} \lambda_1 & \mu_1 & \nu_1 \\ \lambda_2 & \mu_2 & \nu_2 \\ \lambda_3 & \mu_3 & \nu_3 \end{vmatrix} = 1.$$

Gli esempi dal 9 al 14, eccettuato il 13, sono gruppi di trasformazioni. Osserviamo anche che, nella lista dei monoidi a p. 29 dell'originale, gli esempi 1, 2, 3 e 5 non sono gruppi e il 7 può esserlo oppure no. I due esempi geometrici 11 e 12 illustrano un principio generale. Se G è un gruppo di trasformazioni di un insieme S e A è un sottoinsieme di S, le trasformazioni di G che mandano A su sé stesso ($\eta(A) = A$) formano un sottogruppo G_A di G. La verifica è immediata.

Consideriamo ora una costruzione generale, detta *prodotto diretto*, che a partire da monoidi o gruppi assegnati ne produce di nuovi. Siano $M_1, M_2, \ldots, M_n$ monoidi e poniamo $M = M_1 \times M_2 \times \cdots \times M_n$. Introduciamo in M il prodotto

$$(a_1, a_2, \ldots, a_n)(b_1, b_2, \ldots, b_n) = (a_1b_1, a_2b_2, \ldots, a_nb_n),$$

dove $a_i, b_i \in M_i$, e poniamo $1 = (1_1, 1_2, \ldots, 1_n)$, con 1_i elemento neutro di M_i. Scrivendo (a_i) per $(a_1, \ldots, a_n)$ e così via, abbiamo $((a_i)(b_i))(c_i) = ((a_ib_i)c_i)$ e $(a_i)((b_i)(c_i)) = (a_i(b_ic_i))$. Vale perciò la legge associativa. Inoltre $1(a_i) = (a_i) = (a_i)1$, quindi 1 è l'elemento neutro. Abbiamo così un monoide, il *prodotto diretto* $M_1 \times \cdots \times M_n$ dei monoidi M_i. Se ciascun M_i è un gruppo G_i, anche $G_1 \times \cdots \times G_n$ è un gruppo, poiché (a_i) ha per inverso (a_i^{-1}). Si parla allora di *prodotto diretto dei gruppi* G_i. L'esempio 8 è un caso particolare: si può intendere come il prodotto diretto di tre copie di $(\mathbb{R}, +, 0)$. Come mostra l'esempio, non occorre che i monoidi M_i (o i gruppi G_i) siano distinti. È anzi di particolare interesse prendere tutti gli M_i uguali a un monoide fissato N. Otteniamo il prodotto diretto di N con sé stesso n volte, o *n-esima potenza diretta* di N. La indicheremo solitamente con $N^{(n)}$.

Esercizi

1. Determinare $\alpha\beta$, $\beta\alpha$ e α^{-1} in S_5 per

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix}, \qquad \beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 2 \end{pmatrix}.$$

2. Verificare che le permutazioni $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$, $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ e $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$ formano un sottogruppo di S_3.

3. Determinare una tavola di moltiplicazione per S_3.

4. Sia G l'insieme delle coppie di reali (a, b) con $a \neq 0$. Ponendo $(a, b)(c, d) = (ac, ad + b)$ e $1 = (1, 0)$, verificare che si ottiene un gruppo.

5. Sia G l'insieme delle trasformazioni della retta reale $\mathbb{R}$ definite da $x \mapsto x' = ax + b$, con a, b reali e $a \neq 0$. Verificare che G è un gruppo di trasformazioni di $\mathbb{R}$.

6. Verificare che le traslazioni $x \mapsto x' = x + b$ costituiscono un sottogruppo del gruppo dell'esercizio 5.

7. Dimostrare che se un elemento a di un monoide ha un inverso destro b, ossia $ab = 1$, e un inverso sinistro c, ossia $ca = 1$, allora $b = c$ e a è invertibile con $a^{-1} = b$. Dimostrare che a è invertibile di inverso b se e solo se $aba = a$ e $ab^2a = 1$.

8. Sia α una rotazione del piano intorno all'origine e ρ la riflessione rispetto all'asse x. Dimostrare che $\rho\alpha\rho^{-1} = \alpha^{-1}$.

9. Sia G un sottoinsieme non vuoto di un monoide M. Dimostrare che G è un sottogruppo se e solo se ogni $g \in G$ è invertibile in M e $g_1^{-1}g_2 \in G$ per ogni $g_1, g_2 \in G$.

10. Sia G un semigruppo con le seguenti proprietà: (a) G contiene un elemento neutro destro 1_r, ossia $a1_r = a$ per ogni $a \in G$; (b) ogni elemento $a \in G$ ha un inverso destro rispetto a 1_r ($ab = 1_r$). Dimostrare che G è un gruppo.[3]

11. Dimostrare che in un gruppo le equazioni $ax = b$ e $ya = b$ sono risolubili per ogni $a, b \in G$. Viceversa, dimostrare che ogni semigruppo con questa proprietà contiene un elemento neutro ed è un gruppo.

12. Dimostrare che in un gruppo valgono entrambe le leggi di cancellazione: $ax = ay \Rightarrow x = y$ e $xa = ya \Rightarrow x = y$. Dimostrare che un semigruppo finito in cui valgano entrambe è un gruppo. (Suggerimento: usare il principio dei cassetti e l'esercizio 11.)

13. Dimostrare che un gruppo finito di ordine pari contiene un elemento $a \neq 1$ tale che $a^2 = 1$.

14. Dimostrare che un gruppo G non può essere unione di due sottogruppi propri.

15. Sia G un insieme finito dotato di una composizione binaria con elemento neutro. Dimostrare che G è un gruppo se e solo se la tavola di moltiplicazione (costruita come per i monoidi) ha le proprietà seguenti: (i) ogni riga e ogni colonna contengono tutti gli elementi di G; (ii) per ogni coppia $x \neq 1$, $y \neq 1$ di elementi di G, si prenda nel corpo della tavola un rettangolo R che abbia 1 in un vertice, x nel vertice sulla stessa riga di 1 e y nel vertice sulla stessa colonna di 1. Il quarto vertice del rettangolo dipende solo da (x, y), indipendentemente dalla posizione di 1.

1.3 Isomorfismo. Il teorema di Cayley

A questo punto il lettore potrebbe sentirsi un po' sopraffatto dalla moltitudine di esempi di monoidi e gruppi. Sarà dunque rassicurante sapere che certi gruppi, benché in apparenza differenti, possono considerarsi essenzialmente uguali: sono «isomorfi» nel senso che ora preciseremo. Vedremo anche che ogni monoide è isomorfo a un monoide di trasformazioni e ogni gruppo a un gruppo di trasformazioni. Nella classe dei monoidi (gruppi) di trasformazioni ritroviamo così, a meno di isomorfismo, tutti i monoidi (gruppi). Per i gruppi il risultato è dovuto a Cayley. Diamo anzitutto la definizione.

Definizione 1.3. *Due monoidi $(M, p, 1)$ e $(M', p', 1')$ si dicono isomorfi se esiste un'applicazione biiettiva $\eta : M \to M'$ tale che*

$$\eta(1) = 1', \qquad \eta(xy) = \eta(x)\eta(y) \qquad (x, y \in M). \tag{5}$$

[3] I semigruppi che soddisfano (a) e la condizione (b′) ottenuta da (b) sostituendo «inverso destro» con «inverso sinistro» non sono necessariamente gruppi. La loro struttura è stata determinata da A. H. Clifford, *Annals of Mathematics*, vol. 34 (1933), pp. 865–871.

Scriveremo $M \cong M'$ per indicare che M è isomorfo a M'. Una tale applicazione η si chiama *isomorfismo* di M su M'. In realtà, la prima condizione in (5) è superflua. Infatti, se η soddisfa la seconda, si ha $\eta(x)\eta(1) = \eta(x) = \eta(1)\eta(x)$. Poiché η è suriettiva, $\eta(1)$ agisce come l'elemento neutro $1'$ in M'; dall'unicità dell'elemento neutro segue $\eta(1) = 1'$. Preferiamo tuttavia includere la prima condizione nella definizione, poiché ne avremo bisogno in un contesto più generale che considereremo più avanti.

Forse il primo esempio significativo di isomorfismo fra gruppi che sia stato scoperto è quello fra il gruppo additivo dei numeri reali e il gruppo moltiplicativo dei reali positivi. Indichiamoli rispettivamente con $(\mathbb{R}, +, 0)$ e $(\mathbb{R}^+, \cdot, 1)$. Un isomorfismo del primo sul secondo è dato dall'esponenziale $x \mapsto e^x$. È una biiezione con inversa $y \mapsto \log y$ (il logaritmo naturale), e vale l'«equazione funzionale»

$$e^{x+y} = e^x e^y,$$

che esprime appunto la seconda condizione in (5), poiché la composizione in $(\mathbb{R}, +, 0)$ è $+$.

Se M e M' sono isomorfi vi possono essere molti isomorfismi fra di essi. Per esempio, per ogni reale positivo $a \neq 1$, $x \mapsto a^x$ è un isomorfismo fra i gruppi appena considerati. L'isomorfismo è una relazione di equivalenza. Ogni monoide è isomorfo a sé stesso mediante l'identità; se $\eta : M \to M'$ è un isomorfismo, applicando η^{-1} alla seconda condizione in (5) si ottiene $xy = \eta^{-1}(\eta(x)\eta(y))$. Scrivendo $\eta(x) = x'$, $\eta(y) = y'$, risulta $\eta^{-1}(x')\eta^{-1}(y') = \eta^{-1}(x'y')$, per tutti gli $x', y' \in M'$ in virtù della suriettività di η. Dunque η^{-1} è un isomorfismo da M' a M. Infine, se $\zeta : M' \to M''$ è un isomorfismo, allora $(\zeta\eta)(xy) = \zeta(\eta(xy)) = \zeta(\eta(x)\eta(y)) = \zeta(\eta(x))\zeta(\eta(y))$. Dunque $\zeta\eta : M \to M''$ è un isomorfismo. Dimostriamo ora il risultato annunciato.

Teorema di Cayley per i monoidi e i gruppi. *(1) Ogni monoide è isomorfo a un monoide di trasformazioni. (2) Ogni gruppo è isomorfo a un gruppo di trasformazioni.*

Dimostrazione. (1) Dato un monoide $(M, p, 1)$, costruiremo un isomorfismo con un monoide di trasformazioni dello stesso insieme M. Per $a \in M$ definiamo $a_L : M \to M$ mediante $x \mapsto ax$. Chiamiamo a_L la *traslazione sinistra* (o *moltiplicazione a sinistra*) definita da a. Affermiamo anzitutto che $M_L = \{a_L \mid a \in M\}$ è un monoide di trasformazioni; come abbiamo visto, ciò significa che contiene l'identità ed è chiuso per composizione. Poiché 1_L è $x \mapsto 1x = x$, si ha $1_L = 1\,(= 1_M) \in M_L$. Inoltre $a_L b_L$ manda x in $a(bx)$. Per la legge associativa $a(bx) = (ab)x = (ab)_L(x)$, dunque $a_L b_L = (ab)_L \in M_L$. Ora l'applicazione $a \mapsto a_L$ è un isomorfismo di $(M, p, 1)$ sul monoide di trasformazioni M_L. Le uguaglianze $1_L = 1$ e $a_L b_L = (ab)_L$ sono le condizioni (5), e l'applicazione è evidentemente suriettiva. È anche iniettiva: se $a_L = b_L$, si ha in particolare $a = a_L(1) = b_L(1) = b$. Pertanto $a \mapsto a_L$ è un isomorfismo.

(2) Sia ora $(G, p, 1)$ un gruppo. Tutto seguirà dalla dimostrazione di (1) una volta provato che G_L è un gruppo di trasformazioni. Occorrono due ulteriori fatti: le applicazioni a_L sono biiettive e G_L è chiuso rispetto al passaggio agli inversi. Entrambi seguono da $1_L = (a^{-1}a)_L = (a^{-1})_L a_L$ e da $1_L = a_L(a^{-1})_L$. Dunque a_L ammette inversa $(a^{-1})_L$, che appartiene a G_L. □

Se M (o G) è finito, M_L opera sull'insieme finito M. In particolare, se $|G| = n$, G_L è un sottogruppo di S_n, il gruppo simmetrico su un insieme di n elementi. Otteniamo così il seguente corollario.

Corollario. *Ogni gruppo finito di ordine n è isomorfo a un sottogruppo del gruppo simmetrico S_n.*

Esempi

1. Sia $(\mathbb{R}, +, 0)$ il gruppo additivo dei reali. Per $a \in \mathbb{R}$ la traslazione sinistra a_L è $x \mapsto a + x$.

2. Sia G il gruppo delle coppie di reali (a, b) con $a \neq 0$, prodotto $(a, b)(c, d) = (ac, ad + b)$ ed elemento neutro $(1, 0)$ (esercizio 4 del § 1.2). Qui $(a, b)_L$ è l'applicazione $(x, y) \mapsto (ax, ay + b)$. Un altro gruppo di trasformazioni isomorfo a G è quello delle applicazioni di $\mathbb{R}$ in sé $x \mapsto ax + b$ con $a \neq 0$. L'applicazione che manda (a, b) nella trasformazione $T_{(a,b)} : x \mapsto ax + b$ è un isomorfismo.

Esercizi

1. Usare una tavola di moltiplicazione di S_3 (esercizio 3 del § 1.2) e l'isomorfismo $a \mapsto a_L$, con a_L traslazione sinistra definita da a, per ottenere un sottogruppo di S_6 isomorfo a S_3.

2. Dimostrare che i gruppi degli esempi 11 e 13 del § 1.2 sono isomorfi. Trovare un sottogruppo di S_n isomorfo a questi gruppi.

3. Sia G un gruppo. Per $a \in G$, si definisca la traslazione destra a_R come l'applicazione $x \mapsto xa$ in G. Dimostrare che $G_R = \{a_R \mid a \in G\}$ è un gruppo di trasformazioni dell'insieme G e che $a \mapsto a_R^{-1}$ è un isomorfismo di G su G_R.

4. Il gruppo additivo degli interi è isomorfo a quello dei razionali (esempi 1 e 2 del § 1.2)?

5. Il gruppo additivo dei razionali è isomorfo al gruppo moltiplicativo dei razionali non nulli (esempi 2 e 5 del § 1.2)?

6. In $\mathbb{Z}$ porre $a \circ b = a + b - ab$. Mostrare che $(\mathbb{Z}, \circ, 0)$ è un monoide e che $a \mapsto 1 - a$ è un isomorfismo del monoide moltiplicativo $(\mathbb{Z}, \cdot, 1)$ su $(\mathbb{Z}, \circ, 0)$.

1.4 Associatività generalizzata. Commutatività

Sia $a_1, a_2, \ldots, a_n$ una successione finita di elementi di un monoide M. Iterando la composizione binaria si ottengono diversi prodotti a partire dalla successione. Per $n = 4$, per esempio, si hanno le possibilità

$$((a_1a_2)a_3)a_4, \quad (a_1(a_2a_3))a_4, \quad (a_1a_2)(a_3a_4), \quad a_1((a_2a_3)a_4), \quad a_1(a_2(a_3a_4)).$$

In generale dividiamo la successione in due parti $a_1, \ldots, a_m$ e $a_{m+1}, \ldots, a_n$, con $1 \leq m \leq n - 1$. Supponendo di saper già formare i prodotti di ciascuna parte, applichiamo la composizione binaria ai due risultati per ottenere un elemento di M, che è un prodotto associato alla successione $a_1, \ldots, a_n$. Variando m da 1 a $n - 1$ e prendendo tutti i prodotti delle due parti si ottengono i vari prodotti possibili. Affermiamo che la legge associativa garantisce l'uguaglianza di tutti questi prodotti. Ciò è chiaro per $n = 1$, se per «prodotto» intendiamo in tal caso a_1 stesso. Per la dimostrazione generale procederemo per induzione su n, premettendo un piccolo lemma.

Lemma. *Si definisca* $\prod_{i=1}^{n} a_i$ *ponendo* $\prod_{i=1}^{1} a_i = a_1$ *e* $\prod_{i=1}^{r+1} a_i = (\prod_{i=1}^{r} a_i)a_{r+1}$. *Allora*

$$\left(\prod_{i=1}^{n} a_i\right)\left(\prod_{j=1}^{m} a_{n+j}\right) = \prod_{k=1}^{n+m} a_k.$$

Dimostrazione. Per $m = 1$ l'uguaglianza vale per definizione. Supponiamola vera per $m = r$ e consideriamo $m = r + 1$. Abbiamo

$$\begin{aligned}\left(\prod_{i=1}^{n} a_i\right)\left(\prod_{j=1}^{r+1} a_{n+j}\right) &= \left(\prod_{i=1}^{n} a_i\right)\left(\left(\prod_{j=1}^{r} a_{n+j}\right) a_{n+r+1}\right) \\ &= \left(\left(\prod_{i=1}^{n} a_i\right)\left(\prod_{j=1}^{r} a_{n+j}\right)\right) a_{n+r+1} \\ &= \left(\prod_{k=1}^{n+r} a_k\right) a_{n+r+1} = \prod_{k=1}^{n+r+1} a_k. \qquad \square\end{aligned}$$

Consideriamo ora un prodotto qualsiasi associato ad $a_1, \ldots, a_n$. Ha la forma uv, dove u è un prodotto associato ad $a_1, \ldots, a_m$ e v uno associato ad $a_{m+1}, \ldots, a_n$. Per l'ipotesi induttiva su n possiamo supporre $u = \prod_{i=1}^{m} a_i$ e $v = \prod_{j=1}^{n-m} a_{m+j}$. Dal lemma segue $uv = \prod_{k=1}^{n} a_k$. Tutti i prodotti determinati dalla successione sono quindi uguali a $\prod_{k=1}^{n} a_k$. D'ora innanzi indicheremo questo prodotto univocamente determinato con $a_1 a_2 \cdots a_n$, omettendo le parentesi.

Se tutti gli a_i coincidono con a, denotiamo il prodotto con a^n e lo chiamiamo *potenza n-esima* di a. Basta contare i fattori per vedere che

$$a^m a^n = a^{m+n}, \qquad (a^m)^n = a^{mn}. \tag{6}$$

Ponendo $a^0 = 1$, la (6) vale per ogni $m, n \in \mathbb{N}$. Se a è invertibile, si definisce per $n \in \mathbb{N}$ la potenza $a^{-n} = (a^{-1})^n$, cioè il prodotto di n fattori a^{-1}. È chiaro che $a^{-n} = (a^n)^{-1}$, e si dimostra facilmente che la (6) vale per tutti gli $m, n \in \mathbb{Z}$. Lasciamo la verifica al lettore.

Per due elementi a, b di un monoide M può ben accadere che $ab \neq ba$. Per esempio, nel monoide $M(S)$ per $S = \{1, 2\}$, la cui tavola è la (1), si ha $\alpha\beta = \gamma$ e $\beta\alpha = \beta$. Se $ab = ba$ diciamo che a e b *commutano*; se ciò avviene per ogni $a, b \in M$, il monoide si chiama *commutativo*. I gruppi commutativi si chiamano di solito *abeliani* in onore di Niels Henrik Abel, grande matematico norvegese del primo Ottocento.[4] Adotteremo questa terminologia.

Per $a \in M$ definiamo il *centralizzante* $C(a)$, o $C_M(a)$ se occorre precisare M, come il sottoinsieme degli elementi b di M che commutano con a. È un sottomonoide: $1 \in C(a)$ poiché $1a = a = a1$, e per $b_1, b_2 \in C(a)$ vale

$$a(b_1 b_2) = (ab_1)b_2 = (b_1 a)b_2 = b_1(ab_2) = b_1(b_2 a) = (b_1 b_2)a.$$

Se $b \in C(a)$ è invertibile, anche $b^{-1} \in C(a)$, poiché moltiplicando l'uguaglianza $ab = ba$ a sinistra e a destra per b^{-1} si ottiene $b^{-1}a = ab^{-1}$. Se $M = G$ è un gruppo, $C(a)$ è dunque un sottogruppo.

È immediato che l'intersezione $\bigcap M_\alpha$ di una famiglia $\{M_\alpha\}$ di sottomonoidi di un monoide è un sottomonoide. Analogamente l'intersezione di una famiglia di sottogruppi di un gruppo è un sottogruppo. Per un sottoinsieme A di M definiamo il *centralizzante di A* mediante $C(A) = \bigcap_{a \in A} C(a)$. È un sottomonoide e, se M è un gruppo, un sottogruppo. Il sottomonoide $C(M)$ è detto *centro* di M.

Supponiamo che gli elementi $a_1, \ldots, a_n$ commutino a due a due: $a_i a_j = a_j a_i$ per ogni i, j. Consideriamo un prodotto $a_{1'} a_{2'} \cdots a_{n'}$ dove $1', 2', \ldots, n'$ è una permutazione di $1, 2, \ldots, n$. Se a_n è in posizione h, ossia $a_{h'} = a_n$, poiché gli a_i appartengono a $C(a_n)$, anche $a_{(h+1)'} \cdots a_{n'}$ vi appartiene. Si ha pertanto

$$a_{1'} \cdots a_{n'} = (a_{1'} \cdots a_{(h-1)'} a_{(h+1)'} \cdots a_{n'}) a_n.$$

[4] Una bella biografia è quella di Øystein Ore, *Niels Henrik Abel*, Minneapolis, University of Minnesota Press, 1957.

La successione $1', \dots, (h-1)', (h+1)', \dots, n'$ è una permutazione di $1, \dots, n-1$. Per induzione possiamo dunque supporre che il prodotto fra parentesi sia $a_1 a_2 \cdots a_{n-1}$. Ne ricaviamo $a_{1'} a_{2'} \cdots a_{n'} = a_1 a_2 \cdots a_n$. Il prodotto è quindi invariante rispetto a ogni permutazione degli argomenti. In particolare, se $ab = ba$, si ha

$$(ab)^n = a^n b^n. \tag{7}$$

Poiché $a^{-n} = (a^{-1})^n$, la (7) vale anche per esponenti negativi se a e b sono invertibili.

Per un monoide commutativo M si denota spesso la composizione con $+$ e si scrive $a + b$ in luogo di ab, e 0 in luogo di 1. La composizione si chiama allora *addizione* e 0 è l'*elemento zero*. Nella notazione additiva, al posto di a^{-1} si scrive $-a$ e lo si chiama *opposto* di a. La potenza a^n diventa na, l'n-esimo *multiplo* di a. Le regole delle potenze diventano le seguenti regole per i multipli:

$$ma + na = (m+n)a, \qquad m(na) = (mn)a, \tag{8}$$

$$n(a+b) = na + nb. \tag{9}$$

Se M è un gruppo abeliano, esse valgono per tutti gli interi m, n.

Esercizi

1. Sia A un monoide, $M(A)$ il monoide delle trasformazioni di A in sé, A_L l'insieme delle traslazioni sinistre a_L e A_R quello delle traslazioni destre a_R. Dimostrare che A_L, rispettivamente A_R, è il centralizzante di A_R, rispettivamente A_L, in $M(A)$ e che $A_L \cap A_R = \{c_R = c_L \mid c \in C\}$, dove C è il centro di A.

2. Dimostrare che, per $n \geq 3$, il centro di S_n ha ordine 1.

3. Dimostrare che è abeliano ogni gruppo nel quale $a^2 = 1$ per ogni a. Che cosa accade se $a^3 = 1$ per ogni a?

4. Data una composizione binaria, definire per induzione un *prodotto semplice* della successione $a_1, \dots, a_n$ come un prodotto $a_1 u$, con u prodotto semplice di $a_2, \dots, a_n$, oppure $v a_n$, con v prodotto semplice di $a_1, \dots, a_{n-1}$. Dimostrare che ogni prodotto di almeno 2^r elementi si può scrivere come prodotto semplice di r elementi (che sono a loro volta prodotti).

1.5 Sottomonoidi e sottogruppi generati da un sottoinsieme. Gruppi ciclici

Dato un sottoinsieme S di un monoide M o di un gruppo G, occorre spesso considerare il «più piccolo» sottomonoide di M, o sottogruppo di G, contenente S. Cerchiamo un sottomonoide (o sottogruppo) che contenga S e sia contenuto in ogni sottomonoide (o sottogruppo) che contiene S. Se un tale oggetto esiste, è unico: se $H(S)$ e $H'(S)$ soddisfano entrambi le condizioni, allora $H(S) \supset H'(S)$ e $H'(S) \supset H(S)$, dunque $H(S) = H'(S)$. Anche l'esistenza si stabilisce subito. Sia $\{M_\alpha\}$ (oppure $\{G_\alpha\}$) la famiglia di tutti i sottomonoidi di M (o sottogruppi di G) contenenti S. Formiamo l'intersezione $\langle S \rangle$ di tutti questi M_α (o G_α). È un sottomonoide (sottogruppo), poiché tale è l'intersezione di sottomonoidi (sottogruppi). Naturalmente $\langle S \rangle \supset S$. Inoltre ogni sottomonoide N di M (o sottogruppo di G) contenente S è uno degli M_α (o G_α), dunque contiene la loro intersezione $\langle S \rangle$. Chiameremo $\langle S \rangle$ il sottomonoide (o sottogruppo) *generato* da S. Se $S = \{s_1, \dots, s_r\}$ è finito, scriveremo $\langle s_1, \dots, s_r \rangle$ in luogo del più pesante $\langle \{s_1, \dots, s_r\} \rangle$. Un caso importante è $\langle S \rangle = M$ (o G). Diciamo allora che il monoide M (o il gruppo G) è *generato* da S,

oppure che S è un *insieme di generatori*. Ciò significa semplicemente che nessun sottomonoide proprio di M (o sottogruppo proprio di G) contiene S.

Il carattere non costruttivo della definizione di $\langle S\rangle$ potrà lasciare perplesso il lettore. La matematica moderna è ricca di definizioni di questo genere e bisogna imparare a maneggiarle con disinvoltura. Tuttavia, quando è disponibile, una definizione costruttiva è gradevole e spesso utile. Per $\langle S\rangle$ ne esiste una, come ora vedremo. Cominciamo dai monoidi. Quale forma hanno gli elementi di $\langle S\rangle$? Essendo un sottomonoide che contiene S, esso contiene 1 e ogni prodotto $s_1 \cdots s_r$ con fattori in S, non necessariamente distinti. Quindi

$$\langle S\rangle \supset \langle S\rangle' = \{1,\ s_1 s_2 \cdots s_r \mid s_i \in S\}. \tag{10}$$

Qui $\langle S\rangle'$ è il sottoinsieme di M costituito da 1 e dai prodotti di un numero finito di elementi di S. Affermiamo che $\langle S\rangle = \langle S\rangle'$. Infatti $\langle S\rangle'$ contiene S, essendo ammesso $r = 1$ in (10); contiene l'elemento neutro e il prodotto di due prodotti della forma indicata ha ancora la stessa forma. Dunque $\langle S\rangle'$ è un sottomonoide di M contenente S, sicché $\langle S\rangle' \supset \langle S\rangle$. Con l'inclusione precedente otteniamo l'uguaglianza. Si ha così una descrizione costruttiva di $\langle S\rangle$: esso consiste di 1 e di tutti i prodotti finiti di elementi di S.

Nel caso di un gruppo, prendiamo per $\langle S\rangle'$ il sottoinsieme di G formato da 1 e dai prodotti finiti di elementi di S o dei loro inversi. In altre parole,

$$\langle S\rangle' = \{1,\ s_1 s_2 \cdots s_r \mid s_i \in S \text{ oppure } s_i^{-1} \in S\}. \tag{11}$$

È immediato che $\langle S\rangle \supset \langle S\rangle'$, che $\langle S\rangle' \supset S$ e che $\langle S\rangle'$ è un sottogruppo. Quindi $\langle S\rangle' = \langle S\rangle$.

Restringiamoci ora ai gruppi e consideriamo i più semplici: quelli che hanno un unico generatore. Se $G = \langle a\rangle$, diciamo che G è *ciclico* di generatore a. La discussione precedente, o le regole delle potenze, mostrano che $\langle a\rangle = \{a^k \mid k \in \mathbb{Z}\}$ è abeliano. Un esempio è il gruppo additivo degli interi $(\mathbb{Z}, +, 0)$, generato da 1 (o da -1).

Consideriamo l'applicazione $n \mapsto a^n$ di $\mathbb{Z}$ in $\langle a\rangle$. Essa è suriettiva per la descrizione di $\langle a\rangle$. Inoltre $m + n \mapsto a^{m+n} = a^m a^n$ e $0 \mapsto 1$. Se è iniettiva, è dunque un isomorfismo. Supponiamo che non sia un isomorfismo. Esistono allora $m \neq n$ con $a^m = a^n$; possiamo supporre $n > m$. Segue $a^{n-m} = a^n a^{-m} = a^m a^{-m} = 1$, sicché esistono interi positivi p con $a^p = 1$. Sia r il minimo di essi. Affermiamo che

$$\langle a\rangle = \{1, a, a^2, \ldots, a^{r-1}\} \tag{12}$$

e che gli elementi elencati sono distinti: dunque $|\langle a\rangle| = r$. Per ogni $a^m \in \langle a\rangle$ la divisione fra interi dà $m = rq + p$ con $0 \leq p < r$. Allora $a^m = a^{rq+p} = (a^r)^q a^p = 1^q a^p = a^p$, uno degli elementi in (12). Se k e l sono distinti fra $0, 1, \ldots, r-1$, non si può avere $a^k = a^l$: per $l > k$ ne seguirebbe $a^{l-k} = 1$ con $0 < l - k < r$, contro la scelta di r. Quando $n \mapsto a^n$ non è un isomorfismo, $\langle a\rangle$ è dunque finito. Pertanto ogni gruppo ciclico infinito è isomorfo a $(\mathbb{Z}, +, 0)$ e due gruppi ciclici infiniti sono isomorfi.

Mostriamo che sono isomorfi anche due gruppi ciclici finiti dello stesso ordine. Sia $\langle b\rangle$ di ordine r. Come sopra, $\langle b\rangle = \{1, b, \ldots, b^{r-1}\}$, dove r è il minimo intero positivo con $b^r = 1$. Osserviamo che $a^h = 1$ implica $r \mid h$ (r divide h). Scrivendo $h = qr + s$, $0 \leq s < r$, si trova $1 = a^h = (a^r)^q a^s = 1^q a^s = a^s$; per minimalità di r, $s = 0$, dunque $h = qr$. Ora $a^m = a^n$ implica anche $b^m = b^n$. Infatti $a^{m-n} = 1$, quindi $m - n = qr$ e $b^{m-n} = (b^r)^q = 1$, ossia $b^m = b^n$. Per simmetria vale l'implicazione inversa. Vi è dunque una corrispondenza biunivoca fra $\langle a\rangle$ e $\langle b\rangle$ che associa a^n a b^n. Poiché ad $a^m a^n = a^{m+n}$ corrisponde $b^{m+n} = b^m b^n$, tale corrispondenza è un isomorfismo. Abbiamo dimostrato il seguente risultato.

Teorema 1.2. *Due gruppi ciclici dello stesso ordine, finito o infinito, sono isomorfi.*

Il gruppo $(\mathbb{Z}, +, 0)$ è dunque un modello del gruppo ciclico di ordine infinito. Per ogni intero positivo r il gruppo moltiplicativo U_r delle radici r-esime dell'unità (esempio 13 del § 1.2) è un modello per l'ordine r. I suoi elementi sono

$$e^{2k\pi i/r} = \cos(2k\pi/r) + i\sin(2k\pi/r), \qquad k = 0, 1, \ldots, r-1.$$

Da $e^{i\theta}e^{i\theta'} = e^{i(\theta+\theta')}$ segue che $a = e^{2\pi i/r}$ genera U_r.

La nozione di gruppo ciclico permette di classificare gli elementi di un gruppo qualsiasi G. Per $a \in G$ diciamo che a è di *ordine infinito*, oppure di *ordine finito* r, secondo che $\langle a\rangle$ sia infinito o finito di ordine r. Nel primo caso $a^m \neq 1$ per $m \neq 0$; nel secondo $a^r = 1$ e r è il minimo intero positivo con questa proprietà. Inoltre, $a^m = 1$ comporta che m sia multiplo di r. Indicheremo l'ordine di a, finito o infinito, con $o(a)$. Se $o(a) = r = st$ con s, t interi positivi, si vede che $o(a^s) = t$. Più in generale, se $o(a) = r < \infty$, per ogni intero $k \neq 0$ l'ordine di a^k è $r/(r,k)$, equivalente per $k > 0$ a $[r,k]/k$, dove le parentesi quadre denotano il minimo comune multiplo e le tonde il massimo comune divisore (esercizio 4).

I gruppi ciclici sono i gruppi più semplici. Non sorprende quindi che per essi molti problemi della teoria dei gruppi abbiano una soluzione agevole. Per esempio, se ne possono determinare tutti i sottogruppi, un compito in genere arduo per gli altri gruppi. Dimostreremo ora il seguente teorema.

Teorema 1.3. *Ogni sottogruppo di un gruppo ciclico $\langle a\rangle$ è ciclico. Se $\langle a\rangle$ è infinito, i sottogruppi diversi da 1 sono infiniti e $s \mapsto \langle a^s\rangle$ è una biiezione di $\mathbb{N}$ sull'insieme dei sottogruppi di $\langle a\rangle$. Se $\langle a\rangle$ è finito di ordine r, l'ordine di ogni suo sottogruppo divide r e per ciascun divisore positivo q di r esiste uno e un solo sottogruppo di ordine q.*

Dimostrazione. Sia H un sottogruppo di $\langle a\rangle$. Se $H = 1$ (ossia $H = \{1\}$), allora $H = \langle 1\rangle$. Sia $H \neq 1$. Esiste $n \neq 0$ in $\mathbb{Z}$ con $a^n \in H$; poiché anche $a^{-n} = (a^n)^{-1} \in H$, possiamo supporre $n > 0$. Sia s il minimo intero positivo con $a^s \in H$. Affermiamo che $H = \langle a^s\rangle$. Per $a^m \in H$ si scriva $m = qs + t$, $0 \leq t < s$. Allora $a^t = a^m(a^s)^{-q} \in H$. La minimalità di s dà $t = 0$, cosicché $a^m = (a^s)^q \in \langle a^s\rangle$. Ciò vale per ogni elemento di H, e l'altra inclusione è immediata. Questo dimostra la prima affermazione.

Se $\langle a\rangle$ è infinito, abbiamo visto che $a^m \neq a^n$ per $m \neq n$. Per ogni $s > 0$, le potenze a^{sm} al variare di $m = 0, \pm1, \pm2, \ldots$ sono dunque distinte e $\langle a^s\rangle$ è infinito. Inoltre s è il minimo intero positivo con $a^s \in \langle a^s\rangle$. Ogni sottogruppo diverso da 1 è quindi infinito, e $s \mapsto \langle a^s\rangle$ dà una corrispondenza biunivoca fra gli interi positivi e tali sottogruppi.

Supponiamo ora $\langle a\rangle$ finito di ordine r, cosicché $\langle a\rangle = \{1, a, \ldots, a^{r-1}\}$. Abbiamo visto che ogni sottogruppo $H \neq 1$ ha forma $\langle a^s\rangle$, con s minimo positivo tale che $a^s \in H$. Si ha $s \mid r$: infatti, posto $r = qs + t$ con $0 \leq t < s$, otteniamo $1 = a^r = (a^s)^q a^t$, da cui $a^t = (a^s)^{-q} \in H$. La minimalità di s impone $t = 0$ e quindi $r = qs$. Gli elementi di H sono

$$1,\ a^s,\ a^{2s}, \ldots, a^{(q-1)s}, \tag{13}$$

con $a^{sq} = a^r = 1$. Il caso $H = 1$ è compreso prendendo $s = r$. Abbiamo una biiezione $s \mapsto \langle a^s\rangle$ fra i divisori positivi di r e i sottogruppi di $\langle a\rangle$. L'ordine del sottogruppo corrispondente a s è $q = r/s$ e, quando s percorre i divisori positivi di r, anche q li percorre tutti. Dunque l'ordine di ogni sottogruppo divide r, e per ogni divisore positivo q esiste un unico sottogruppo di tale ordine. □

Il sottogruppo di ordine q del gruppo ciclico di ordine r si può elencare come nella (13). Un'altra sua caratterizzazione spesso utile è la seguente.

Corollario. *Se $\langle a\rangle$ ha ordine finito r, il suo sottogruppo H di ordine $q \mid r$ è l'insieme degli elementi $b \in \langle a\rangle$ tali che $b^q = 1$.*

Dimostrazione. Ogni elemento di H ha forma a^{sk} con $s = r/q$. Allora $(a^{sk})^q = a^{rk} = 1$. Viceversa, se $b = a^m$ soddisfa $b^q = 1$, si ha $a^{mq} = 1$, quindi $mq = kr$. Ne segue $m = ks$ e $b = (a^s)^k \in H$. □

Dopo i gruppi ciclici, i gruppi più semplici sono quelli abeliani finitamente generati, ossia dotati di un insieme finito di generatori. Ne fanno parte i gruppi abeliani finiti. Nel capitolo 3 determineremo la struttura dei gruppi di questa classe e ne daremo una classificazione completa mediante invarianti numerici. Indipendentemente dalla teoria della struttura, ricaviamo ora un criterio affinché un gruppo abeliano finito sia ciclico. Servirà per dimostrare un importante teorema sui campi (teorema 2.18). Per enunciarlo introduciamo l'*esponente* $\exp G$ di un gruppo finito G: è il minimo intero positivo e tale che $x^e = 1$ per ogni $x \in G$. Per esempio $\exp S_3 = 6 = |S_3|$.

Teorema 1.4. *Un gruppo abeliano finito G è ciclico se e solo se* $\exp G = |G|$.

La dimostrazione si basa su due lemmi interessanti anche di per sé.

Lemma 1. *Siano g, h elementi di un gruppo abeliano G di ordini finiti rispettivamente m, n, primi fra loro. Allora $o(gh) = mn$.*

Dimostrazione. Se $(gh)^r = 1$, allora $k = g^r = h^{-r} \in \langle g\rangle \cap \langle h\rangle$. Dunque $o(k)$ divide sia m sia n, cosicché $o(k) = 1$. Da $(gh)^r = 1$ segue pertanto $g^r = h^r = 1$, quindi $m \mid r$, $n \mid r$ e $mn = [m, n] \mid r$. D'altra parte $(gh)^{mn} = g^{mn}h^{mn} = 1$. Ne segue $o(gh) = mn$. □

Lemma 2. *Sia G un gruppo abeliano finito e sia g un suo elemento di ordine massimo. Allora* $\exp G = o(g)$.

Dimostrazione. Dobbiamo provare che $h^{o(g)} = 1$ per ogni $h \in G$. Scriviamo $o(g) = p_1^{e_1} \cdots p_s^{e_s}$ e $o(h) = p_1^{f_1} \cdots p_s^{f_s}$, con i p_i primi distinti ed $e_i, f_i \geq 0$. Se $h^{o(g)} \neq 1$, per qualche i si ha $f_i > e_i$ e possiamo supporre $f_1 > e_1$. Poniamo $g' = g^{p_1^{e_1}}$ e $h' = h^{p_2^{f_2} \cdots p_s^{f_s}}$. Allora $o(g') = p_2^{e_2} \cdots p_s^{e_s}$ e $o(h') = p_1^{f_1}$. Per il lemma 1, $o(g'h') = p_1^{f_1} p_2^{e_2} \cdots p_s^{e_s} > o(g)$, contro la massimalità di $o(g)$. □

Dimostrazione del teorema 1.4. Se $G = \langle g\rangle$, allora $|G| = o(g)$ e $\exp G = |G|$. Viceversa, sia G un gruppo abeliano finito con $\exp G = |G|$. Per il lemma 2 esiste un elemento g tale che $\exp G = o(g)$. Allora $|G| = o(g) = |\langle g\rangle|$, sicché $G = \langle g\rangle$. □

Esercizi

1. Come nel § 1.4 sia $C(A)$ il centralizzante di A nel monoide M (o gruppo G). Osservare che $C(C(A)) \supset A$ e che $A \subset B$ implica $C(A) \supset C(B)$. Deducerne $C(C(C(A))) = C(A)$. Senza usare la forma esplicita degli elementi di $\langle A\rangle$, dimostrare che $C(A) = C(\langle A\rangle)$. (Suggerimento: se $c \in C(A)$, allora $A \subset C(c)$ e quindi $\langle A\rangle \subset C(c)$.) Usare quest'ultimo risultato per dimostrare che un monoide o un gruppo generato da elementi che commutano a due a due è commutativo.

2. Sia M un monoide generato da un insieme S i cui elementi sono tutti invertibili. Dimostrare che M è un gruppo.

3. Sia G un gruppo abeliano con un insieme finito di generatori. Se G è periodico, nel senso che tutti i suoi elementi hanno ordine finito, dimostrare che è finito.

4. Se $o(g) = n$, dimostrare che, per $k \neq 0$, g^k ha ordine $n/(n, k)$, ossia $[n, |k|]/|k|$. Dimostrare che il numero di generatori di $\langle g\rangle$ è il numero degli interi positivi minori o uguali a n e primi con n. Si denota tale numero con $\varphi(n)$ e si chiama φ *funzione di Eulero.*

5. Dimostrare che ogni sottogruppo finitamente generato del gruppo additivo dei razionali $(\mathbb{Q}, +, 0)$ è ciclico. Dedurne che questo gruppo non è isomorfo al prodotto diretto di due sue copie.

6. Siano a, b come nel lemma 1. Dimostrare che $\langle a\rangle \cap \langle b\rangle = 1$ e $\langle a, b\rangle = \langle ab\rangle$.

7. Se $o(a) = n = rs$ con $(r, s) = 1$, dimostrare che $\langle a\rangle \cong \langle b\rangle \times \langle c\rangle$, dove $o(b) = r$ e $o(c) = s$. Dedurne che ogni gruppo ciclico finito è isomorfo a un prodotto diretto di gruppi ciclici di ordine potenza di primo.

1.6 Decomposizione delle permutazioni in cicli

Una permutazione γ di $\{1, 2, \ldots, n\}$ che permuta ciclicamente $r > 1$ elementi $i_1, i_2, \ldots, i_r$, nel senso che

$$\gamma(i_1) = i_2, \quad \gamma(i_2) = i_3, \quad \ldots, \quad \gamma(i_{r-1}) = i_r, \quad \gamma(i_r) = i_1, \tag{14}$$

e fissa, cioè lascia invariati, gli altri numeri si chiama *ciclo* o *r-ciclo*. Scriviamo

$$\gamma = (i_1 i_2 \cdots i_r). \tag{15}$$

Si può scrivere altrettanto bene $\gamma = (i_2 i_3 \cdots i_r i_1) = (i_3 i_4 \cdots i_r i_1 i_2)$ e così via. La permutazione γ^2 manda i_1 in i_3, i_2 in i_4, …, i_r in i_2. In generale, per $1 \le k \le r$,

$$\gamma^k(i_j) = \begin{cases} i_{j+k} & j + k \le r, \\ i_{j+k-r} & j + k > r. \end{cases} \tag{16}$$

Dunque $\gamma^r = 1$, ma $\gamma^k \neq 1$ per $1 \le k < r$, cosicché γ ha ordine r.

Due cicli γ, γ' sono *disgiunti* se i loro simboli non hanno lettere comuni. In tal caso ogni numero mosso da uno dei due è fissato dall'altro. Se $\gamma(i) \neq i$, si ha perciò $\gamma\gamma'(i) = \gamma(i)$ e, poiché anche $\gamma^2(i) \neq \gamma(i)$, si ha $\gamma'\gamma(i) = \gamma(i)$. Similmente, se $\gamma'(i) \neq i$, allora $\gamma'\gamma(i) = \gamma'(i) = \gamma\gamma'(i)$. Infine, se $\gamma(i) = i = \gamma'(i)$, allora $\gamma\gamma'(i) = \gamma'\gamma(i)$. Dunque $\gamma\gamma' = \gamma'\gamma$: due cicli disgiunti commutano. Sia ora α un prodotto di cicli disgiunti,

$$\alpha = (i_1 i_2 \cdots i_r)(j_1 j_2 \cdots j_s) \cdots (l_1 l_2 \cdots l_u). \tag{17}$$

Indichiamo con m il minimo comune multiplo di $r, s, \ldots, u$. Affermiamo che α ha ordine m. Poniamo $\gamma_1 = (i_1 \cdots i_r)$, $\gamma_2 = (j_1 \cdots j_s)$, …, $\gamma_k = (l_1 \cdots l_u)$. Si ha $\alpha^m = \gamma_1^m \gamma_2^m \cdots \gamma_k^m = 1$. D'altra parte α e le sue potenze permutano fra loro $i_1, \ldots, i_r$ e la restrizione di α a questi elementi è quella di γ_1. Se dunque $\alpha^n = 1$, si ha $\gamma_1^n = 1$ e $r \mid n$. Analogamente $s, \ldots, u$ dividono n, quindi anche il loro minimo comune multiplo con r. L'ordine di α è perciò tale minimo comune multiplo.

Conviene estendere la definizione e la notazione ai cicli di lunghezza uno: per ogni i, (i) indica l'identità. Con questa convenzione ogni permutazione è un prodotto di cicli disgiunti. Ad esempio, se

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 6 & 5 & 4 & 8 & 2 & 7 & 1 \end{pmatrix},$$

allora $\alpha(1) = 3$, $\alpha(3) = 5$, $\alpha(5) = 8$, $\alpha(8) = 1$; inoltre $\alpha(2) = 6$, $\alpha(6) = 2$, $\alpha(4) = 4$ e $\alpha(7) = 7$. Ne segue $\alpha = (7)(4)(26)(1358)$.

Per un'α qualsiasi cominciamo da un numero i_1 a piacere fra $1, \ldots, n$ e poniamo $\alpha(i_1) = i_2$, $\alpha(i_2) = i_3$, e così via, finché compare un numero già incontrato. La prima ripetizione si ha

quando $i_{r+1} = \alpha(i_r) = i_1$: infatti $i_k = \alpha^{k-1}(i_1)$ e, se $i_k = i_l$ con $l > k$, allora $\alpha^{l-k}(i_1) = i_1$. Dunque la successione $i_1, \dots, i_r$ viene permutata ciclicamente da α. Se $r < n$, scegliamo j_1 fuori da $\{i_1, \dots, i_r\}$. Se fosse $\alpha^m(j_1) = \alpha^q(i_1)$, allora $j_1 = \alpha^{q-m}(i_1)$ apparterrebbe a $\{i_1, \dots, i_r\}$, contro la scelta fatta. Otteniamo così una nuova successione di numeri $j_1, \dots, j_s$, permutati ciclicamente da α, disgiunta dalla prima. Procedendo in questo modo si esaurisce $\{1, \dots, n\}$. Confrontando le immagini di ogni i mediante α e mediante $(l_1 \cdots l_u) \cdots (i_1 \cdots i_r)$, otteniamo

$$\alpha = (l_1 \cdots l_u) \cdots (i_1 \cdots i_r),$$

un prodotto di cicli disgiunti. I diversi cicli della fattorizzazione commutano e possiamo aggiungere od omettere i cicli banali di lunghezza uno. A meno dell'ordine dei fattori e dell'inserzione o omissione degli 1-cicli, la fattorizzazione è unica. Infatti, se ve ne fosse una essenzialmente diversa dalla (17), per qualche $i \neq j$ tali che j segua immediatamente i in un ciclo della (17), ciò non avverrebbe nell'altra. La prima fattorizzazione darebbe allora $\alpha(i) = j$ e la seconda $\alpha(i) \neq j$: una contraddizione.

Un ciclo della forma (ab) è detto *trasposizione*. Si verifica facilmente che

$$(i_1 i_2 \cdots i_r) = (i_1 i_r) \cdots (i_1 i_3)(i_1 i_2), \tag{18}$$

un prodotto di $r - 1$ trasposizioni. Ogni $\alpha \in S_n$ è dunque un prodotto di trasposizioni. Più precisamente, dalla (17) ricaviamo un prodotto di $(r-1) + (s-1) + \cdots + (u-1)$ trasposizioni. Indichiamo questo numero, univocamente determinato da α, con $N(\alpha)$. È chiaro che $N(1) = 0$. La fattorizzazione in trasposizioni non è unica: per esempio $(123) = (13)(12) = (12)(23) = (23)(13)$. Tuttavia, come vedremo, tutte le fattorizzazioni di una data α in trasposizioni hanno un tratto comune: il numero dei fattori ha sempre la stessa parità, ossia è sempre pari oppure sempre dispari. La dimostrazione si basa su una semplice formula, interessante di per sé:

$$(ab)(ac_1 \cdots c_h bd_1 \cdots d_k) = (bd_1 \cdots d_k)(ac_1 \cdots c_h). \tag{19}$$

Ammettiamo $h = 0$ o $k = 0$, nel senso che nel ciclo manchino le c o le d. Basta confrontare le immagini di ogni $i \in \{1, \dots, n\}$ per verificare la (19). Poiché $(ab)^{-1} = (ab)$, moltiplicando entrambi i membri a sinistra per (ab) si ottiene

$$(ab)(bd_1 \cdots d_k)(ac_1 \cdots c_h) = (ac_1 \cdots c_h bd_1 \cdots d_k). \tag{20}$$

Con la definizione precedente $N((ac_1 \cdots c_h bd_1 \cdots d_k)) = h + k + 1$ e $N((bd_1 \cdots d_k)(ac_1 \cdots c_h)) = h+k$. Ne segue che $N((ab)\alpha) = N(\alpha) - 1$ se a, b compaiono nello stesso ciclo della decomposizione di α, mentre $N((ab)\alpha) = N(\alpha) + 1$ se compaiono in cicli diversi. Se α è un prodotto di m trasposizioni, da $N(1) = 0$ segue perciò $N(\alpha) = \sum_{i=1}^{m} \varepsilon_i$, con $\varepsilon_i = \pm 1$. Cambiare un termine -1 in 1 equivale ad aggiungere 2 alla somma: la parità rimane invariata. Effettuando la sostituzione per tutti i termini negativi otteniamo m. Dunque m e $N(\alpha)$ hanno la stessa parità, e il numero dei fattori in due qualunque fattorizzazioni in trasposizioni ha la parità di $N(\alpha)$.

Chiamiamo α *pari* o *dispari* secondo che si fattorizzi in un numero pari o dispari di trasposizioni (equivalentemente, secondo la parità di $N(\alpha)$). Definiamo il suo *segno*, $\operatorname{sg}\alpha$, ponendo

$$\operatorname{sg}\alpha = 1 \quad \text{se } \alpha \text{ è pari}, \qquad \operatorname{sg}\alpha = -1 \quad \text{se } \alpha \text{ è dispari}. \tag{21}$$

Allora $\operatorname{sg} 1 = 1$; se $\alpha = (ab) \cdots (kl)$ e $\beta = (pq) \cdots (uv)$, il prodotto è $\alpha\beta = (ab) \cdots (kl)(pq) \cdots (uv)$. Esso è pari se e solo se entrambi i fattori sono pari o entrambi dispari; è dispari se hanno parità diversa. Quindi

$$\operatorname{sg}(\alpha\beta) = (\operatorname{sg}\alpha)(\operatorname{sg}\beta). \tag{22}$$

È chiaro inoltre che le permutazioni pari costituiscono un sottogruppo A_n di S_n, detto *gruppo alterno* (di grado n). Se ne elenchiamo gli elementi $\alpha_1, \dots, \alpha_m$ e $n \geq 2$, otteniamo m permutazioni dispari distinte $\alpha_1(ab), \dots, \alpha_m(ab)$. Sono tutte: per β dispari, $\beta(ab)$ è pari, dunque $\beta(ab) = \alpha_i$ per qualche i e $\beta = \alpha_i(ab)$. Pertanto $|S_n| = 2m = 2|A_n|$ e $|A_n| = n!/2$ per $n \geq 2$.

Esercizi

1. Scrivere (456)(567)(671)(123)(234)(345) come prodotto di cicli disgiunti.
2. Dimostrare che per $n \geq 3$ il gruppo A_n è generato dai 3-cicli (abc).
3. Determinare il segno di $\begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ n & n-1 & \cdots & 2 & 1 \end{pmatrix}$.
4. Dimostrare che, per ogni permutazione α, si ha $\alpha(i_1 i_2 \cdots i_r)\alpha^{-1} = (\alpha(i_1)\,\alpha(i_2)\cdots\alpha(i_r))$.
5. Dimostrare che S_n è generato dalle $n-1$ trasposizioni $(12), (13), \ldots, (1n)$ e anche dalle $n-1$ trasposizioni $(12), (23), \ldots, (n-1\ \ n)$.

1.7 Orbite. Classi laterali di un sottogruppo

Sia G un gruppo di trasformazioni di un insieme S. Esso definisce una relazione di equivalenza su S: scriviamo $x \sim_G y$ (si legge «x è G-equivalente a y») se $y = \alpha(x)$ per qualche $\alpha \in G$. La riflessività, la simmetria e la transitività seguono subito dalla definizione di gruppo di trasformazioni. Infatti $x = 1_S(x)$; da $y = \alpha(x)$ segue $x = \alpha^{-1}(y)$; infine da $y = \alpha(x)$ e $z = \beta(y)$ segue $z = (\beta\alpha)(x)$. Ora $1_S \in G$, e $\alpha^{-1}, \beta\alpha \in G$ se $\alpha, \beta \in G$. La classe di G-equivalenza di x è $Gx = \{\alpha(x) \mid \alpha \in G\}$ ed è detta *G-orbita* di x. Se, ad esempio, G è il gruppo delle rotazioni del piano intorno all'origine, l'orbita del punto P è la circonferenza passante per P e avente centro nell'origine. Come per ogni relazione di equivalenza, le orbite costituiscono una partizione di S. Può accadere che vi sia una sola orbita: $S = Gx$ per qualche x e quindi per ogni x. In tal caso G si chiama gruppo di trasformazioni *transitivo* su S. È chiaro che S_n è transitivo su $\{1, 2, \ldots, n\}$. Il lettore verificherà senza difficoltà che ciò vale anche per A_n quando $n \geq 3$.

Sia, d'altra parte, $\alpha \in S_n$ e sia $\alpha = (i_1 \cdots i_r)(j_1 \cdots j_s)\cdots(l_1 \cdots l_u)$ la sua fattorizzazione in cicli disgiunti, compresi gli 1-cicli, in modo che ogni lettera di $\{1, \ldots, n\}$ compaia una e una sola volta fra tutte le lettere scritte. Gli insiemi

$$\{i_1, \ldots, i_r\}, \quad \{j_1, \ldots, j_s\}, \quad \ldots, \quad \{l_1, \ldots, l_u\}$$

sono le orbite di $\{1, \ldots, n\}$ determinate dal sottogruppo ciclico $\langle\alpha\rangle$ di S_n. Si ottiene così un'altra interpretazione del numero $N(\alpha)$ del § 1.6: $N(\alpha) = \sum(k-1)$, dove k percorre le cardinalità delle orbite determinate da $\langle\alpha\rangle$.

Sia ora G un gruppo qualsiasi e H un suo sottogruppo. Ricordiamo i gruppi di trasformazioni G_L delle traslazioni sinistre $g_L : x \mapsto gx$ e G_R delle traslazioni destre g_R, entrambi operanti su G. Poiché, assegnati y, x, sia $y = gx$ sia $y = xg$ sono risolubili rispetto a g, entrambi i gruppi sono transitivi. Indichiamo con $H_L(G)$ il sottoinsieme di G_L costituito dalle applicazioni h_L su G per $h \in H$. Poiché H è un sottogruppo e $g \mapsto g_L$ è un isomorfismo, $H_L(G)$ è un sottogruppo di G_L, dunque un gruppo di trasformazioni dell'insieme G. Quali ne sono le orbite? Per $x \in G$ l'orbita è chiaramente

$$Hx = \{hx \mid h \in H\}. \tag{23}$$

Nella letteratura sui gruppi essa viene chiamata ora classe laterale sinistra, ora classe laterale destra di x rispetto a H. La maggioranza sembra preferire la seconda terminologia; la adotteremo e chiameremo Hx *classe laterale destra* di x rispetto a H. Abbiamo la partizione $G = \bigcup_{x \in G} Hx$. Due classi laterali destre Hx e Hy qualsiasi hanno la stessa cardinalità, poiché $(x^{-1}y)_R : z \mapsto z(x^{-1}y)$ è una biiezione di Hx su Hy. Essendo $H = H1$ una delle classi, $|Hx| = |H|$.

In particolare, supponiamo G finito, con $|G| = n$ e $|H| = m$. La partizione è

$$G = Hx_1 \cup Hx_2 \cup \cdots \cup Hx_r, \tag{24}$$

dove abbiamo indicato le classi distinte, sicché $Hx_i \cap Hx_j = \varnothing$ per $i \neq j$. Il loro numero r si chiama *indice* di H in G e si denota con $[G : H]$. Da $|Hx_i| = m$ e dalla (24) segue $n = mr$. È così dimostrato un teorema fondamentale dovuto a Lagrange.

Teorema 1.5. *L'ordine di un sottogruppo H di un gruppo finito G divide l'ordine di G. Più precisamente,*

$$|G| = |H|[G : H].$$

Abbiamo inoltre il seguente corollario.

Corollario. *Se G è finito di ordine n, allora $x^n = 1$ per ogni $x \in G$.*

Dimostrazione. Sia m l'ordine di $\langle x \rangle$. Abbiamo $x^m = 1$ e $n = mr$, quindi $x^n = 1$. □

Ai risultati sulle classi destre corrispondono quelli per le classi sinistre, cioè le orbite di G determinate dal gruppo di trasformazioni $H_R(G)$. In questo caso l'orbita di x è $xH = \{xh \mid h \in H\}$, detta *classe laterale sinistra* di x rispetto a H. Per una classe destra Hx, l'insieme degli inversi $(hx)^{-1} = x^{-1}h^{-1}$ dei suoi elementi è la classe sinistra $x^{-1}H$. L'applicazione $Hx \mapsto x^{-1}H$ è evidentemente una biiezione dell'insieme delle classi destre su quello delle classi sinistre. Ne segue che le due famiglie hanno la stessa cardinalità. Come per i gruppi finiti, questa cardinalità si chiama *indice di H in G* e si indica con $[G : H]$.

Esercizi

1. Determinare le classi laterali di $\langle \alpha \rangle$ in S_4, dove $\alpha = (1234)$.

2. Se G è finito e H, K sono sottogruppi con $H \supset K$, dimostrare che $[G : K] = [G : H][H : K]$.

3. Siano H_1, H_2 sottogruppi di G. Dimostrare che ogni classe laterale destra rispetto a $H_1 \cap H_2$ è intersezione di una classe destra di H_1 con una classe destra di H_2. Dedurne il *teorema di Poincaré*: se H_1, H_2 hanno indice finito in G, lo stesso vale per $H_1 \cap H_2$.

4. Sia G un gruppo finitamente generato e H un sottogruppo di indice finito. Dimostrare che H è finitamente generato.

5. Siano H, K sottogruppi di un gruppo G. Dimostrare che le applicazioni $x \mapsto hxk$, per $h \in H$ e $k \in K$, costituiscono un gruppo di trasformazioni di G. Mostrare che l'orbita di x è $HxK = \{hxk \mid h \in H,\ k \in K\}$, detta *classe laterale doppia* di x rispetto a (H, K). Se G è finito, dimostrare che

$$|HxK| = |H|\,[K : x^{-1}Hx \cap K] = |K|\,[H : xKx^{-1} \cap H].$$

6. Sia H un sottogruppo di un gruppo finito G. Dimostrare che esiste un sottoinsieme $\{z_1, \ldots, z_r\}$ di G che è simultaneamente un sistema di rappresentanti delle classi laterali sinistre e destre di H: G è unione disgiunta degli z_iH e anche degli Hz_i, $1 \leq i \leq r$. (Suggerimento: per $g \in G$ scrivere $HgH = \bigcup_{j=1}^{s} x_j gH$, con $x_j \in H$ e $x_j gH \cap x_k gH = \varnothing$ per $j \neq k$. Osservare che il numero di classi destre di H contenute in HgH è s e scrivere $HgH = \bigcup_{j=1}^{s} Hgy_j$ con $y_j \in H$. Porre $z_j = x_j g y_j$ e verificare $HgH = \bigcup z_jH = \bigcup Hz_j$.)

1.8 Congruenze. Monoidi e gruppi quoziente

Nella teoria elementare dei numeri due interi a, b si dicono congruenti modulo l'intero m, e si scrive $a \equiv b \pmod m$, se $a - b$ è un multiplo di m, ossia $a - b = km$, $k \in \mathbb{Z}$.[5] Per m fissato si ottiene una relazione di equivalenza. Infatti $a \equiv a \pmod m$, poiché $a - a = 0 = 0m$. Da $a \equiv b \pmod m$ segue $b \equiv a \pmod m$, poiché $a - b = km$ implica $b - a = (-k)m$. Infine $a \equiv b \pmod m$ e $b \equiv c \pmod m$ implicano $a \equiv c \pmod m$, poiché da $a - b = km$ e $b - c = lm$ segue $a - c = (k + l)m$. Nel gruppo additivo $(\mathbb{Z}, +, 0)$ si possono sommare le congruenze modulo m: se $a \equiv a' \pmod m$ e $b \equiv b' \pmod m$, allora $a+b \equiv a'+b' \pmod m$. Infatti $a-a' = km$ e $b-b' = lm$ danno $a + b - (a' + b') = (k + l)m$. Nel monoide $(\mathbb{Z}, \cdot, 1)$ si possono anche moltiplicare: $a \equiv a' \pmod m$, $b \equiv b' \pmod m$ implicano $ab \equiv a'b' \pmod m$, poiché da $a = a' + km$, $b = b' + lm$ si ricava $ab = a'b' + (a'l + b'k + klm)m$. Le congruenze modulo m in $(\mathbb{Z}, +, 0)$ e in $(\mathbb{Z}, \cdot, 1)$ illustrano la nozione generale che ora definiamo.

Definizione 1.4. Sia $(M, \cdot, 1)$ un monoide. Una *congruenza* (o *relazione di congruenza*) $\equiv$ in M è una relazione di equivalenza in M tale che, per ogni a, a', b, b' con $a \equiv a'$ e $b \equiv b'$, si abbia $ab \equiv a'b'$. (In altre parole, le congruenze sono relazioni di equivalenza che si possono moltiplicare.)

Sia $\equiv$ una congruenza nel monoide M e consideriamo l'insieme quoziente $\overline{M} = M/\equiv$ di M rispetto a $\equiv$. Ricordiamo che $\overline{M}$ è il sottoinsieme dell'insieme delle parti $\mathcal{P}(M)$ costituito dalle classi di equivalenza $\bar{a} = \{b \in M \mid b \equiv a\}$. Per esempio, in $(\mathbb{Z}, +, 0)$, prendendo la congruenza modulo m definita sopra, si ha $\bar{a} = \{a + km \mid k \in \mathbb{Z}\}$. Poiché le congruenze si possono moltiplicare, nel caso generale da $\bar{a} = \bar{a}'$ e $\bar{b} = \bar{b}'$ segue $\overline{ab} = \overline{a'b'}$. Pertanto

$$(\bar{a}, \bar{b}) \longmapsto \overline{ab}$$

è un'applicazione ben definita di $\overline{M} \times \overline{M}$ in $\overline{M}$, ossia una composizione binaria in $\overline{M}$. La indichiamo ancora con un punto. Mostriamo che $(\overline{M}, \cdot, \bar{1})$ è un monoide. Anzitutto $(\bar{a}\bar{b})\bar{c} = \bar{a}(\bar{b}\bar{c})$: il primo membro è $\overline{ab}\,\bar{c} = \overline{(ab)c}$ e il secondo è $\bar{a}\,\overline{bc} = \overline{a(bc)}$, dunque l'uguaglianza segue dall'associatività in M. Inoltre $\bar{a}\bar{1} = \overline{a1} = \bar{a}$ e $\bar{1}\bar{a} = \overline{1a} = \bar{a}$, quindi $\bar{1}$ è un'unità. Il monoide $(\overline{M}, \cdot, \bar{1})$ si chiama *monoide quoziente di M rispetto alla congruenza $\equiv$*.

Nel caso particolare $M = (\mathbb{Z}, +, 0)$, con la congruenza modulo m, $m > 0$, ogni $a \in \mathbb{Z}$ si può scrivere $a = qm + r$, con $0 \le r < m$, e quindi $a \equiv r \pmod m$. Se $0 \le r_i < m$ per $i = 1, 2$, da $r_1 \equiv r_2 \pmod m$ segue $r_1 = r_2$. Il monoide quoziente, che indicheremo con $\mathbb{Z}/\mathbb{Z}m$ (caso particolare di una notazione generale che introdurremo fra poco), consiste dunque di m elementi:

$$\begin{aligned}
\bar{0} &= \{0, \pm m, \pm 2m, \pm 3m, \ldots\},\\
\bar{1} &= \{1, 1 \pm m, 1 \pm 2m, 1 \pm 3m, \ldots\},\\
&\vdots\\
\overline{m-1} &= \{m-1, m-1 \pm m, m-1 \pm 2m, \ldots\}.
\end{aligned}$$

Nel caso moltiplicativo $M = (\mathbb{Z}, \cdot, 1)$ il monoide $(\mathbb{Z}/\mathbb{Z}m, \cdot, \bar{1})$ ha lo stesso insieme soggiacente.

Se $M = G$ è un gruppo e $\equiv$ è una congruenza su G, si può dire molto di più. Anzitutto il monoide quoziente $(\overline{G}, \cdot, \bar{1})$ è un gruppo, poiché $\bar{a}\,\overline{a^{-1}} = \bar{1} = \overline{a^{-1}}\,\bar{a}$. Dunque ogni $\bar{a}$ è invertibile, con inverso $\overline{a^{-1}}$. Inoltre possiamo determinare tutte le congruenze su un gruppo o, più precisamente, ridurre questo problema a quello di determinare certi sottogruppi, specificati dalla definizione seguente.

[5] È interessante leggere la discussione delle congruenze fra interi all'inizio del grande classico della teoria dei numeri, le *Disquisitiones Arithmeticae* di Carl Friedrich Gauss. L'opera, pubblicata nel 1801, fu scritta quando Gauss aveva diciannove anni. Traduzione inglese di A. A. Clarke, Yale University Press, New Haven, 1966.

Definizione 1.5. Un sottogruppo K del gruppo G si dice *normale* (talvolta *invariante*, e nella letteratura più antica *autoconiugato*) se

$$g^{-1}kg \in K$$

per ogni $g \in G$ e $k \in K$.

Fra le congruenze sul gruppo G e i suoi sottogruppi normali sussiste la seguente relazione fondamentale.

Teorema 1.6. *Siano G un gruppo e $\equiv$ una congruenza su G. La classe di congruenza $K = \bar{1}$ dell'unità è un sottogruppo normale di G e per ogni $g \in G$ si ha $\bar{g} = Kg = gK$, la classe laterale destra o sinistra di g rispetto a K. Viceversa, per ogni sottogruppo normale K di G, la relazione definita da*

$$a \equiv b \pmod{K} \quad \text{se} \quad a^{-1}b \in K$$

è una congruenza in G, le cui classi sono le classi laterali sinistre (o destre) gK.

Dimostrazione. Sia anzitutto $\equiv$ una congruenza su G e sia $K = \bar{1}$. Se $k_1, k_2 \in K$, allora $k_1k_2 \in K$, poiché $\overline{k_1k_2} = \bar{k}_1\bar{k}_2 = \bar{1}\bar{1} = \bar{1}$. Inoltre $1 \in K$ e $k_1^{-1} \in K$, poiché, come si è visto, $\overline{k_1^{-1}} = \bar{k}_1^{-1} = \bar{1}^{-1} = \bar{1}$. Dunque K è un sottogruppo di G. Sia $g \in G$ e consideriamone la classe $\bar{g}$. Se $a \in \bar{g}$, allora $g^{-1}a$ e ag^{-1} appartengono a K: infatti $\overline{g^{-1}a} = \overline{g^{-1}}\,\bar{a} = \bar{g}^{-1}\bar{g} = \bar{1} = K$, e analogamente per ag^{-1}. Pertanto $a \in Kg$ e $a \in gK$. Viceversa, se $a \in Kg$, allora $a = kg$ con $k \in K$ e $\bar{a} = \overline{kg} = \bar{1}\bar{g} = \bar{g}$, ossia $a \equiv g$. Lo stesso accade se $a \in gK$. Così

$$\bar{g} = gK = Kg, \qquad g \in G. \tag{25}$$

Ne segue che K è normale. Lo si può vedere direttamente o, meglio, osservando che per un sottogruppo K la condizione $gK = Kg$ per ogni g equivale alla normalità. Se vale questa condizione, per $g \in G$, $k \in K$ si ha $kg \in gK$, ossia $kg = gk'$ con $k' \in K$. Quindi $g^{-1}kg \in K$ e K è normale. Reciprocamente, se K è normale, ripercorrendo i passaggi in senso inverso si trova $kg \in gK$ per $k \in K$, $g \in G$, dunque $Kg \subseteq gK$. Sostituendo g^{-1} a g otteniamo $Kg^{-1} \subseteq g^{-1}K$, da cui $gK \subseteq Kg$. Pertanto $Kg = gK$ per ogni $g \in G$.

Viceversa, sia K un sottogruppo normale di G e poniamo $a \equiv b \pmod{K}$ quando $a^{-1}b \in K$. Ciò equivale a dire che $b \in aK$, o che b appartiene all'orbita di a rispetto al gruppo di trasformazioni $K_R(G)$. Nel paragrafo precedente abbiamo visto che la relazione così definita è di equivalenza per qualunque sottogruppo K di G. Mostriamo che la normalità di K permette di moltiplicare le equivalenze, così che si ha una congruenza. Siano $a \equiv g \pmod{K}$, $b \equiv h \pmod{K}$. Allora $a = gk_1$, $b = hk_2$ con $k_i \in K$ e, poiché $Kh = hK$, si ha $k_1h = hk_3$, $k_3 \in K$. Pertanto $ab = gk_1hk_2 = ghk_3k_2$, ossia $ab \equiv gh \pmod{K}$. Per questa congruenza $\bar{1} = \{k \mid 1^{-1}k \in K\} = K$ e, per ogni g, $\bar{g} = \{a \mid g^{-1}a \in K\} = gK$. La verifica è conclusa. □

D'ora in poi scriveremo G/K per $\overline{G} = G/\equiv \pmod{K}$ e lo chiameremo *gruppo fattore* (o *gruppo quoziente*) di G rispetto al sottogruppo normale K. Per definizione il prodotto in G/K è

$$(gK)(hK) = ghK, \tag{26}$$

l'unità è $K = 1K$ e l'inverso di gK è $g^{-1}K$.

Ogni gruppo $G \neq 1$ ha due sottogruppi normali, G e 1. Si dice *semplice* se questi sono i suoi soli sottogruppi normali. Equivalentemente, G è semplice se le sole congruenze su G sono quelle banali: l'uguaglianza e la relazione in cui due elementi qualsiasi sono equivalenti. È chiaro dalla definizione che ogni sottogruppo di un gruppo abeliano è normale. Se ne deduce facilmente che i soli gruppi abeliani semplici sono i gruppi ciclici di ordine primo. Lasciamo la dimostrazione al lettore. Osserviamo anche che, se C è il centro di G, ogni sottogruppo di C è normale in G.

Si possono anche considerare i gruppi fattore dal punto di vista della moltiplicazione dei sottoinsiemi. Se A, B sono sottoinsiemi di un gruppo G (o, analogamente, di un monoide), si pone

$$AB = \{ab \mid a \in A,\ b \in B\}.$$

Con questo prodotto e con $1 = \{1\}$, i sottoinsiemi non vuoti di G formano un monoide: $(AB)C$ è l'insieme degli elementi $(ab)c$ e $A(BC)$ quello degli $a(bc)$, per $a \in A, b \in B, c \in C$. L'associatività segue dunque da quella in G. Inoltre $1A = A = A1$. Un sottoinsieme H di G è un sottogruppo se e solo se (1) $H^2 \subseteq H$, (2) $1 \in H$, (3) $H^{-1} = \{h^{-1} \mid h \in H\} \subseteq H$; le condizioni (1) e (2) insieme implicano $H^2 = H$. La classe Hg (rispettivamente gH) è il prodotto di H e $\{g\}$ (rispettivamente di $\{g\}$ e H). Un sottogruppo K è normale se e solo se vale una qualsiasi delle condizioni equivalenti $g^{-1}Kg \subseteq K$, $Kg = gK$, $g^{-1}Kg = K$ per ogni $g \in G$. In tal caso il prodotto tra insiemi appena definito dà $(gK)(hK) = g(Kh)K = g(hK)K = ghK^2 = ghK$. Il prodotto in G/K definito dalla (26) coincide quindi col prodotto dei sottoinsiemi gK e hK.

Esercizi

1. Determinare le tavole di addizione di $(\mathbb{Z}/\mathbb{Z}3, +)$ e $(\mathbb{Z}/\mathbb{Z}6, +)$. Determinare tutti i sottogruppi di $(\mathbb{Z}/\mathbb{Z}6, +)$.

2. Determinare la tavola di moltiplicazione di $(\mathbb{Z}/\mathbb{Z}6, \cdot)$.

3. Sia G il gruppo delle coppie di numeri reali (a, b), $a \neq 0$, col prodotto $(a, b)(c, d) = (ac, ad + b)$ (esercizio 4, p. 36 dell'originale). Verificare che $K = \{(1, b) \mid b \in \mathbb{R}\}$ è un sottogruppo normale di G. Mostrare che $G/K \cong (\mathbb{R}^*, \cdot, 1)$, gruppo moltiplicativo dei reali non nulli.

4. Dimostrare che ogni sottogruppo di indice due è normale. Dedurne che A_n è normale in S_n.

5. Verificare che l'intersezione di una famiglia qualsiasi di sottogruppi normali è normale. Se H, K sono sottogruppi normali, dimostrare che HK è un sottogruppo normale.

6. Siano G_1, G_2 gruppi semplici. Dimostrare che ogni sottogruppo normale di $G = G_1 \times G_2$, distinto da G e da 1, è isomorfo a G_1 oppure a G_2.

7. Sia $\equiv$ una relazione di equivalenza su un monoide M. Mostrare che è una congruenza se e solo se il sottoinsieme di $M \times M$ che la definisce (p. 10 dell'originale) è un sottomonoide di $M \times M$.

8. Sia $\{\equiv_i\}$ una famiglia di congruenze su M. Definirne l'intersezione come intersezione dei corrispondenti sottoinsiemi di $M \times M$. Verificare che si ottiene una congruenza su M.

9. Siano G_1, G_2 sottogruppi del gruppo G e $\alpha : G_1 \times G_2 \to G$ l'applicazione definita da $\alpha(g_1, g_2) = g_1 g_2$. Mostrare che la fibra sopra $g_1 g_2$, ossia $\alpha^{-1}(g_1 g_2)$, è l'insieme delle coppie $(g_1 k, k^{-1} g_2)$ con $k \in K = G_1 \cap G_2$. Dedurne che tutte le fibre sopra l'immagine hanno la stessa cardinalità, quella di K. Se G_1, G_2 sono finiti, ricavarne

$$|G_1 G_2| = \frac{|G_1|\,|G_2|}{|G_1 \cap G_2|}.$$

10. Sia G un gruppo finito e siano A, B suoi sottoinsiemi non vuoti. Dimostrare che $G = AB$ se $|A| + |B| > |G|$.

11. Sia G un gruppo di ordine $2k$ con k dispari. Dimostrare che contiene un sottogruppo di indice 2. (Suggerimento: considerare il gruppo di permutazioni G_L delle traslazioni sinistre e applicare l'esercizio 13, p. 36 dell'originale.)

1.9 Omomorfismi

Quando si studiano strutture matematiche quali monoidi, gruppi, spazi vettoriali, spazi topologici e così via, è importante precisare quali applicazioni siano, in un certo senso, naturali nel contesto considerato. Per gli spazi vettoriali sono le applicazioni lineari, per gli spazi topologici le applicazioni continue. Quasi tutti i risultati interessanti dell'algebra lineare riguardano le trasformazioni lineari o, equivalentemente, le matrici. In effetti, ben poco si può dire degli spazi vettoriali senza ricorrere esplicitamente all'idea di trasformazione lineare o di matrice.[6]

Le applicazioni naturali per i monoidi (e per i gruppi) si chiamano *omomorfismi*. Si ottengono semplicemente sopprimendo la richiesta di biiettività nella definizione di isomorfismo. Il concetto di omomorfismo si affermò piuttosto tardi nella teoria dei gruppi e divenne uno strumento importante per il loro studio soltanto in tempi relativamente recenti, negli ultimi quaranta o cinquant'anni. Il concetto si applica a tutti i tipi di strutture algebriche. Nel caso dei monoidi la definizione formale è la seguente.

Definizione 1.6. Se M, M' sono monoidi, un'applicazione η di M in M' si dice *omomorfismo* se

$$\eta(ab) = \eta(a)\eta(b), \qquad \eta(1) = 1', \qquad a, b \in M.$$

Se M' è un gruppo, la seconda condizione è superflua. Infatti, se vale la prima, $\eta(1) = \eta(1^2) = \eta(1)^2$, e moltiplicando per $\eta(1)^{-1}$ si ottiene $1' = \eta(1)$. Abbiamo già incontrato diversi esempi di omomorfismi che possono non essere isomorfismi. Uno è l'applicazione

$$\eta_a : n \longmapsto a^n$$

del gruppo additivo degli interi in un gruppo qualsiasi G, determinata da un elemento fissato $a \in G$. Poiché $\eta_a(n+m) = a^{n+m} = a^n a^m = \eta_a(n)\eta_a(m)$, questo è un omomorfismo di $(\mathbb{Z}, +, 0)$ in G. Un altro esempio già incontrato è l'applicazione

$$\pi \longmapsto \operatorname{sgn} \pi$$

del gruppo simmetrico S_n nel gruppo moltiplicativo $\{1, -1\}$: che sia un omomorfismo segue dalla (22). Ecco altri esempi di omomorfismi, insieme a un falso esempio.

Esempi

1. Siano M, M' monoidi. Mandando ogni $a \in M$ nell'unità $1'$ di M' si ottiene un omomorfismo di M in M'.

2. Sia $M = (\mathbb{Z}, \cdot, 1)$ il monoide moltiplicativo degli interi e mandiamo ogni $a \in M$ in 0. Vale $\eta(ab) = \eta(a)\eta(b)$, ma l'applicazione non è un omomorfismo, poiché $1 \mapsto 0 \neq 1$.

3. Siano $G = (\mathbb{R}, +, 0)$ e $G' = (\mathbb{C}^*, \cdot, 1)$ il gruppo moltiplicativo dei numeri complessi non nulli. L'applicazione $\eta : \theta \mapsto e^{i\theta}$ è un omomorfismo di G in G'.

4. Sia G il gruppo delle coppie (a, b), $a \neq 0$, dell'esercizio 4, p. 36 dell'originale. L'applicazione $(a, b) \mapsto a$ di G in $G' = (\mathbb{R}^*, \cdot, 1)$ è un omomorfismo.

[6] Forse il risultato più profondo dell'algebra lineare che faccia a meno delle trasformazioni lineari è il teorema dell'invarianza della dimensione: due basi qualsiasi hanno la stessa cardinalità.

5. Sia G un gruppo di trasformazioni di S e sia T un sottoinsieme di S stabilizzato da G, nel senso che $\alpha(T) \subseteq T$ per ogni $\alpha \in G$. Sia $\alpha|_T$ la restrizione di α a T. Allora $\alpha \mapsto \alpha|_T$ è un omomorfismo di G in $\operatorname{Sym} T$, detto *omomorfismo di restrizione*.

Sottolineiamo che, come negli esempi precedenti, un omomorfismo η può non essere né suriettivo né iniettivo. Se risulta suriettivo, lo chiamiamo *epimorfismo*; se è iniettivo, *monomorfismo*. Naturalmente, se è biiettivo, è un isomorfismo.

Se η è un omomorfismo del monoide M nel monoide M', si vede per induzione che $\eta(a^k) = \eta(a)^k$ per $a \in M$, $k \in \mathbb{N}$. Se a è invertibile, applicando η ad $aa^{-1} = 1 = a^{-1}a$ si trova $\eta(a)\eta(a^{-1}) = 1' = \eta(a^{-1})\eta(a)$. Dunque $a' = \eta(a)$ è invertibile in M' e $\eta(a^{-1}) = \eta(a)^{-1}$. Ne segue $\eta(a^k) = \eta(a)^k$ per ogni $k \in \mathbb{Z}$. Ci occorrerà abbastanza spesso un altro risultato da giustificarne l'enunciato come teorema.

Teorema 1.7. *Siano η, ζ omomorfismi di un monoide M (o di un gruppo G) in un monoide M' e sia S un insieme di generatori di M (del gruppo G). Se $\eta(s) = \zeta(s)$ per ogni $s \in S$, allora $\eta = \zeta$.*

Dimostrazione. Consideriamo anzitutto i monoidi e poniamo

$$M_1 = \{a \in M \mid \eta(a) = \zeta(a)\}.$$

Si ha $1 \in M_1$, poiché $\eta(1) = 1' = \zeta(1)$, e $M_1 \supseteq S$. Inoltre, se $a, b \in M_1$, allora $ab \in M_1$, poiché $\eta(ab) = \eta(a)\eta(b) = \zeta(a)\zeta(b) = \zeta(ab)$. Dunque M_1 è un sottomonoide e, contenendo un insieme di generatori, coincide con M. Pertanto $\eta(a) = \zeta(a)$ per ogni a, ossia $\eta = \zeta$.

Nel caso di un gruppo G la dimostrazione è analoga. L'argomento precedente mostra che $G_1 = \{a \in G \mid \eta(a) = \zeta(a)\}$ è un sottomonoide. Se $a \in G_1$, allora $\eta(a^{-1}) = \eta(a)^{-1} = \zeta(a)^{-1} = \zeta(a^{-1})$. Dunque $a^{-1} \in G_1$ e G_1 è un sottogruppo. Poiché contiene un insieme di generatori di G come gruppo, $G_1 = G$. □

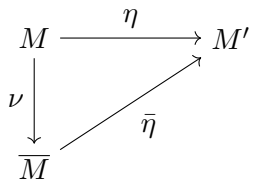
Un omomorfismo di M in sé si chiama *endomorfismo*; un isomorfismo di M su sé stesso si chiama *automorfismo* di M. L'applicazione identica è un automorfismo. Applicando il teorema 1.7 a un endomorfismo η e a $\zeta = 1$ vediamo che, se un endomorfismo di un monoide o di un gruppo coincide con l'identità su un insieme di generatori, allora $\eta = 1$. Osserviamo inoltre che l'insieme degli elementi fissi di η ($\eta(a) = a$) è un sottomonoide se M è un monoide e un sottogruppo se $M = G$ è un gruppo. Lo si vede dalla dimostrazione del teorema 1.7.

Siano $\eta : M \to M'$ e $\zeta : M' \to M''$ omomorfismi di monoidi. Per $a, b \in M$ si ha $\zeta\eta(ab) = \zeta(\eta(ab)) = \zeta(\eta(a)\eta(b)) = (\zeta\eta(a))(\zeta\eta(b))$. Inoltre $\zeta\eta(1) = \zeta(1') = 1''$, l'unità di M''. Quindi $\zeta\eta : M \to M''$ è un omomorfismo. Se η è biiettivo, abbiamo già visto che η^{-1} è un isomorfismo di M' su M. Poiché l'identità è un automorfismo, l'insieme $\operatorname{Aut} M$ degli automorfismi di un monoide è un gruppo di trasformazioni del monoide, detto *gruppo degli automorfismi di M*. L'insieme più grande $\operatorname{End} M$ degli endomorfismi è un monoide di trasformazioni, il *monoide degli endomorfismi di M*.

Sia M un monoide, $\equiv$ una congruenza su M e $\overline{M}$ il monoide quoziente da essa determinato. L'applicazione naturale $\nu : a \mapsto \bar{a}$ (la classe di congruenza di a) è un omomorfismo, poiché $\nu(1) = \bar{1}$ è l'unità di $\overline{M}$ e $\nu(ab) = \overline{ab} = \bar{a}\bar{b} = \nu(a)\nu(b)$ per definizione del prodotto in $\overline{M}$. Ricaviamo ora il risultato principale sugli omomorfismi di monoidi e di gruppi.

Teorema fondamentale degli omomorfismi di monoidi e di gruppi. *Sia η un omomorfismo di un monoide M in un monoide M'. L'immagine $\eta(M)$ è un sottomonoide di M' e, se M è un gruppo, è un sottogruppo di M'. La relazione di equivalenza E_η determinata da η ($aE_\eta b$ significa $\eta(a) = \eta(b)$) è una congruenza in M. Esiste un unico omomorfismo $\bar{\eta}$ del*

monoide quoziente $\overline{M} = M/E_\eta$ *in* M' *che renda commutativo il diagramma*



L'applicazione ν *è un epimorfismo e* $\bar{\eta}$ *un monomorfismo. Nel caso dei gruppi,* $\bar{1} = K = \eta^{-1}(1')$ *è un sottogruppo normale di* M, $\overline{M} = M/K$, ν *è* $a \mapsto aK$ *e* $\bar{\eta}$ *è* $aK \mapsto \eta(a)$.

Dimostrazione. Come spesso accade quando si espongono i fondamenti, la dimostrazione è poco più lunga dell'enunciato e consiste nel verificare direttamente le sue varie affermazioni. Sia $\eta : M \to M'$ un omomorfismo di monoidi. Allora $1' = \eta(1) \in \eta(M)$ e l'uguaglianza $\eta(a)\eta(b) = \eta(ab)$ mostra che $\eta(M)$ è chiuso rispetto al prodotto in M'. Pertanto $\eta(M)$ è un sottomonoide. Se M è un gruppo, $\eta(a)$ è invertibile con inverso $\eta(a^{-1})$, sicché $\eta(M)$ è un sottogruppo di M'. Consideriamo ora la relazione di equivalenza E_η in M. Siano $a_1 E_\eta a_2$ e $b_1 E_\eta b_2$, cioè $\eta(a_1) = \eta(a_2)$ e $\eta(b_1) = \eta(b_2)$. Allora $\eta(a_1 b_1) = \eta(a_1)\eta(b_1) = \eta(a_2)\eta(b_2) = \eta(a_2 b_2)$, e quindi $a_1 b_1 E_\eta a_2 b_2$. Dunque E_η è una congruenza. Per i risultati sulle applicazioni fra insiemi (§ 0.3), esiste un'unica applicazione indotta $\bar{\eta}$ di $\overline{M} = M/E_\eta$ in M' tale che $\bar{\eta}\nu = \eta$. Abbiamo visto che ν è un omomorfismo. Per il caso dei monoidi resta da verificare che anche $\bar{\eta}$ lo sia. Si ha $\bar{\eta}(\bar{a}) = \eta(a)$, dunque $\bar{\eta}(\bar{a}\bar{b}) = \bar{\eta}(\overline{ab}) = \eta(ab) = \eta(a)\eta(b) = \bar{\eta}(\bar{a})\bar{\eta}(\bar{b})$, e $\bar{\eta}(\bar{1}) = \eta(1) = 1'$, come richiesto. Nel § 0.3 si è visto che ν è suriettiva e $\bar{\eta}$ iniettiva. Si tratta dunque, rispettivamente, di un epimorfismo di M e di un monomorfismo di $\overline{M}$. Siano ora M, M' gruppi. Poiché E_η è una congruenza nel gruppo M, la classe K di 1 è un sottogruppo normale di M e la classe di a è $Ka = aK$ (§ 1.8). Per definizione la classe di 1 è

$$K = \{a \in M \mid \eta(a) = \eta(1) = 1'\},$$

ossia $K = \eta^{-1}(1')$. Il resto segue dal teorema 1.6. □

Abbiamo ricavato i risultati sui gruppi da quelli sui monoidi. Per questi ultimi sono essenziali le nozioni di congruenza e di monoide quoziente definito da una congruenza. I risultati fondamentali sugli omomorfismi di gruppi si possono però ottenere anche direttamente senza ricorrere alle congruenze. Seguiamo ora questa strada per chiarire il caso, particolarmente importante, dei gruppi.

Ripartiamo da un omomorfismo η del gruppo G nel gruppo G'. È immediato che l'immagine $\operatorname{im}\eta$ è un sottogruppo di G'. Consideriamo poi $K = \eta^{-1}(1')$, analogo al sottospazio nullo di un'applicazione lineare da uno spazio vettoriale in un altro. Una verifica diretta mostra che K è un sottogruppo normale di G. Lo chiamiamo *nucleo* di η e lo indichiamo anche con $\ker\eta$. Anzitutto η è iniettivo se e solo se $\ker\eta = 1$. Infatti, se $\ker\eta \neq 1$, esiste $b \neq 1$ in G tale che $\eta(b) = 1' = \eta(1)$. D'altra parte, se η non è iniettivo, esistono $a \neq b$ in G con $\eta(a) = \eta(b)$. Allora $a^{-1}b \neq 1$ e $\eta(a^{-1}b) = \eta(a)^{-1}\eta(b) = 1'$, cosicché $\ker\eta \neq 1$.

Sia L un sottogruppo normale di G contenuto in K. Possiamo formare il gruppo fattore $\overline{G} = G/L$, costituito dalle classi $aL = La$, $a \in G$, con prodotto $(aL)(bL) = abL$ e unità $\bar{1} = L$ (si veda l'ultimo capoverso a p. 56 dell'originale). Per definizione $\nu : a \mapsto aL$ è un omomorfismo di G su $\overline{G} = G/L$. Supponiamo $aL = bL$. Allora $b = al$ con $l \in L$, e $\eta(b) = \eta(a)\eta(l) = \eta(a)1' = \eta(a)$ poiché $L \subseteq \ker\eta$. Si ha perciò un'applicazione ben definita $\bar{\eta} : aL \mapsto \eta(a)$ di G/L in G'. Poiché $\bar{\eta}((aL)(bL)) = \bar{\eta}(abL) = \eta(ab) = \eta(a)\eta(b) = \bar{\eta}(aL)\bar{\eta}(bL)$, essa è un omomorfismo, detto *omomorfismo di* $\overline{G} = G/L$ *indotto da* η. Per $a \in G$ si ha $\bar{\eta}\nu(a) = \bar{\eta}(aL) = \eta(a)$. Quindi $\eta = \bar{\eta}\nu$ e si ha un diagramma commutativo come quello precedente.

Evidentemente $\operatorname{im}\bar{\eta} = \operatorname{im}\eta$. Qual è il nucleo di $\bar{\eta}$? Per definizione è l'insieme delle classi aL tali che $\bar{\eta}(aL) = 1'$. Poiché $\bar{\eta}(aL) = \eta(a)$, la condizione è $\eta(a) = 1'$. Pertanto $\ker\bar{\eta} = \{aL \mid a \in \ker\eta\} = (\ker\eta)/L$ (chiaramente L è un sottogruppo normale di K). Un omomorfismo è iniettivo se e solo se il suo nucleo è 1; dunque $\bar{\eta}$ è iniettivo se e solo se $L = \ker\eta$.

Questi risultati vanno oltre l'enunciato del teorema fondamentale, poiché sostituiscono a $K = \ker\eta$ un sottogruppo normale qualsiasi L di G contenuto in K. Supponiamo ora $K = L$ e η suriettivo. Allora $\bar{\eta}$ da $\overline{G} = G/K$ in G' è suriettivo e iniettivo, quindi è un isomorfismo. Abbiamo così il seguente

Corollario. *Se η è un epimorfismo del gruppo G sul gruppo G' con nucleo K, l'applicazione indotta $\bar{\eta} : aK \mapsto \eta(a)$ è un isomorfismo. Ogni immagine omomorfa di un gruppo G è dunque isomorfa a un gruppo fattore G/K rispetto a un sottogruppo normale K.*

Esercizi

1. Siano $G = (\mathbb{Q}, +, 0)$ e $K = \mathbb{Z}$. Mostrare che G/K è isomorfo al gruppo moltiplicativo dei numeri complessi della forma $e^{2\pi i\theta}$, $\theta \in \mathbb{Q}$.

2. Sia G l'insieme delle terne di interi (k, l, m). Si ponga $(k_1, l_1, m_1)(k_2, l_2, m_2) = (k_1 + k_2 + l_1m_2, l_1 + l_2, m_1 + m_2)$. Verificare che si ottiene un gruppo con unità $(0, 0, 0)$. Mostrare che $C = \{(k, 0, 0) \mid k \in \mathbb{Z}\}$ è un sottogruppo normale e che G/C è isomorfo al gruppo $\mathbb{Z}^{(2)} = \{(l, m) \mid l, m \in \mathbb{Z}\}$ con l'addizione ordinaria.

3. Mostrare che $a \mapsto a^{-1}$ è un automorfismo del gruppo G se e solo se G è abeliano e che, se G è abeliano, $a \mapsto a^k$ è un endomorfismo per ogni $k \in \mathbb{Z}$.

4. Determinare $\operatorname{Aut} G$ se G è (i) un gruppo ciclico infinito, (ii) un gruppo ciclico di ordine sei, (iii) un qualsiasi gruppo ciclico finito.

5. Determinare $\operatorname{Aut} S_3$.

6. Per $a \in G$ si definisca l'*automorfismo interno* (o *coniugio*) I_a come $x \mapsto axa^{-1}$. Verificare che I_a è un automorfismo. Mostrare che $a \mapsto I_a$ è un omomorfismo di G in $\operatorname{Aut} G$ con nucleo il centro C di G. Dedurne che $\operatorname{Inn} G = \{I_a \mid a \in G\}$ è un sottogruppo di $\operatorname{Aut} G$ e che $\operatorname{Inn} G \cong G/C$. Verificare che $\operatorname{Inn} G$ è normale in $\operatorname{Aut} G$. Si chiama $\operatorname{Aut} G/\operatorname{Inn} G$ il *gruppo degli automorfismi esterni*.

7. Sia G_L l'insieme delle traslazioni sinistre a_L, $a \in G$. Mostrare che $G_L \operatorname{Aut} G$ è un gruppo di trasformazioni dell'insieme G e contiene G_R. Si chiama *olomorfo* di G e si indica con $\operatorname{Hol} G$. Se G è finito, mostrare che $|\operatorname{Hol} G| = |G|\,|\operatorname{Aut} G|$.

8. Sia G un gruppo con $\operatorname{Aut} G = 1$. Mostrare che G è abeliano e tutti i suoi elementi soddisfano $x^2 = 1$. Se G è finito, mostrare che $|G| = 1$ oppure 2. (Suggerimento: usare il procedimento di ricerca di una base di uno spazio vettoriale per mostrare che G contiene $a_1, \ldots, a_r$ tali che ogni elemento si scrive in un unico modo come $a_1^{k_1} \cdots a_r^{k_r}$, $k_i = 0, 1$. Mostrare poi che esiste un automorfismo che scambia a_1 e a_2.)

9. Sia α un automorfismo di G che fissa soltanto l'unità ($\alpha(a) = a \Rightarrow a = 1$). Mostrare che $a \mapsto \alpha(a)a^{-1}$ è iniettiva. Dedurne che se G è finito ogni suo elemento è della forma $\alpha(a)a^{-1}$.

10. Siano G, α come nell'esercizio precedente, con G finito e $\alpha^2 = 1$. Mostrare che G è abeliano di ordine dispari.

11. Siano G un gruppo finito e α un suo automorfismo. Si ponga $I = \{g \in G \mid \alpha(g) = g^{-1}\}$. Se $|I| > \frac{3}{4}|G|$, mostrare che G è abeliano. Se $|I| = \frac{3}{4}|G|$, mostrare che G ha un sottogruppo abeliano di indice 2.

1.10 Sottogruppi di un'immagine omomorfa. Due teoremi fondamentali di isomorfismo

Stabiliremo una corrispondenza biunivoca fra i sottogruppi dell'immagine omomorfa $\overline{G}$ di un gruppo G e i sottogruppi di G che contengono il nucleo dell'omomorfismo dato. Poiché ogni immagine omomorfa è isomorfa a un gruppo fattore, possiamo supporre $\overline{G} = G/K$, con K normale in G.

Teorema 1.8. *Sia K un sottogruppo normale di G e sia H un sottogruppo di G contenente K. Allora $\overline{H} = H/K$ è un sottogruppo di $\overline{G} = G/K$, e $H \mapsto \overline{H}$ è una biiezione fra l'insieme dei sottogruppi di G contenenti K e l'insieme dei sottogruppi di $\overline{G}$. Il sottogruppo H ($\supseteq K$) è normale in G se e solo se $\overline{H}$ è normale in $\overline{G}$. In tal caso*

$$\frac{G}{H} \cong \frac{\overline{G}}{\overline{H}} = \frac{G/K}{H/K}.$$

Dimostrazione. Che H/K sia un sottogruppo di G/K segue dalla definizione di G/K. Siano H_1, H_2 due sottogruppi di G che contengono K e supponiamo $H_1/K = H_2/K$. Per ogni $h_1 \in H_1$ si ha $h_1K \in H_2/K$, dunque $h_1K = h_2K$ per qualche $h_2 \in H_2$. Allora $h_2^{-1}h_1 \in K$, ossia $h_1 = h_2k$ con $k \in K$. Poiché $K \subseteq H_2$, ne segue $h_1 \in H_2$. Pertanto $H_1 \subseteq H_2$ e, analogamente, $H_2 \subseteq H_1$. Dunque $H_1 = H_2$ e l'applicazione $H \mapsto H/K$ è iniettiva. Per verificarne la suriettività, sia $\overline{H}$ un sottogruppo di $\overline{G}$, dunque una famiglia di classi laterali. Sia H l'unione in G delle classi di questa famiglia. Se $h_1, h_2 \in H$, allora $h_1K, h_2K \in \overline{H}$ e $h_1h_2K = (h_1K)(h_2K) \in \overline{H}$, dunque $h_1h_2 \in H$. Analogamente $h_1^{-1}K = (h_1K)^{-1} \in \overline{H}$, quindi $h_1^{-1} \in H$. Ne segue che H è un sottogruppo di G. Chiaramente $\overline{H} = H/K$. È evidente che se H è normale in G, allora $\overline{H}$ è normale in $\overline{G}$. Viceversa, se $\overline{H}$ è normale in $\overline{G}$, per $h \in H$ e $g \in G$ si ha $(g^{-1}hg)K = (gK)^{-1}(hK)(gK) = h'K$ per qualche $h' \in H$. Dunque $g^{-1}hg \in H$ e H è normale in G. In questo caso possiamo formare il gruppo fattore $\overline{G}/\overline{H}$ e abbiamo l'omomorfismo naturale $\bar{\nu} : \bar{g} \mapsto \bar{g}\overline{H}$ di $\overline{G}$ su $\overline{G}/\overline{H}$. Abbiamo anche l'omomorfismo naturale $g \mapsto \bar{g}$ di G su $\overline{G}$, e quindi l'omomorfismo composto $g \mapsto \bar{g}\overline{H}$ di G su $\overline{G}/\overline{H}$. Il suo nucleo è l'insieme dei $g \in G$ tali che $\bar{g} \in \overline{H}$, ossia $gK = hK$ per qualche $h \in H$. Si tratta precisamente di H. Per il teorema fondamentale degli omomorfismi, $gH \mapsto \bar{g}\overline{H}$ è dunque un isomorfismo di G/H su $\overline{G}/\overline{H}$. □

È talvolta utile enunciare il teorema 1.8 nella seguente forma, in apparenza un po' più generale.

Teorema 1.8′. *Sia η un epimorfismo di G su G' e sia Λ l'insieme dei sottogruppi H di G che contengono $K = \ker\eta$. L'applicazione $H \mapsto \eta(H)$ stabilisce una corrispondenza biunivoca fra Λ e l'insieme di tutti i sottogruppi di G'. Inoltre H è normale in G se e solo se $\eta(H)$ è normale in G'. In tal caso*

$$gH \longmapsto \eta(g)\eta(H) \tag{27}$$

è un isomorfismo di G/H su $G'/\eta(H)$.

Lo si può dimostrare direttamente come per il teorema 1.8, oppure dedurlo da quel teorema mediante l'isomorfismo $gK \mapsto \eta(g)$ di G/K su G'. Lasciamo i dettagli al lettore.

L'isomorfismo (27) prende spesso il nome di *primo teorema di isomorfismo* per i gruppi. Vi è anche un importante *secondo teorema di isomorfismo*.

Teorema 1.9. *Siano H, K sottogruppi di G, con K normale in G. Allora $HK = \{hk \mid h \in H,\ k \in K\}$ è un sottogruppo di G contenente K, $H \cap K$ è normale in H e l'applicazione*

$$hK \longmapsto h(K \cap H), \qquad h \in H, \tag{28}$$

è un isomorfismo di HK/K su $H/(K \cap H)$.

Dimostrazione. Poiché K è normale, $hK = Kh$ per $h \in H$. Da $HK = \bigcup_{h\in H} hK$ e $KH = \bigcup_{h\in H} Kh$ si ricava $HK = KH$. Allora $(HK)^2 = HKHK = H^2K^2 = HK$. Inoltre $1 \in HK$ e, per $hk \in HK$, si ha $(hk)^{-1} = k^{-1}h^{-1} \in KH = HK$. Quindi HK è un sottogruppo di G. Chiaramente $HK \supseteq 1K = K$ e K è normale in HK. Consideriamo la restrizione $\nu' = \nu|_H$ dell'applicazione $\nu : g \mapsto gK$. La sua immagine è l'insieme delle classi hK, $h \in H$. Poiché ogni classe hkK con $h \in H$, $k \in K$ coincide con hK, si ha $\operatorname{im}\nu' = HK/K$. Il nucleo è l'insieme degli $h \in H$ tali che $hK = K$, l'unità di HK/K. Siccome $hK = K$ se e solo se $h \in K$, segue $\ker\nu' = H \cap K$. Questo è perciò normale in H e, per il teorema fondamentale degli omomorfismi, $h(H \cap K) \mapsto hK$ è un isomorfismo di $H/(H \cap K)$ su HK/K. Il suo inverso è $hK \mapsto h(H \cap K)$, come nella (28). □

Le dimostrazioni di questo paragrafo illustrano l'efficacia del teorema fondamentale. Come ulteriore esempio del suo impiego e di quello della corrispondenza fra sottogruppi del teorema 1.8, ricaveremo rapidamente i risultati sui gruppi ciclici. Tutto seguirà dalla determinazione dei sottogruppi di $(\mathbb{Z}, +, 0)$ e delle loro relazioni di inclusione. Sia K un sottogruppo non nullo di $\mathbb{Z}$. Se $n \in K$, anche $-n \in K$, quindi K contiene interi positivi e perciò un minimo intero positivo k. Sia n un elemento qualsiasi di K. L'algoritmo della divisione in $\mathbb{Z}$ dà $n = qk + r$, con $0 \le r < k$. Chiaramente $qk \in K$ e, poiché $n \in K$, anche $r = n - qk \in K$. Deve dunque essere $r = 0$, poiché k è il minimo intero positivo in K. Ogni elemento di K è così multiplo di k e, viceversa, ogni multiplo di k appartiene a K. Dunque $K = \mathbb{Z}k = \{mk \mid m \in \mathbb{Z}\}$. Per ogni $k \ge 0$, $\mathbb{Z}k$ è chiaramente un sottogruppo, compreso il sottogruppo nullo $\mathbb{Z}0$. Tutti i sottogruppi di $\mathbb{Z}$ sono quindi gli insiemi $\mathbb{Z}k$, $k \in \mathbb{N}$. Se $k, l \in \mathbb{N}$ e $\mathbb{Z}l \supseteq \mathbb{Z}k$, allora $k \in \mathbb{Z}l$, cioè $k = lm$ e $l \mid k$. Il viceversa è chiaro. Quindi

$$\mathbb{Z}l \supseteq \mathbb{Z}k \quad \Longleftrightarrow \quad l \mid k. \tag{29}$$

Se $k = 0$, si ha $\mathbb{Z}/\mathbb{Z}k \cong \mathbb{Z}$. Se $k > 0$, $\mathbb{Z}/\mathbb{Z}k$ è l'insieme delle classi di congruenza modulo k, ossia

$$\bar{0} = \mathbb{Z}k, \quad \bar{1} = \{1 + mk \mid m \in \mathbb{Z}\}, \quad \bar{2} = \{2 + mk \mid m \in \mathbb{Z}\},$$
$$\ldots, \quad \overline{k-1} = \{(k-1) + mk \mid m \in \mathbb{Z}\}.$$

Dunque $\mathbb{Z}/\mathbb{Z}k$ ha ordine k ed è chiaramente ciclico con generatore $\bar{1}$.

Sia ora $G = \langle a \rangle$ un gruppo ciclico con generatore a. Poiché $a^m a^n = a^{m+n}$, abbiamo l'epimorfismo $n \mapsto a^n$ di $(\mathbb{Z}, +, 0)$ su G. Quindi $G \cong \mathbb{Z}/\mathbb{Z}k$ per qualche $k \in \mathbb{N}$. Se $k = 0$, $G \cong \mathbb{Z}$; se $k > 0$, G è finito di ordine k. È dunque chiaro che due gruppi ciclici dello stesso ordine sono isomorfi.

Possiamo determinare anche i sottogruppi di $\mathbb{Z}/\mathbb{Z}k$. Per $k = 0$ ritroviamo $\mathbb{Z}$, di cui abbiamo già determinato i sottogruppi: sono gli $\mathbb{Z}l$ con $l \ge 0$, e $\mathbb{Z}l$ è ciclico con generatore l. Per $k > 0$ dal teorema 1.8 segue che i sottogruppi di $\mathbb{Z}/\mathbb{Z}k$ hanno la forma $\mathbb{Z}l/\mathbb{Z}k$ con $l \ge 0$ e $\mathbb{Z}l \supseteq \mathbb{Z}k$. Allora $l \mid k$, sia dunque $k = lm$. Poiché $(\mathbb{Z}/\mathbb{Z}k)/(\mathbb{Z}l/\mathbb{Z}k) \cong \mathbb{Z}/\mathbb{Z}l$, si ha $|\mathbb{Z}l/\mathbb{Z}k| = |\mathbb{Z}/\mathbb{Z}k|/|\mathbb{Z}/\mathbb{Z}l| = k/l = m$. Per ogni divisore m di k, il gruppo ciclico $\mathbb{Z}/\mathbb{Z}k$ di ordine k ha dunque uno e un solo sottogruppo di ordine m. Tale sottogruppo, $\mathbb{Z}l/\mathbb{Z}k$, è a sua volta ciclico con generatore $l + \mathbb{Z}k$.

Esercizi

1. Mostrare che $\mathbb{Z}l \cap \mathbb{Z}k = \mathbb{Z}[l, k]$ e $\mathbb{Z}l + \mathbb{Z}k = \{a + b \mid a \in \mathbb{Z}l,\ b \in \mathbb{Z}k\} = \mathbb{Z}(l, k)$.

2. Sia $\{H_\alpha\}$ una famiglia di sottogruppi contenenti il sottogruppo normale K. Mostrare che $\bigcap_\alpha (H_\alpha/K) = (\bigcap_\alpha H_\alpha)/K$.

1.11 Oggetti liberi. Generatori e relazioni

Il metodo del paragrafo precedente, che studia i gruppi ciclici come immagini omomorfe del gruppo ciclico «universale» $(\mathbb{Z}, +, 0)$, può essere generalizzato per determinare la struttura dei gruppi abeliani finitamente generati. Svilupperemo questo programma nel capitolo 3. Per ora definiamo questi gruppi abeliani universali finitamente generati, detti *gruppi abeliani liberi*, e consideriamo le nozioni analoghe per i monoidi commutativi, per i monoidi arbitrari e per i gruppi arbitrari.

Costruiamo anzitutto, per ogni intero positivo r, un gruppo abeliano $\mathbb{Z}^{(r)}$ con r generatori $x_1, \dots, x_r$, tale che, per ogni gruppo abeliano G e per ogni scelta di elementi $a_1, \dots, a_r$ in G, esista un unico omomorfismo di $\mathbb{Z}^{(r)}$ in G che mandi

$$x_i \longmapsto a_i, \qquad 1 \leq i \leq r.$$

Sia $\mathbb{Z}^{(r)}$ la potenza diretta r-esima di $\mathbb{Z}$: è l'insieme delle r-uple $(n_1, \dots, n_r)$ di interi, con addizione componente per componente, $(m_i) + (n_i) = (m_i + n_i)$, e $0 = (0, \dots, 0)$. Si tratta di un gruppo abeliano. Poniamo

$$x_i = (0, \dots, 0, \overset{i}{1}, 0, \dots, 0), \qquad 1 \leq i \leq r. \tag{30}$$

Allora $(n_1, \dots, n_r) = \sum_{i=1}^r n_i x_i$, dunque gli x_i generano $\mathbb{Z}^{(r)}$. Sia $a_1, \dots, a_r$ una successione di r elementi di un gruppo abeliano qualsiasi G e consideriamo l'applicazione

$$\eta : (n_1, \dots, n_r) \longmapsto a_1^{n_1} a_2^{n_2} \cdots a_r^{n_r}. \tag{31}$$

Poiché gli a_i commutano, si ha

$$(a_1^{m_1} a_2^{m_2} \cdots a_r^{m_r})(a_1^{n_1} a_2^{n_2} \cdots a_r^{n_r}) = a_1^{m_1+n_1} a_2^{m_2+n_2} \cdots a_r^{m_r+n_r},$$

quindi η è un omomorfismo di $\mathbb{Z}^{(r)}$ in G. Inoltre

$$\eta(x_i) = \eta(0, \dots, 0, \overset{i}{1}, 0, \dots, 0) = a_1^0 \cdots a_{i-1}^0 a_i^1 a_{i+1}^0 \cdots a_r^0 = a_i.$$

Poiché gli x_i generano $\mathbb{Z}^{(r)}$, un solo omomorfismo di $\mathbb{Z}^{(r)}$ manda x_i in a_i per $1 \leq i \leq r$ (teorema 1.7). Chiamiamo $\mathbb{Z}^{(r)}$ il *gruppo abeliano libero con r generatori (liberi) x_i*.

Identiche considerazioni si applicano ai monoidi commutativi. Sia $\mathbb{N}^{(r)}$ la potenza diretta r-esima del monoide $(\mathbb{N}, +, 0)$. È un monoide commutativo generato dagli r elementi x_i della (30). Inoltre, come nel caso dei gruppi, se $a_1, \dots, a_r$ sono elementi di un monoide commutativo M, esiste un unico omomorfismo di $\mathbb{N}^{(r)}$ in M tale che $x_i \mapsto a_i$, $1 \leq i \leq r$. Chiamiamo $\mathbb{N}^{(r)}$ il *monoide commutativo libero con r generatori (liberi) x_i*.

Togliamo ora la richiesta di commutatività. Vogliamo costruire prima un monoide, poi un gruppo, generati da r elementi x_i, in modo che, assegnati r elementi a_i di un monoide M (o di un gruppo G), esista un unico omomorfismo dal monoide (o gruppo) costruito che mandi x_i in a_i, $1 \leq i \leq r$.

Cominciamo dai monoidi. Poniamo $X^1 = X = \{x_1, \dots, x_r\}$ e $X^j = X \times \cdots \times X$ (j fattori), per $j = 2, 3, \dots$. Indichiamo con $FS^{(r)}$ l'unione disgiunta degli insiemi $X^1, X^2, \dots$. I suoi

elementi sono le «parole nell'alfabeto X», ossia successioni $(x_{i_1}, x_{i_2}, \ldots, x_{i_m})$, con $x_{i_j} \in X$ e $m = 1, 2, 3, \ldots$. Introduciamo in $FS^{(r)}$ la moltiplicazione per giustapposizione:

$$(x_{i_1}, \ldots, x_{i_m})(x_{j_1}, \ldots, x_{j_n}) = (x_{i_1}, \ldots, x_{i_m}, x_{j_1}, \ldots, x_{j_n}). \tag{32}$$

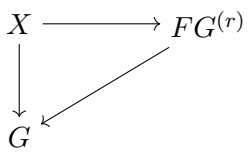
Il prodotto è chiaramente associativo, ma manca un'unità. Possiamo aggiungerne una, che chiameremo 1 (esercizio 5, p. 30 dell'originale), ottenendo un monoide $FM^{(r)}$. Per la (32), $(x_{i_1}, \ldots, x_{i_m}) = x_{i_1} \cdots x_{i_m}$, dunque gli x_i generano $FM^{(r)}$. Siano $a_1, \ldots, a_r$ elementi di un monoide qualsiasi M. Ogni elemento di $FM^{(r)}$ diverso da 1 si scrive in modo unico come $(x_{i_1}, \ldots, x_{i_m})$, quindi

$$1 \longmapsto 1, \qquad (x_{i_1}, \ldots, x_{i_m}) \longmapsto a_{i_1} \cdots a_{i_m}$$

è un'applicazione ben definita. Dalla (32) è chiaro che si tratta di un omomorfismo di $FM^{(r)}$ in M che manda x_i in a_i, $1 \leq i \leq r$. Poiché gli x_i generano $FM^{(r)}$, è l'unico omomorfismo con questa proprietà. Chiamiamo $FM^{(r)}$ il *monoide libero (liberamente) generato dagli r elementi x_i*, o il *monoide delle parole negli x_i*.

Per costruire un gruppo libero osserviamo anzitutto che, in un gruppo, il sottogruppo generato da un sottoinsieme X coincide con il sottomonoide generato dall'unione di X con l'insieme degli inversi dei suoi elementi. Ciò suggerisce di formare $X \cup X'$, dove $X = \{x_1, \ldots, x_r\}$ è l'insieme dato e $X' = \{x'_1, \ldots, x'_r\}$ è un altro insieme disgiunto da X e in corrispondenza biunivoca $x_i \leftrightarrow x'_i$ con esso. Formiamo il monoide libero $FM^{(2r)}$ generato da $X \cup X'$. Siano G un gruppo e $a_1, \ldots, a_r$ una successione di elementi di G. Esiste un unico omomorfismo η di $FM^{(2r)}$ in G che mandi x_i in a_i e x'_i in a_i^{-1}, $1 \leq i \leq r$. Per il teorema fondamentale degli omomorfismi, si ottiene una congruenza E_η su $FM^{(2r)}$ ponendo $aE_\eta b$ quando $\eta(a) = \eta(b)$. Allora $x_i x'_i E_\eta 1$ e $x'_i x_i E_\eta 1$. Consideriamo pertanto l'insieme Γ di tutte le congruenze $\equiv_\alpha$ su $FM^{(2r)}$ per cui $x_i x'_i \equiv_\alpha 1$ e $x'_i x_i \equiv_\alpha 1$, $1 \leq i \leq r$, e formiamone l'intersezione $\equiv$. Per definizione, $a \equiv b$ significa $a \equiv_\alpha b$ per ogni $\equiv_\alpha$. Anche questa è una congruenza (esercizio 8 del § 1.8), sicché possiamo formare il monoide quoziente $FM^{(2r)}/\equiv$, che indicheremo con $FG^{(r)}$. Osserviamo anzitutto che $FG^{(r)}$ è un gruppo generato dalle classi $\bar{x}_i$, $1 \leq i \leq r$. Infatti la classe $\bar{x}_i$ ha inverso $\bar{x}'_i$ in $FG^{(r)}$, e $FG^{(r)}$ è generato come monoide dagli elementi $\bar{x}_i, \bar{x}'_i$. Siano di nuovo G un gruppo e $a_1, \ldots, a_r$ elementi di G. L'unico omomorfismo η di $FM^{(2r)}$ in G che manda x_i in a_i e x'_i in a_i^{-1} dà una congruenza E_η con $x_i x'_i E_\eta 1$ e $x'_i x_i E_\eta 1$. Pertanto $a \equiv b$ in $FM^{(2r)}$ implica $aE_\eta b$, e si ottiene un'applicazione ben definita di $FG^{(r)}$ che manda $\bar{a}$ in $\eta(a)$. È un omomorfismo di $FG^{(r)}$ che manda $\bar{x}_i$ in a_i, $1 \leq i \leq r$; poiché gli $\bar{x}_i$ generano $FG^{(r)}$, è l'unico con questa proprietà.

Riassumendo, dato $X = \{x_1, \ldots, x_r\}$, abbiamo ottenuto un'applicazione $x_i \mapsto \bar{x}_i$ di X in un gruppo $FG^{(r)}$ tale che, per ogni gruppo G e per ogni applicazione $x_i \mapsto a_i$ di X in G, esiste un unico omomorfismo di $FG^{(r)}$ in G che renda commutativo il diagramma



Mostriamo che l'applicazione $x_i \mapsto \bar{x}_i$ è iniettiva. Prendiamo come G nel diagramma il gruppo abeliano libero $\mathbb{Z}^{(r)}$ generato dagli elementi $(0, \ldots, 0, 1, 0, \ldots, 0)$ e scegliamo come freccia verticale l'applicazione $x_i \mapsto (0, \ldots, 0, \overset{i}{1}, 0, \ldots, 0)$. Questa è iniettiva e l'iniettività della composta $\beta\alpha$ di due applicazioni implica quella di α; ne segue l'iniettività di $x_i \mapsto \bar{x}_i$. Identifichiamo infine x_i con la sua immagine $\bar{x}_i$. Possiamo così dire che $FG^{(r)}$ è generato dagli x_i. Inoltre, dati $a_i \in G$, esiste un unico omomorfismo di $FG^{(r)}$ in G tale che $x_i \mapsto a_i$, $1 \leq i \leq r$. Chiamiamo $FG^{(r)}$ il *gruppo libero (liberamente) generato dagli r elementi x_i*.[7]

[7] Un'altra costruzione dei gruppi liberi è data a p. 89 di *Basic Algebra II*.

Un gruppo G si dice *finitamente generato* se contiene un insieme finito di generatori $\{a_i \mid 1 \leq i \leq r\}$. Si ha allora l'omomorfismo η di $FG^{(r)}$ che manda x_i in a_i. Poiché gli a_i generano G, si tratta di un epimorfismo e $G \cong FG^{(r)}/K$, dove K è il nucleo di η. Il sottogruppo normale K si dice *insieme delle relazioni fra i generatori* a_i. Se S è un sottoinsieme di un gruppo, si definisce il *sottogruppo normale generato da* S come l'intersezione di tutti i sottogruppi normali che contengono S. È un sottogruppo normale che contiene S ed è contenuto in ogni sottogruppo normale contenente S. Per $S \subseteq FG^{(r)}$ si dice che G è *definito dalle relazioni* S se $G \cong FG^{(r)}/K$, dove K è il sottogruppo normale generato da S. Se S è finito, G si dice *gruppo finitamente presentato*.

Mostriamo, come esempio, che il gruppo diedrale D_n costituito dalle n rotazioni e dalle n riflessioni che mandano in sé un n-agono regolare (esempio 12, p. 34 dell'originale) è definito dalle relazioni

$$x^n, \quad y^2, \quad xyxy \tag{33}$$

nel gruppo libero generato da x, y. È chiaro che D_n è generato dalla rotazione R di angolo $2\pi/n$ e dalla riflessione S rispetto all'asse x. Valgono le relazioni

$$R^n = 1, \qquad S^2 = 1, \qquad SRS = R^{-1}. \tag{34}$$

Pertanto D_n è un'immagine omomorfa di $FG^{(2)}/K$, ove K è il sottogruppo normale generato dagli elementi (33). Mostriamo che $|FG^{(2)}/K| \leq 2n$, da cui seguirà $D_n \cong FG^{(2)}/K$. Siano $\bar{x} = xK$, $\bar{y} = yK$ in $FG^{(2)}/K$. Poiché $x^n, y^2, xyxy \in K$, abbiamo $\bar{x}^n = 1$, $\bar{y}^2 = 1$, $\bar{x}\bar{y}\bar{x}\bar{y} = 1$. Allora $\bar{y}\bar{x} = \bar{x}^{-1}\bar{y}$, da cui $\bar{y}\bar{x}^k = \bar{x}^{-k}\bar{y}$. Ne segue che il prodotto di due elementi della forma $\bar{x}^k, \bar{x}^k\bar{y}$, $k = 0, 1, \ldots, n-1$, è ancora uno di tali elementi. L'insieme considerato contiene 1 ed è chiuso rispetto agli inversi, dunque è un sottogruppo di $FG^{(2)}/K$. Poiché contiene i generatori $\bar{x}, \bar{y}$, si ha $FG^{(2)}/K = \{\bar{x}^k, \bar{x}^k\bar{y} \mid 0 \leq k \leq n-1\}$. Perciò $|FG^{(2)}/K| \leq 2n$ e $D_n \cong FG^{(2)}/K$.

Esercizi

1. Sia S un sottoinsieme del gruppo G tale che $g^{-1}Sg \subseteq S$ per ogni $g \in G$. Mostrare che il sottogruppo $\langle S \rangle$ generato da S è normale. Per un sottoinsieme qualsiasi T di G sia $S = \bigcup_{g \in G} g^{-1}Tg$. Mostrare che $\langle S \rangle$ è il sottogruppo normale generato da T.

2. Sia G il gruppo definito in $FG^{(3)}$ dalle relazioni $x_2x_1 = x_3x_1x_2$, $x_3x_1 = x_1x_3$, $x_3x_2 = x_2x_3$. Mostrare che G è isomorfo al gruppo definito nell'esercizio 2 del § 1.9.

I tre esercizi seguenti sono tratti da Burnside, *The Theory of Groups of Finite Order*, seconda edizione, 1911 (ristampa Dover, pp. 464–465).

3. Servendosi dei generatori $(12), (13), \ldots, (1n)$ di S_n (esercizio 5, p. 51 dell'originale), mostrare che S_n è definito dalle seguenti relazioni sugli $x_1, \ldots, x_{n-1}$ di $FG^{(n-1)}$:

$$x_i^2, \quad (x_ix_j)^3, \quad (x_ix_jx_ix_k)^2, \qquad i, j, k \text{ distinti.}$$

4. Servendosi dei generatori $(12), (23), \ldots, (n-1\ n)$ di S_n, mostrare che il gruppo è definito da $x_1, \ldots, x_{n-1}$ soggetti alle relazioni

$$x_i^2, \qquad (x_ix_{i+1})^3, \qquad (x_ix_j)^2 \quad (j > i+1).$$

5. Mostrare che A_n può essere definito dalle seguenti relazioni su $x_1, \ldots, x_{n-2}$:

$$x_1^3, \qquad x_i^2 \ (i > 1), \qquad (x_ix_{i+1})^3, \qquad (x_ix_j)^2 \ (j > i+1).$$

1.12 Gruppi che agiscono su insiemi

Storicamente, la teoria dei gruppi si occupò dapprima soltanto di gruppi di trasformazioni. Il concetto di gruppo astratto fu introdotto più tardi per concentrare l'attenzione sulle proprietà dei gruppi di trasformazioni che riguardano soltanto la composizione e non fanno riferimento all'insieme su cui le trasformazioni agiscono. In geometria interessano tuttavia anzitutto i gruppi di trasformazioni, e anche nella teoria astratta conviene spesso tornare a questo punto di vista concreto. Per esempio, i gruppi di trasformazioni forniscono una tecnica di conteggio importante nella teoria dei gruppi finiti. Ne abbiamo visto un esempio nella dimostrazione del teorema di Lagrange. In questo paragrafo e nel prossimo incontreremo altri notevoli risultati ottenuti contando.

È utile disporre di un mezzo per passare dal punto di vista astratto a quello concreto delle trasformazioni. A ciò serve il concetto, che ora definiamo, di gruppo che agisce su un insieme. L'idea è semplice. Si parte da un gruppo astratto G e si considerano le sue diverse «realizzazioni» mediante gruppi di trasformazioni. Dapprima si è tentati di considerare soltanto quelle «fedeli», ossia gli isomorfismi di G su gruppi di trasformazioni. L'esperienza mostra presto che conviene allargare la prospettiva includendo gli omomorfismi di G in gruppi di trasformazioni.

Consideriamo dunque un gruppo G e un omomorfismo T di G in $\operatorname{Sym} S$, gruppo delle trasformazioni biiettive di un insieme S. Se $T(g)$ è la trasformazione corrispondente a $g \in G$, le condizioni imposte a T sono:

1. $T(1) = 1$ ($= 1_S$, l'applicazione identica di S).
2. $T(g_1 g_2) = T(g_1)T(g_2)$ per $g_i \in G$.

La prima può essere omessa se, come stiamo supponendo, ogni $T(g)$ è biiettiva. D'altra parte, se si conserva la condizione 1, l'ipotesi che $T(g)$ sia biiettiva è superflua. Se infatti T è un'applicazione del gruppo G nel monoide $M(S)$ delle trasformazioni di S che soddisfa entrambe le condizioni, è un omomorfismo di G in $M(S)$. L'immagine di G è perciò un sottogruppo di $M(S)$ ed è contenuta in $\operatorname{Sym} S$. Conviene pensare all'immagine $T(g)x$ di x mediante $T(g)$ semplicemente come a un prodotto gx dell'elemento $g \in G$ per l'elemento $x \in S$. Si ottiene così un'applicazione

$$(g, x) \longmapsto gx \quad (= T(g)x)$$

di $G \times S$ in S. Quali proprietà possiede? Le condizioni 1 e 2 implicano rispettivamente

$$\begin{aligned} &\text{(i)} \quad 1x = x, \qquad x \in S, \\ &\text{(ii)} \quad (g_1 g_2)x = g_1(g_2 x). \end{aligned}$$

Invertiamo ora l'ordine e poniamo la seguente definizione.

Definizione 1.7. Si dice che un gruppo G *agisce* (o *opera*) sull'insieme S se è data un'applicazione $(g, x) \mapsto gx$ da $G \times S$ in S che soddisfa (i) e (ii).

Abbiamo visto che un omomorfismo T di G in $M(S)$ definisce un'azione di G su S ponendo semplicemente $gx = T(g)x$. Viceversa, supponiamo che G agisca su S. Definiamo $T(g)$ come l'applicazione $x \mapsto gx$, $x \in S$. Le (i) e (ii) implicano le condizioni 1 e 2, dunque $T : g \mapsto T(g)$ è un omomorfismo di G in $\operatorname{Sym} S$.

Chiameremo T l'*omomorfismo associato all'azione* e $T(G)$ il *gruppo di trasformazioni associato*. Se T è un monomorfismo, diremo che G agisce *effettivamente* su S. Chiameremo inoltre il nucleo di T il *nucleo dell'azione*. Pertanto G agisce effettivamente se e solo se il nucleo dell'azione è 1.

Esempi

1. Sia $S = G$ l'insieme soggiacente al gruppo G. Per $g \in G$ e $x \in S$ si definisca gx come il prodotto di g e x in G. Le (i) e (ii) sono evidenti. È l'*azione di G su sé stesso per traslazioni sinistre* (o *moltiplicazioni a sinistra*), usata per dimostrare il teorema di Cayley. La dimostrazione di quel teorema consisteva appunto nel verificare che l'azione è effettiva.

2. Definiamo ora l'azione di G su sé stesso per traslazioni destre. Prendiamo ancora $S = G$. Per evitare confusione col prodotto gx del gruppo, indichiamo l'azione di $g \in G$ su $x \in S$ con $g \circ x$ e la definiamo come xg^{-1}. Allora $1 \circ x = x1 = x$ e $(g_1g_2) \circ x = x(g_1g_2)^{-1} = xg_2^{-1}g_1^{-1} = g_1 \circ (g_2 \circ x)$. Si tratta quindi di un'azione di G su sé stesso, detta *azione per traslazioni destre*. È effettiva.

3. Un'altra azione di G su sé stesso è quella *per coniugio*. Indichiamo l'azione di $g \in G$ su $x \in S = G$ con ${}^g x$, definito come gxg^{-1}. Allora ${}^1x = x$ e ${}^{g_1g_2}x = (g_1g_2)x(g_1g_2)^{-1} = g_1(g_2xg_2^{-1})g_1^{-1} = {}^{g_1}({}^{g_2}x)$. Il nucleo è l'insieme dei c tali che ${}^c x = x$ per ogni x, ossia $cxc^{-1} = x$, o $cx = xc$. Dunque il nucleo è il centro C e l'azione è effettiva se e solo se il centro è banale ($C = 1$).

4. Da un'azione di G su S si ottiene per restrizione un'azione su S di qualunque sottogruppo H di G. In particolare si hanno le azioni di H su G per traslazioni sinistre e destre.

5. Sia H un sottogruppo e denotiamo con G/H l'insieme delle classi laterali sinistre xH, $x \in G$. Finora questa notazione era riservata al caso di H normale in G, in cui G/H denotava il gruppo fattore. Chiameremo G/H lo *spazio delle classi laterali (sinistre)* di G rispetto a H. Per $g \in G$ definiamo $g(xH)$ come il prodotto insiemistico di $\{g\}$ e xH, sicché $g(xH) = gxH$. Si ottiene chiaramente un'azione di G su G/H. Il nucleo è l'insieme dei g tali che $gxH = xH$ per ogni $x \in G$, ossia $x^{-1}gx \in H$ per ogni x. Ciò equivale a $g \in xHx^{-1}$ per ogni x, ossia $g \in \bigcap_{x \in G} xHx^{-1}$. Facilmente si vede che quest'intersezione è il massimo sottogruppo normale di G contenuto in H. Quindi l'azione di G su G/H è effettiva se e solo se H non contiene sottogruppi diversi da 1 normali in G.

6. Come nel punto 5 si ottiene un'azione di G sull'insieme delle classi laterali destre Hx ponendo $g \circ (Hx) = (Hx)g^{-1} = Hxg^{-1}$.

7. Supponiamo che G agisca su S e che T sia un sottoinsieme stabilizzato dall'azione, nel senso che $gT \subseteq T$ per ogni $g \in G$. Per restrizione si ottiene un'azione di G su T. Consideriamo, per esempio, il coniugio di G su sé stesso. Se K è un sottogruppo normale di G, allora ${}^g K = K$ per $g \in G$ e il coniugio si restringe a un'azione di G su K.

8. Se G agisce su S, si ha un'azione indotta sull'insieme delle parti $\mathcal{P}(S)$. Per un sottoinsieme non vuoto A poniamo $gA = \{gx \mid x \in A\}$ e per $A = \varnothing$ poniamo $g\varnothing = \varnothing$. Allora $1A = A$ e $(g_1g_2)A = g_1(g_2A)$, quindi si ottiene un'azione di G su $\mathcal{P}(S)$. Chiaramente $|gA| = |A|$. Si hanno dunque azioni indotte anche sugli insiemi di sottoinsiemi di S di una cardinalità fissata.

Per le azioni di un medesimo gruppo G vi è una definizione naturale di equivalenza. Due azioni di G su S e S' sono *equivalenti* se esiste una biiezione $x \mapsto x'$ di S su S' tale che

$$(gx)' = gx', \qquad g \in G, \quad x \in S. \tag{35}$$

Se denotiamo con α l'applicazione $x \mapsto x'$ e con $T(g), T'(g)$ le trasformazioni $x \mapsto gx$ e $x' \mapsto gx'$, la (35) equivale a

$$\alpha T(g) = T'(g)\alpha, \qquad g \in G. \tag{36}$$

In altre parole, per ogni $g \in G$ è commutativo il diagramma

$$\begin{array}{ccc} S & \xrightarrow{T(g)} & S \\ {\scriptstyle\alpha}\downarrow & & \downarrow{\scriptstyle\alpha} \\ S' & \xrightarrow[T'(g)]{} & S' \end{array}$$

Poiché α è biiettiva, possiamo anche scrivere la (36) come

$$T'(g) = \alpha T(g)\alpha^{-1}, \qquad g \in G. \tag{36'}$$

Come esempio consideriamo le due azioni di G su sé stesso per traslazioni sinistre e destre. L'applicazione $x \mapsto x^{-1}$ è un'equivalenza, poiché $(gx)^{-1} = x^{-1}g^{-1} = g \circ x^{-1}$.

La relazione di equivalenza su S definita da un gruppo di trasformazioni di S si estende alle azioni. Se G agisce su S, per $x, y \in S$ poniamo $x \sim_G y$ quando $y = gx$ per qualche $g \in G$. Si tratta evidentemente dell'equivalenza rispetto al gruppo di trasformazioni $T(G)$ definita in precedenza. Otteniamo come prima una partizione di S in orbite, dove la *G-orbita* di x è $Gx = \{gx \mid g \in G\}$. Indichiamo con S/G l'insieme quoziente formato da queste orbite.

Se H è un sottogruppo di G, le H-orbite dell'azione di H su G per traslazioni sinistre (destre) sono le classi laterali destre (sinistre) di H. Agisca ora G su sé stesso per coniugio. L'orbita di $x \in G$ è ${}^Gx = \{gxg^{-1} \mid g \in G\}$, detta *classe di coniugio* di x. Le diverse classi di coniugio formano naturalmente una partizione di G. Osserviamo che ${}^Gx$ consiste di un solo elemento, ${}^Gx = \{x\}$, se e solo se x appartiene al centro. Il centro è dunque l'unione delle classi di coniugio costituite da un solo elemento di G.

Come esempio determiniamo la decomposizione in classi di coniugio per S_n. Abbiamo già osservato (esercizio 4, p. 51 dell'originale) che per $\beta \in S_n$ vale $\beta(i_1 i_2 \dots i_r)\beta^{-1} = (\beta(i_1), \beta(i_2), \dots, \beta(i_r))$. Se α è un prodotto di cicli $\gamma_1, \gamma_2, \dots$ come nella (17), allora $\beta\alpha\beta^{-1} = (\beta\gamma_1\beta^{-1})(\beta\gamma_2\beta^{-1})\cdots$. Per $\alpha = (i_1 \dots i_r)\cdots(l_1 \dots l_u)$ risulta quindi

$$\beta\alpha\beta^{-1} = (\beta(i_1), \dots, \beta(i_r))\cdots(\beta(l_1), \dots, \beta(l_u)). \tag{37}$$

Conviene supporre $r \geq s \geq \cdots \geq u$ e scrivere la decomposizione in cicli disgiunti in modo che ogni numero in $\{1, \dots, n\}$ vi compaia una e una sola volta. In tal modo associamo ad α una successione di interi positivi $(r, s, \dots, u)$ tale che

$$r \geq s \geq \cdots \geq u, \qquad r + s + \cdots + u = n. \tag{38}$$

Chiamiamo questa successione una *partizione di n*. Dalla (37) risulta che due permutazioni sono coniugate se e solo se determinano la stessa partizione. Le classi di coniugio corrispondono quindi biunivocamente alle diverse partizioni di n. Se $p(n)$ denota il numero di partizioni distinte di n, in S_n si hanno dunque $p(n)$ classi di coniugio. La funzione $p(n)$ degli interi positivi è un'interessante funzione aritmetica. I suoi primi valori sono

$$p(2) = 2, \quad p(3) = 3, \quad p(4) = 5, \quad p(5) = 7, \quad p(6) = 11.$$

Se l'azione di G su S ha una sola orbita, cioè $S = Gx$ per qualche $x \in S$ (e perciò per ogni $x \in S$), diciamo che G agisce *transitivamente* su S. Le azioni di G su sé stesso per traslazione sono chiaramente transitive. Più in generale, se H è un sottogruppo, l'azione di G sullo spazio G/H delle classi laterali sinistre è transitiva: per due classi qualsiasi xH, yH si ha $gxH = yH$

prendendo $g = yx^{-1}$. Mostreremo ora che queste sono essenzialmente le sole azioni transitive di un gruppo G. Ci occorre a tale scopo lo *stabilizzatore* $\operatorname{Stab} x$ di $x \in S$, definito come l'insieme dei $g \in G$ tali che $gx = x$. È chiaramente un sottogruppo di G. Per esempio, nell'azione di G su G per coniugio, $\operatorname{Stab} x = C(x)$, il centralizzante di x in G. Se $y = ax$, la condizione $gy = y$ equivale a $gax = ax$, ossia $(a^{-1}ga)x = x$. Pertanto $\operatorname{Stab} x = a^{-1}(\operatorname{Stab} y)a$. Ne segue che, se G agisce transitivamente su S, gli stabilizzatori dei punti di S sono tutti coniugati: $\operatorname{Stab} y = a(\operatorname{Stab} x)a^{-1}$.

Il risultato seguente caratterizza internamente le azioni transitive.

Teorema 1.10. *Supponiamo che G agisca transitivamente su S e sia $H = \operatorname{Stab} x$ per $x \in S$. L'azione di G su S è equivalente all'azione di G sullo spazio delle classi laterali G/H.*

Dimostrazione. Consideriamo $\alpha : g \mapsto gx$ da G in S. È suriettiva per la transitività di G su S, e induce quindi una biiezione $\bar{\alpha}$ dall'insieme quoziente $\overline{G}$ di G definito da α. Ricordiamo che $\overline{G}$ è l'insieme delle classi $\bar{g} = \{a \mid \alpha(a) = \alpha(g)\} = \{a \mid ax = gx\}$. Ora $ax = gx$ equivale a $g^{-1}ax = x$, cioè $g^{-1}a \in \operatorname{Stab} x$. Dunque $\bar{g}$ è la classe laterale $g(\operatorname{Stab} x)$, e abbiamo la biiezione $\bar{\alpha} : g(\operatorname{Stab} x) \mapsto gx$. Per verificare che è un'equivalenza di azioni occorre vedere che, per $g' \in G$, $\bar{\alpha}$ manda $g'(g \operatorname{Stab} x)$ in $g'(gx)$. È chiaro, poiché questi elementi sono rispettivamente $(g'g) \operatorname{Stab} x$ e $(g'g)x$. □

Per i gruppi finiti, una delle conclusioni più importanti del teorema è che se un gruppo finito G agisce transitivamente su S, allora $|S| = [G : \operatorname{Stab} x]$ per ogni $x \in S$. Dunque $|S|$ è finito e divide $|G|$. Più in generale possiamo applicare questo fatto a qualsiasi azione di un gruppo finito G su un insieme finito S. Abbiamo la partizione

$$S = O_1 \cup O_2 \cup \cdots \cup O_r, \tag{39}$$

dove le O_i sono le distinte orbite degli elementi di S sotto l'azione di G. Il gruppo G agisce transitivamente su O_i, quindi per $x_i \in O_i$ si ha $|O_i| = [G : \operatorname{Stab} x_i]$. Si ottiene così il seguente conteggio degli elementi di S:

$$|S| = \sum_i [G : \operatorname{Stab} x_i], \tag{40}$$

dove la somma è estesa a un sistema $\{x_1, \ldots, x_r\}$ di rappresentanti delle orbite. È importante osservare che tutti i termini $[G : \operatorname{Stab} x_i]$ a secondo membro dividono $|G|$. Un'altra osservazione utile, applicabile a ogni gruppo, è

$$\operatorname{Stab}(axa^{-1}) = a(\operatorname{Stab} x)a^{-1}. \tag{41}$$

La verifica è immediata.

Un caso importante della (40) si ottiene facendo agire G su sé stesso per coniugio. La (40) si specializza allora in

$$|G| = \sum_i [G : C(x_i)], \tag{42}$$

dove $C(x_i)$ è il centralizzante di x_i e $\{x_i\}$ è un sistema di rappresentanti delle classi di coniugio di G. La formula prende il nome di *equazione delle classi* del gruppo finito G. Possiamo modificarla leggermente raccogliendo le classi degli x_i per cui $C(x_i) = G$. Sono precisamente gli elementi del centro C di G, e ciascuna delle loro classi contiene un solo elemento. Dunque

$$|G| = |C| + \sum_j [G : C(y_j)], \tag{42'}$$

dove y_j percorre un sistema di rappresentanti delle classi di coniugio con più di un elemento.

Il tipo di conteggio degli elementi di un gruppo finito dato dalle (40) e (42) è uno strumento importante nella teoria dei gruppi finiti. Ne incontreremo alcuni esempi nel prossimo paragrafo, studiando i teoremi di Sylow. Intanto illustriamo il metodo usando l'equazione delle classi per dimostrare quanto segue.

Teorema 1.11. *Ogni gruppo finito G il cui ordine è una potenza positiva di un primo ha centro $C \neq 1$.*

Dimostrazione. Il primo membro della (42′) è divisibile per il primo p, e gli indici a secondo membro sono potenze di p. Poiché $C(y_j) \neq G$, si ha $[G : C(y_j)] > 1$, dunque $[G : C(y_j)]$ è divisibile per p. La (42′) mostra allora che $|C|$ è divisibile per p, e perciò $C \neq 1$. □

Per le azioni transitive è utile distinguere fra *primitività* e *imprimitività*. La distinzione riguarda l'azione indotta sull'insieme delle parti $\mathcal{P}(S)$. Una partizione $\pi(S)$ di S si dice *stabilizzata* dall'azione di G su S se $gA \in \pi(S)$ per ogni $g \in G$ e $A \in \pi(S)$. Due partizioni hanno banalmente questa proprietà: $\pi_1(S) = \{S\}$ e $\pi_2(S)$ formata da tutti i singoletti $\{x\}$, $x \in S$. Diremo che l'azione è *primitiva* se π_1 e π_2 sono le sole partizioni di S stabilizzate da G. La partizione di S nelle orbite di G è stabilizzata da G, poiché $gA = A$ per ogni orbita A e ogni $g \in G$. Se le orbite consistono di singoli punti, G agisce banalmente, nel senso che $gx = x$ per ogni $g \in G$, $x \in S$; se vi è una sola orbita, G è transitivo. Quindi un'azione non banale e intransitiva di G su S è imprimitiva. Il caso interessante è quello in cui G agisce transitivamente su un insieme di più elementi. Si ha allora il criterio seguente.

Teorema 1.12. *Se G agisce transitivamente su S, con $|S| > 1$, l'azione è primitiva se e solo se lo stabilizzatore* $\operatorname{Stab} x$ *di un qualsiasi $x \in S$ è un sottogruppo massimale di G, ossia non esiste H con* $\operatorname{Stab} x \subsetneq H \subsetneq G$.

Dimostrazione. Osserviamo anzitutto che G agisce imprimitivamente su S se e solo se esiste un sottoinsieme proprio A con $|A| \geq 2$ tale che, per ogni $g \in G$, o $gA = A$ oppure $gA \cap A = \varnothing$. Se vale questa condizione, per ogni $g_1, g_2 \in G$ o $g_1A = g_2A$ oppure $g_1A \cap g_2A = \varnothing$. Sia B il complementare in S di $\bigcup_{g \in G} gA$. Allora $g_1B \cap g_2A = \varnothing$ per ogni $g_1, g_2 \in G$, da cui $gB = B$ per ogni $g \in G$. I distinti sottoinsiemi gA, $g \in G$, insieme a B se non vuoto, costituiscono dunque una partizione non banale di S stabilizzata da G. Viceversa, supponiamo che G agisca imprimitivamente su S. Esiste allora una partizione $\pi(S)$ stabilizzata da G che contiene un sottoinsieme proprio A con $|A| \geq 2$. Per ogni $g \in G$, o $gA = A$ o $gA \cap A = \varnothing$.

Supponiamo ora che $\operatorname{Stab} x$ non sia massimale per qualche $x \in S$, e scegliamo un sottogruppo H con $\operatorname{Stab} x \subsetneq H \subsetneq G$. Poiché G agisce transitivamente su S, quest'azione equivale a quella consueta sullo spazio delle classi $G/\operatorname{Stab} x$. Due azioni equivalenti sono entrambe primitive oppure entrambe imprimitive, quindi basta mostrare che G agisce imprimitivamente su $G/\operatorname{Stab} x$. Prendiamo l'insieme A delle classi $h \operatorname{Stab} x$ con $h \in H$. Poiché $\operatorname{Stab} x \subsetneq H \subsetneq G$, si ha $|A| \geq 2$ e A è un sottoinsieme proprio di $G/\operatorname{Stab} x$. Se $h' \in H$, $h'A$ è l'insieme delle classi $h'h \operatorname{Stab} x$ con $h \in H$, dunque $h'A = A$. D'altra parte, se $g \notin H$, si ha $gh_1 \operatorname{Stab} x \neq h_2 \operatorname{Stab} x$ per $h_1, h_2 \in H$. Altrimenti $gh_1k_1 = h_2k_2$ per $h_1, h_2 \in H$, $k_1, k_2 \in \operatorname{Stab} x$, da cui $g = h_2k_2k_1^{-1}h_1^{-1} \in H$, contro l'ipotesi. Dunque gA, insieme delle classi $gh \operatorname{Stab} x$ per $h \in H$, è disgiunto da A se $g \notin H$. Ne segue che G agisce imprimitivamente su $G/\operatorname{Stab} x$, e quindi su S.

Supponiamo viceversa che G agisca transitivamente ma non primitivamente su S. Esiste un sottoinsieme proprio A di S, $|A| \geq 2$, tale che per ogni $g \in G$ o $gA = A$ oppure $gA \cap A = \varnothing$. Prendiamo $x \in A$ e $H = \{h \in G \mid hA = A\}$. Allora H è un sottogruppo di G e contiene $\operatorname{Stab} x$, poiché $gx = x \Rightarrow gA \cap A \neq \varnothing \Rightarrow gA = A$. Siccome $A \neq S$ e G è transitivo su S, esiste $g \in G$ tale che $gx \notin A$. Allora $gA \neq A$ e $g \notin H$, dunque $G \neq H$. Sia ora $y \in A$, $y \neq x$ (esiste perché $|A| \geq 2$). Per qualche $g \in G$ si ha $gx = y$. Allora $y \in gA \cap A$, quindi $gA = A$, mentre $gx \neq x$.

Pertanto $g \in H$ e $g \notin \operatorname{Stab} x$, cosicché $H \neq \operatorname{Stab} x$. Lo stabilizzatore non è dunque massimale in G. La dimostrazione è conclusa. □

Esercizi

1. Sia $\gamma = (12\dots n)$ in S_n. Mostrare che la sua classe di coniugio ha $(n-1)!$ elementi e che il centralizzante è $C(\gamma) = \langle\gamma\rangle$.

2. Determinare i rappresentanti delle classi di coniugio di S_5 e il numero di elementi di ciascuna classe. Dedurne che i soli sottogruppi normali di S_5 sono $1, A_5, S_5$.

3. Sia $(n_1, \dots, n_q)$ la partizione associata a una classe di coniugio, dove $n_1 = \cdots = n_{q_1} > n_{q_1+1} = \cdots = n_{q_1+q_2} > n_{q_1+q_2+1} = \cdots$. Mostrare che il numero degli elementi della classe è
$$\frac{n!}{\prod_i q_i! \prod_j n_j}.$$

4. Se un gruppo finito G ha un sottogruppo H di indice n, mostrare che H contiene un sottogruppo normale di G il cui indice in G divide $n!$. (Suggerimento: considerare l'azione di G su G/H per traslazioni sinistre.)

5. Sia p il più piccolo primo che divide l'ordine di un gruppo finito G. Mostrare che ogni suo sottogruppo H di indice p è normale.

6. Mostrare che ogni gruppo di ordine p^2, con p primo, è abeliano. Mostrare che, a meno di isomorfismi, vi sono soltanto due gruppi di quest'ordine.

7. Sia H un sottogruppo proprio di un gruppo finito G. Mostrare che $G \neq \bigcup_{g\in G} gHg^{-1}$.

8. Supponiamo che G agisca su S e H su T, con $S \cap T = \varnothing$. Poniamo $U = S \cup T$ e, per $g \in G$, $h \in H$, $s \in S$, $t \in T$, definiamo $(g,h)s = gs$ e $(g,h)t = ht$. Mostrare che si ottiene un'azione di $G \times H$ su U.

9. Si dice che un gruppo H agisce su un gruppo K *per automorfismi* se H agisce su K e per ogni $h \in H$ l'applicazione $k \mapsto hk$ è un automorfismo di K. Supponiamo che ciò accada e sia G l'insieme prodotto $K \times H$. Vi si definisca la composizione
$$(k_1, h_1)(k_2, h_2) = (k_1(h_1k_2), h_1h_2)$$
e si ponga $1 = (1,1)$, con le unità di K e H. Verificare che si ottiene un gruppo, che $h \mapsto (1,h)$ è un monomorfismo di H in $K \times H$ e che $k \mapsto (k,1)$ è un monomorfismo di K in $K \times H$ la cui immagine è normale. Si dice che G è un *prodotto semidiretto* di K e H. Se K, H sono finiti, osservare che $|K \times H| = |K|\,|H|$.

10. Sia G un gruppo, H un gruppo di trasformazioni di S e G^S l'insieme delle applicazioni di S in G. Allora G^S è un gruppo (la potenza diretta di G indicizzata da S) se si pone $(f_1f_2)(s) = f_1(s)f_2(s)$, $f_i \in G^S$, $s \in S$. Per $h \in H$, $f \in G^S$ si definisca hf mediante $(hf)(s) = f(h^{-1}s)$. Verificare che si ottiene un'azione di H su G^S per automorfismi. Il corrispondente prodotto semidiretto si chiama *prodotto intrecciato (non ristretto)* $G \wr H$ di G con H.

11. Siano G, H, S come nell'esercizio 10 e supponiamo che G agisca su T. Sia $(f,h) \in G \wr H$, dove $f : S \to G$. Il prodotto di due elementi $(f_1,h_1), (f_2,h_2)$ è $(f_1(h_1f_2), h_1h_2)$. Per $(t,s) \in T \times S$ si ponga $(f,h)(t,s) = (f(s)t, hs)$. Verificare che si ottiene un'azione di $G \wr H$

su $T \times S$.[8] Se tutti gli insiemi sono finiti, osservare che $|G \wr H| = |G|^{|S|}|H|$ e che il grado dell'azione, definito come la cardinalità dell'insieme su cui essa si svolge, è il prodotto dei gradi delle azioni di H e di G.

12. Un'azione di G su S si dice k *volte transitiva*, per $k = 1, 2, 3, \ldots$, se per due k-uple qualsiasi $(x_1, \ldots, x_k), (y_1, \ldots, y_k) \in S^{(k)}$, con gli x_i distinti fra loro e gli y_i distinti fra loro, esiste $g \in G$ tale che $gx_i = y_i$, $1 \leq i \leq k$. Mostrare che un'azione doppiamente transitiva è primitiva.

13. Mostrare che, se l'azione di G su S è primitiva ed effettiva, l'azione su S indotta da ogni sottogruppo normale $H \neq 1$ di G è transitiva.

1.13 I teoremi di Sylow

Abbiamo visto che l'ordine di un sottogruppo di un gruppo finito G divide $|G|$ e che, se G è ciclico, per ogni divisore di $|G|$ esiste uno e un solo sottogruppo di quell'ordine. Viene naturale chiedersi se per ogni k che divide $|G|$ esista sempre un sottogruppo di ordine k. Qualche tentativo mostra che la risposta è negativa. Per esempio, il gruppo alterno A_4, di ordine 12, non ha sottogruppi di ordine 6. Mostreremo inoltre nel capitolo 4 che per $n \geq 5$ il gruppo A_n è semplice, ossia non contiene sottogruppi normali distinti da 1 e da A_n. Poiché ogni sottogruppo di indice due è normale, ne segue che A_n, $n \geq 5$, non ha sottogruppi di ordine $n!/4$. Il principale risultato positivo di questo tipo fu scoperto da Sylow: se una potenza di primo p^k divide l'ordine di un gruppo finito G, allora G contiene un sottogruppo di ordine p^k. Sylow dimostrò anche altri risultati importanti sui sottogruppi di ordine p^m, dove p^m è la massima potenza di p che divide $|G|$. Studiamoli ora. Dimostriamo anzitutto il seguente enunciato.

Sylow I. *Se p è primo e p^k, $k \geq 0$, divide $|G|$ (supposto finito), G contiene un sottogruppo di ordine p^k.*

Dimostrazione. Procediamo per induzione su $|G|$. Il risultato è evidente per $|G| = 1$ e possiamo supporlo vero per tutti i gruppi di ordine minore di $|G|$. Dimostriamo anzitutto un caso particolare, che risale a Cauchy: se G è finito abeliano e p è un divisore primo di $|G|$, allora G contiene un elemento di ordine p. Prendiamo $a \neq 1$ in G. Se l'ordine r di a è divisibile per p, poniamo $r = pr'$: $b = a^{r'}$ ha ordine p. Se invece r è primo con p, l'ordine $|G|/r$ di $G/\langle a \rangle$ è divisibile per p ed è minore di $|G|$. Questo gruppo fattore contiene quindi un elemento $b\langle a \rangle$ di ordine p. L'ordine s di b è divisibile per p, poiché $(b\langle a \rangle)^s = b^s\langle a \rangle = \bar{1} = \langle a \rangle$ e dunque l'ordine p di $b\langle a \rangle$ divide s. Da b, che ha ordine divisibile per p, otteniamo come prima un elemento di ordine p.

Dopo questo preliminare la dimostrazione è rapida. Consideriamo l'equazione delle classi $(42')$, $|G| = |C| + \sum_j [G : C(y_j)]$. Se $p \nmid |C|$, allora per qualche j si ha $p \nmid [G : C(y_j)]$. Quindi $p^k \mid |C(y_j)|$, e $C(y_j)$ ha ordine minore di $|G|$ perché y_j non è nel centro. Per ipotesi induttiva, $C(y_j)$ contiene un sottogruppo di ordine p^k. Supponiamo ora $p \mid |C|$. Per il risultato di Cauchy il centro contiene un elemento c di ordine p. Allora $\langle c \rangle$ è un sottogruppo normale di G di ordine p e l'ordine $|G|/p$ di $G/\langle c \rangle$ è divisibile per p^{k-1} (il caso $k = 0$ è immediato). Per induzione $G/\langle c \rangle$ contiene un sottogruppo di ordine p^{k-1}. È della forma $H/\langle c \rangle$, dove H è un sottogruppo di G che contiene $\langle c \rangle$. Allora

$$|H| = |H/\langle c \rangle|\,|\langle c \rangle| = p^{k-1}p = p^k.$$

□

[8] Nota alla traduzione. La formula precedente riproduce quella della scansione. Con la convenzione dell'esercizio 10, affinché valga la legge dell'azione occorre scrivere $(f, h)(t, s) = (f(hs)t, hs)$.

Sia p^m la massima potenza di p che divide $|G|$. Il primo teorema di Sylow assicura l'esistenza di sottogruppi di ordine p^m in G. Questi si chiamano *p-sottogruppi di Sylow* di G. Di essi tratta il prossimo teorema.

Sylow II. *(1) Due p-sottogruppi di Sylow di G sono coniugati in G: se P_1, P_2 sono tali sottogruppi, esiste $a \in G$ con $P_2 = aP_1a^{-1}$. (2) Il numero dei p-sottogruppi di Sylow divide l'indice di uno qualunque di essi ed è congruo a 1 modulo p. (3) Ogni sottogruppo di ordine p^k è contenuto in un p-sottogruppo di Sylow.*

Per dimostrarlo considereremo l'azione di G per coniugio sull'insieme Π dei p-sottogruppi di Sylow. Più in generale, se H è un sottogruppo di un gruppo G e $g \in G$, anche gHg^{-1} è un sottogruppo. Si ha quindi un'azione di G sull'insieme Γ dei sottogruppi di G per coniugio, ${}^gH = gHg^{-1}$. Lo stabilizzatore di H è il sottogruppo $N(H)$ (o $N_G(H)$) $= \{g \in G \mid gHg^{-1} = H\}$, detto *normalizzante* di H in G. Evidentemente $H \subseteq N(H)$ e H è normale in $N(H)$. L'orbita di H per l'azione di coniugio è $\{gHg^{-1} \mid g \in G\}$. La formula di conteggio del § 1.12 dà $|\{gHg^{-1} \mid g \in G\}| = [G : N(H)]$. Se G è finito, $[G : N(H)]$ divide $[G : H]$, poiché $G \supseteq N(H) \supseteq H$ e dunque $[G : H] = [G : N(H)][N(H) : H]$.

Sia ora G finito e sia Π l'insieme dei suoi p-sottogruppi di Sylow. Se $P \in \Pi$, allora $gPg^{-1} \in \Pi$, quindi il coniugio su Γ induce un'azione di G su Π. Ci occorrerà il seguente lemma.

Lemma. *Sia P un p-sottogruppo di Sylow di G. Se H è un sottogruppo di ordine p^k contenuto in $N(P)$, allora $H \subseteq P$.*

Dimostrazione. Poiché H è un sottogruppo di $N(P)$ e P è normale in $N(P)$, HP è un sottogruppo e $HP/P \cong H/(H \cap P)$ (teorema di isomorfismo del § 1.10). Dunque HP/P è isomorfo a un gruppo fattore di H e ha ordine una potenza p^j. Allora $|HP| = p^j|P|$. Poiché P è un p-sottogruppo di Sylow, $j = 0$, $HP = P$ e quindi $H \subseteq P$. □

È evidente che P è un p-sottogruppo di Sylow di $N(P)$. Dal lemma risulta che è l'unico.

Dimostrazione di Sylow II. Agisca G per coniugio sull'insieme Π dei p-sottogruppi di Sylow. Sia Σ un'orbita. Prendiamo $P \in \Sigma$ e restringiamo l'azione di G su Σ a P. Allora Σ si scompone in P-orbite, una delle quali è $\{P\}$. È anzi l'unica P-orbita di cardinalità uno: se $\{P'\}$ è una tale orbita, allora $P \subseteq N(P')$, e quindi $P = P'$, poiché P' è l'unico p-sottogruppo di Sylow di $N(P')$. Ogni P-orbita ha per cardinalità una potenza di p, perché la sua cardinalità divide $|P|$. Pertanto $|\Sigma| \equiv 1 \pmod{p}$. Mostriamo che $\Sigma = \Pi$. Altrimenti esiste $P \in \Pi$ non appartenente a Σ. Applicando a questo P l'argomento precedente, si vede che in Σ non esistono P-orbite di cardinalità uno. Si avrebbe $|\Sigma| \equiv 0 \pmod{p}$, in contraddizione con $|\Sigma| \equiv 1 \pmod{p}$. Dunque $\Sigma = \Pi$, ossia G agisce transitivamente su Π. Questo prova (1). Si ha anche $|\Pi| \equiv 1 \pmod{p}$, la seconda affermazione di (2); la prima segue da $|\Pi| = [G : N(P)]$. Sia infine H un sottogruppo di G di ordine p^k e restringiamo a H l'azione di G su Π. Le H-orbite hanno cardinalità una potenza di p e $|\Pi| \equiv 1 \pmod{p}$, quindi esiste un'orbita $\{P\}$ di un solo elemento. Allora $H \subseteq N(P)$ e, per il lemma, $H \subseteq P$. Questo prova (3). □

Esercizi

1. Mostrare che, se P è un sottogruppo di Sylow, $N(N(P)) = N(P)$.
2. Mostrare che non vi sono gruppi semplici di ordine 148 o di ordine 56.
3. Mostrare che non vi è un gruppo semplice di ordine pq, con p, q primi (cfr. esercizio 5 del § 1.12).
4. Mostrare che ogni gruppo non abeliano di ordine 6 è isomorfo a S_3.

5. Determinare il numero dei gruppi non isomorfi di ordine 15.

Un elemento di ordine 2 in un gruppo si dice *involuzione*. Lo studio delle involuzioni e dei loro centralizzanti fornisce informazioni importanti sulla struttura di un gruppo finito. I prossimi cinque esercizi tracciano un programma per caratterizzare S_5 in questo modo. Mi sono stati comunicati da Walter Feit, che attribuisce i primi quattro a Richard Brauer, osservando tuttavia che John Thompson riconobbe per primo l'importanza del risultato dell'esercizio 9. In questi esercizi e nei successivi G è un gruppo finito.

6. Siano u, v involuzioni distinte di G. Mostrare che $\langle u, v\rangle$ è isomorfo a un gruppo diedrale.

7. Siano u, v involuzioni di G. Se uv ha ordine dispari, mostrare che u, v sono coniugate in G ($v = gug^{-1}$).

8. Siano u, v involuzioni in G e supponiamo che uv abbia ordine pari $2n$, sicché $w = (uv)^n$ è un'involuzione. Mostrare che $u, v \in C(w)$.

9. Supponiamo che G contenga esattamente due classi di coniugio di involuzioni. Siano u_1, u_2 involuzioni non coniugate e $c_i = |C(u_i)|$, $i = 1, 2$. Sia S_i l'insieme delle coppie ordinate (x, y) con x coniugato a u_1, y coniugato a u_2 e $(xy)^n = u_i$ per qualche n. Sia $s_i = |S_i|$. Dimostrare che $|G| = c_1 s_2 + c_2 s_1$. (Suggerimento: contare in due modi le coppie (x, y) con x coniugato a u_1 e y coniugato a u_2. Anzitutto il numero è $(|G|/c_1)(|G|/c_2)$. Poiché x e y non sono coniugati, gli esercizi 7 e 8 implicano che, per $n = o(xy)/2$, $(xy)^n$ è coniugato a u_1 oppure a u_2. Dunque $(|G|/c_1)(|G|/c_2) = (|G|/c_1)s_1 + (|G|/c_2)s_2$.)

10. (*Una caratterizzazione astratta di S_5.*) Sia G un gruppo con esattamente due classi di coniugio di involuzioni, rappresentate da u_1, u_2. Supponiamo $C_1 = C(u_1) \cong \langle u_1\rangle \times S_3$ e $C_2 = C(u_2)$ diedrale di ordine 8. Allora $G \cong S_5$.

    Traccia della dimostrazione.

    (i) Poiché una qualche involuzione appartiene al centro di un sottogruppo di Sylow, C_2 è un 2-sottogruppo di Sylow.

    (ii) Sostituendo u_1 con un suo coniugato, possiamo supporre $u_1 \in C_2$; allora $u_2 \in C_1$.

    (iii) C_2 contiene tre classi di involuzioni. Se x è un'involuzione di C_2, $x \neq u_2$, x è coniugata a xu_2. Poiché G contiene due classi di involuzioni, dedurre che o $s_2 = 0$, oppure $s_2 = 4$ e C_2 contiene un gruppo non ciclico V di ordine 4 tutte le cui involuzioni sono coniugate in G a u_2.

    (iv) C_1 contiene tre classi di coniugio di involuzioni. Se x è un'involuzione di C_1, $x \neq u_1$, allora x non è coniugata a xu_1 in C_1. Poiché G contiene due classi di involuzioni, dal punto (iii) segue che per ogni involuzione x di C_1 diversa da u_1 esattamente una fra x e xu_1 è coniugata a u_1. Dedurne che $s_1 = 9$ nella notazione dell'esercizio 9.

    (v) Usare la formula dell'esercizio 9 per mostrare che o $s_2 = 4$, $|G| = 120$, oppure $s_2 = 0$, $|G| = 72$.

    (vi) Mostrare che $|G| \neq 72$ nel modo seguente. Sia P un 3-sottogruppo di Sylow di C_1 e supponiamo $|G| = 72$. Sia Q un 3-sottogruppo di Sylow di G contenente P. Allora $|Q| = 9$ e $\langle C_1, Q\rangle \subseteq N(P)$, dunque $36 \mid |N(P)|$. Esiste perciò H con $C(P) \subseteq H$ e $|H| = 36$. Ne segue $u_1 \in H$ e, poiché u_2 è un quadrato, $u_2 \in H$. Poiché $[G : H] = 2$, H è normale in G e contiene quindi tutte le involuzioni di G. Allora $C_2 \cap H$ contiene tutte le involuzioni di C_2. Ciò è impossibile perché $|C_2 \cap H| = 4$ e C_2 contiene cinque involuzioni.

(vii) Per (iii), C_2 contiene un gruppo non ciclico V di ordine 4 con $u_2 \in V$, e tutte le involuzioni di V sono coniugate in G. Sia $x \in G$ tale che $x^{-1}u_2x \neq u_2$ e $x^{-1}u_2x \in V$. Allora $x^{-1}C_2x \neq C_2$ e $u_2 \in C(x^{-1}u_2x) = x^{-1}C_2x$.

(viii) $C(V) = V$. Per (vii), $N(V)$ contiene almeno due 2-sottogruppi di Sylow di G.

(ix) $N(V)/V \cong \operatorname{Aut} V \cong S_3$. Quindi $|N(V)| = 24$.

(x) $[G : N(V)] = 5$. Mostrare che G agisce effettivamente sullo spazio delle classi $G/N(V)$ e che perciò $G \cong S_5$.

I quattro esercizi seguenti hanno lo scopo di dimostrare un'estensione del primo teorema di Sylow. Se p è un primo e $p^k \mid |G|$, il numero dei sottogruppi di ordine p^k è congruo a 1 modulo p. Il teorema è dovuto a Frobenius. La dimostrazione che indicheremo, particolarmente elegante, è di P. X. Gallagher (*Archiv der Mathematik*, vol. XXIII, 1967, p. 469). Si basa sull'azione di G sull'insieme S dei sottoinsiemi di cardinalità p^k. Le dimostrazioni di questo tipo del teorema di Sylow hanno avuto una storia curiosa. Sembra che il metodo fosse stato scoperto da G. A. Miller più di cinquant'anni prima (*Annals of Math.*, vol. 16, 1915, pp. 169–171), poi completamente dimenticato fino alla sua riscoperta da parte di H. Wielandt nel 1959.

11. Sia $|G| = p^k m$, con p primo, e sia n il numero dei sottogruppi di G di ordine p^k. Sia S l'insieme dei sottoinsiemi di G di cardinalità p^k, sul quale G agisce per traslazione sinistra. Per $A \in S$ poniamo $H_A = \operatorname{Stab} A$. Allora H_A agisce su A per traslazioni sinistre. Osservare che le orbite in A per quest'azione sono classi laterali destre. Dedurne che $|H_A| \mid p^k$.

12. Sia S_0 l'insieme degli $A \in S$ tali che $|H_A| = p^k$ e $\overline{S}_0$ quello dei $B \in S$ tali che $|H_B| = p^l$ con $l < k$. Osservare che l'orbita di ciascuno di questi B sotto G ha cardinalità divisibile per pm. Dimostrare così che $|S| \equiv |S_0| \pmod{pm}$.

13. Siano $A \in S_0$ e $x \in A$. Si ha $H_A x \subseteq A$ e, poiché $|H_A| = p^k = |A|$, risulta $H_A x = A$. Dunque A è una classe laterale destra di H_A, che è un sottogruppo di ordine p^k. Viceversa, siano H un qualsiasi sottogruppo di ordine p^k e Hx una sua classe destra. Allora $H(Hx) = Hx$ e $\operatorname{Stab}(Hx)$ contiene H. Dall'esercizio 11 segue $\operatorname{Stab}(Hx) = H$, quindi $Hx \in S_0$. Concludere che $|S_0| = nm$, dove n è il numero dei sottogruppi di ordine p^k.

14. Osservare che $|S|$ dipende soltanto da $|G|$ e p^k e che, per gli esercizi 12 e 13, $n = |S_0|/m \equiv |S|/m \pmod{p}$. La classe di n modulo p dipende quindi soltanto da $|G|$ e p^k ed è indipendente dal gruppo G. Considerare un gruppo ciclico di ordine $|G|$: in questo caso esiste esattamente un sottogruppo di ordine p^k. Dunque $n \equiv 1 \pmod{p}$.

Nei due esercizi seguenti si costruisce un gruppo isomorfo a un qualsiasi p-sottogruppo di Sylow di S_n, per p primo non superiore a n.

15. Mostrare che l'ordine di un p-sottogruppo di Sylow di S_n è $p^{v_p(n!)}$, dove

$$v_p(n!) = \left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \cdots,$$

con $[k/l]$ il massimo intero $\leq k/l$. Mostrare inoltre che, scrivendo $n = a_0 + a_1 p + a_2 p^2 + \cdots + a_k p^k$ con $0 \leq a_i < p$ (rappresentazione di n in base p), si ha

$$v_p(n!) = \sum_{i=1}^{k} a_i (1 + p + \cdots + p^{i-1}).$$

16. Sia Z_p il sottogruppo di S_p generato dal ciclo $(12\dots p)$. Osservare che il prodotto intrecciato $Z_p \wr Z_p$ ha ordine p^{p+1} ed è isomorfo a un sottogruppo di S_{p^2} (esercizi 10 e 11 del § 1.12). Per $r \geq 1$ si definisca induttivamente $Z_p^{\wr r}$ ponendo $Z_p^{\wr 1} = Z_p$ e $Z_p^{\wr(k+1)} = Z_p^{\wr k} \wr Z_p$. Mostrare che $Z_p^{\wr r}$ ha ordine $p^{p^{r-1}+p^{r-2}+\cdots+1}$ ed è isomorfo a un sottogruppo di S_{p^r}. Dedurne che, se $n = a_0 + a_1 p + \cdots + a_k p^k$, $0 \leq a_i < p$, ogni p-sottogruppo di Sylow di S_n è isomorfo a

$$\underbrace{Z_p^{\wr 1} \times \cdots \times Z_p^{\wr 1}}_{a_1} \times \underbrace{Z_p^{\wr 2} \times \cdots \times Z_p^{\wr 2}}_{a_2} \times \cdots \times \underbrace{Z_p^{\wr k} \times \cdots \times Z_p^{\wr k}}_{a_k}.$$

Capitolo 2
Anelli

In questo capitolo iniziamo lo studio di un secondo tipo di struttura algebrica, detto *anello*. Il prototipo di queste strutture è l'anello $\mathbb{Z}$ degli interi, che nel capitolo precedente abbiamo considerato dal punto di vista dei monoidi, ricavandone i due monoidi $(\mathbb{Z}, +, 0)$ e $(\mathbb{Z}, \cdot, 1)$. La teoria degli anelli considera simultaneamente queste due strutture e le collega mediante la legge distributiva. Mentre la teoria dei gruppi ha avuto essenzialmente un'unica origine, lo studio delle trasformazioni biiettive rispetto alla composizione, la teoria degli anelli è nata dalla fusione di diverse teorie particolari. Per questo apparirà meno ordinata e unitaria della teoria dei gruppi. La moltitudine degli esempi, molti dei quali familiari al lettore, dovrebbe tuttavia testimoniare in modo convincente la ricchezza di questo ramo dell'algebra. Nel capitolo successivo vedremo che anche gli anelli nascono in modo analogo ai gruppi di trasformazioni, precisamente come anelli di endomorfismi di gruppi abeliani. Abbiamo inoltre il concetto di modulo, che per gli anelli è l'esatto analogo del concetto di gruppo che opera su un insieme.

Inizieremo con le definizioni e con esempi dei vari tipi di anelli: domini, anelli con divisione, anelli commutativi e campi. Studieremo poi le nozioni fondamentali di ideale, anello quoziente e omomorfismo, analoghe rispettivamente a quelle di sottogruppo normale, gruppo quoziente e omomorfismo nella teoria dei gruppi. Nella seconda metà del capitolo ci occuperemo principalmente degli anelli commutativi, considerando dapprima costruzioni e caratterizzazioni di alcune loro estensioni: i campi delle frazioni dei domini commutativi e gli anelli di polinomi in un'indeterminata x. Passeremo quindi alla teoria elementare della fattorizzazione nei domini commutativi. Di tanto in tanto indicheremo applicazioni, specialmente alla teoria dei numeri. L'ultimo paragrafo, che si può considerare facoltativo, sarà dedicato agli «anelli senza unità» e alla loro immersione in «anelli», che per convenzione ammettono sempre un'unità.

Gran parte della materia sembrerà familiare. Lo studente osservi però che il nostro punto di vista differisce sotto certi aspetti da altri che può avere incontrato. Per esempio, trattiamo i polinomi formalmente anziché come funzioni, e ammettiamo matrici con elementi in un anello qualsiasi, anziché soltanto nell'anello $\mathbb{R}$ dei reali. Diamo inoltre risalto alle proprietà fondamentali degli omomorfismi associate ad alcune costruzioni di estensioni di un anello dato. In casi importanti, queste proprietà caratterizzano l'estensione e avranno una parte di rilievo nel seguito.

2.1 Definizione e proprietà elementari

Definizione 2.1. Un *anello* è una struttura costituita da un insieme non vuoto R, da due operazioni binarie $+, \cdot$ in R e da due elementi distinti per il loro ruolo, $0, 1 \in R$, tali che:

1. $(R, +, 0)$ è un gruppo abeliano;
2. $(R, \cdot, 1)$ è un monoide;
3. valgono, per ogni $a, b, c \in R$, le leggi distributive[1]

$$\text{(D)} \qquad \begin{aligned} a(b+c) &= ab + ac, \\ (b+c)a &= ba + ca. \end{aligned}$$

[1]Il termine «anello» sembra essere stato usato per primo da A. Fraenkel, che diede un sistema di assiomi per questo concetto in un articolo nel *Journal für die reine und angewandte Mathematik*, vol. 145 (1914). La sua definizione risentiva però di alcune ipotesi particolari, inadatte a una teoria generale. La nozione qui definita si deve a Emmy Noether, che la formulò in un articolo nei *Mathematische Annalen*, vol. 83 (1921). In precedenza il termine *Zahlring* era già comparso nella teoria algebrica dei numeri.

Le ipotesi 1 e 2 richiedono dunque che $a+b$ e ab appartengano a R e che valgano le seguenti condizioni:

(A1) $$(a+b)+c = a+(b+c),$$

(A2) $$a+b = b+a,$$

(A3) $$a+0 = a = 0+a,$$

(A4) per ogni a esiste $-a$: $$a+(-a) = 0 = (-a)+a,$$

(M1) $$(ab)c = a(bc),$$

(M2) $$a1 = a = 1a.$$

La struttura $(R,+,0)$ è il *gruppo additivo* di R; $(R,\cdot,1)$ è il suo *monoide moltiplicativo*. Un sottoinsieme S di un anello R è un *sottoanello* se è un sottogruppo del gruppo additivo e un sottomonoide del monoide moltiplicativo di R. L'intersezione di una famiglia qualsiasi di sottoanelli di R è chiaramente un sottoanello. Per ogni sottoinsieme A di R possiamo quindi definire il *sottoanello generato da* A come l'intersezione di tutti i sottoanelli di R che contengono A. Lo caratterizzano le seguenti proprietà: è un sottoanello, contiene A ed è contenuto in ogni sottoanello che contiene A.

Esempi

1. $\mathbb{Z}$, con gli usuali $+,\cdot,0,1$. Nell'introduzione abbiamo osservato che è un anello.
2. $\mathbb{Q}$, i numeri razionali, con gli usuali $+,\cdot,0,1$.
3. $\mathbb{R}$, l'anello dei numeri reali.
4. $\mathbb{C}$, l'anello dei numeri complessi. $\mathbb{R}$, $\mathbb{Q}$ e $\mathbb{Z}$ sono sottoanelli di $\mathbb{C}$.
5. L'insieme $\mathbb{Z}[\sqrt{2}]$ dei numeri reali della forma $m+n\sqrt{2}$, con $m,n\in\mathbb{Z}$. La differenza di due numeri di $\mathbb{Z}[\sqrt{2}]$ vi appartiene ancora. Inoltre $1\in\mathbb{Z}[\sqrt{2}]$ e, per $m,n,m',n'\in\mathbb{Z}$,
$$(m+n\sqrt{2})(m'+n'\sqrt{2}) = (mm'+2nn')+(mn'+nm')\sqrt{2}\in\mathbb{Z}[\sqrt{2}].$$
Dunque $\mathbb{Z}[\sqrt{2}]$ è un sottoanello di $\mathbb{R}$.
6. Come in 5, sostituendo $\mathbb{Z}$ con $\mathbb{Q}$. Gli stessi calcoli mostrano che si ottiene un sottoanello di $\mathbb{R}$.
7. Analogamente, $\mathbb{Z}[\sqrt{-1}]$ e $\mathbb{Q}[\sqrt{-1}]$, costituiti dai complessi $m+n\sqrt{-1}$ con $m,n\in\mathbb{Z}$ nel primo caso e $m,n\in\mathbb{Q}$ nel secondo, sono sottoanelli di $\mathbb{C}$. Sono i sottoanelli generati rispettivamente da $\mathbb{Z}$ e $\sqrt{-1}$ e da $\mathbb{Q}$ e $\sqrt{-1}$. Il primo si chiama *anello degli interi di Gauss*.
8. L'insieme Γ delle funzioni continue a valori reali sull'intervallo $[0,1]$. Si definiscono nel modo usuale $(f+g)(x) = f(x)+g(x)$ e $(fg)(x) = f(x)g(x)$. Siano 0 e 1 le funzioni costanti di valore rispettivamente zero e uno. Allora $(\Gamma,+,\cdot,0,1)$ è un anello.
9. L'insieme $\{0,1,2\}$ con lo zero e l'uno indicati e con l'addizione e la moltiplicazione definite dalle tavole

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

è un anello. Si può verificarlo direttamente. Fra poco sarà chiaro anche senza questa verifica (forse lo è già).

Diverse proprietà elementari degli anelli seguono dal fatto che un anello è un gruppo abeliano rispetto all'addizione e un monoide rispetto alla moltiplicazione. Per esempio, $-(a+b) = -a-b = (-a)+(-b)$ e, se per $n \in \mathbb{Z}$ si definisce na come in precedenza, valgono le regole dei multipli (o delle potenze) in un gruppo abeliano:

$$n(a+b) = na + nb, \qquad (n+m)a = na + ma, \qquad (nm)a = n(ma).$$

Valgono inoltre le leggi associative generalizzate per l'addizione e la moltiplicazione e la legge commutativa generalizzata per l'addizione (si vedano le pp. 40 e 41 dell'originale). Osserviamo alcune semplici conseguenze delle leggi distributive. In primo luogo, per induzione su m e n si ottiene la generalizzazione

$$\begin{aligned}(a_1 + a_2 + \cdots + a_m)(b_1 + b_2 + \cdots + b_n) \\ = a_1b_1 + a_1b_2 + \cdots + a_1b_n + a_2b_1 + a_2b_2 + \cdots + a_2b_n + \cdots \\ + a_mb_1 + a_mb_2 + \cdots + a_mb_n,\end{aligned}$$

ovvero

$$\left(\sum_{i=1}^{m} a_i\right)\left(\sum_{j=1}^{n} b_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_ib_j.$$

Osserviamo poi che $a0 = 0 = 0a$ per ogni a. Infatti $a0 = a(0+0) = a0 + a0$; sommando $-a0$ si ottiene $a0 = 0$. Analogamente $0a = 0$. Dall'uguaglianza

$$0 = 0b = (a + (-a))b = ab + (-a)b$$

segue $(-a)b = -ab$. Allo stesso modo $a(-b) = -ab$, e pertanto

$$(-a)(-b) = -a(-b) = -(-ab) = ab.$$

Se a e b commutano, cioè $ab = ba$, allora $a^mb^n = b^na^m$. Per induzione si dimostra anche il *teorema del binomio*

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + b^n, \tag{1}$$

dove il coefficiente binomiale è

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}. \tag{2}$$

Il passo induttivo si ricava dalla formula

$$\binom{r}{k} + \binom{r}{k-1} = \frac{r!}{k!(r-k)!} + \frac{r!}{(k-1)!(r-k+1)!} = \frac{(r+1)!}{k!(r-k+1)!} = \binom{r+1}{k}.$$

Il lettore svolga la dimostrazione osservando in quale preciso punto interviene la commutatività della moltiplicazione.

Esercizi

1. Sia C l'insieme delle funzioni continue a valori reali sulla retta reale $\mathbb{R}$. Mostrare che C, con la consueta addizione di funzioni e con 0, è un gruppo abeliano e che C, con il prodotto $(f \cdot g)(x) = f(g(x))$ e con 1 uguale all'applicazione identica, è un monoide. Con queste operazioni e con 0 e 1, C è un anello?

2. Mostrare che in un anello R vale $a(b-c) = ab - ac$, dove $b - c = b + (-c)$, e che $n(ab) = (na)b = a(nb)$ per $n \in \mathbb{Z}$.

3. Mostrare che, se si ammettono tutti gli assiomi di anello tranne la commutatività dell'addizione, questa ne segue e si ottiene dunque un anello.

4. Sia I l'insieme dei complessi della forma $m + n\sqrt{-3}$, dove o $m, n \in \mathbb{Z}$ oppure m e n sono entrambi metà di interi dispari. Mostrare che I è un sottoanello di $\mathbb{C}$.

5. Siano a, b elementi di un anello. Si definiscano $a^{(0)} = a$, $a' = [a, b] = ab - ba$ e, per induzione, $a^{(k)} = [a^{(k-1)}, b]$ (per semplicità nella notazione $a^{(k)}$ non indichiamo la dipendenza da b). Dimostrare la formula

$$\sum_{i=0}^{k} b^i a b^{k-i} = \sum_{j=0}^{k} \binom{k+1}{j+1} b^{k-j} a^{(j)}.$$

2.2 Tipi di anelli

Imponendo condizioni particolari al monoide moltiplicativo otteniamo vari tipi di anelli. Per esempio, un anello R si dice *commutativo* se $(R, \cdot, 1)$ è commutativo. Questa proprietà vale per tutti gli esempi del paragrafo precedente. Nei prossimi due paragrafi daremo esempi di anelli non commutativi. Un anello è detto *dominio* (anche *dominio d'integrità*) se l'insieme R^* dei suoi elementi non nulli è un sottomonoide di $(R, \cdot, 1)$. Nella definizione di dominio è implicito che $R \neq 0$. Oltre a ciò, la condizione richiesta è che da $a \neq 0$, $b \neq 0$ in R segua $ab \neq 0$. Ogni sottoanello di un dominio è evidentemente un dominio. Tutti gli esempi del § 2.1, tranne l'ottavo, sono domini. Nell'esempio 8 si possono invece prendere gli elementi f e g così definiti:

$$f(x) = \begin{cases} 0 & 0 \leq x \leq \frac{1}{2}, \\ x - \frac{1}{2} & \frac{1}{2} < x \leq 1, \end{cases} \qquad g(x) = \begin{cases} -x + \frac{1}{2} & 0 \leq x \leq \frac{1}{2}, \\ 0 & \frac{1}{2} < x \leq 1. \end{cases}$$

Allora $f \neq 0$ (la funzione costante zero) e $g \neq 0$, ma $fg = 0$. Dunque l'anello delle funzioni continue a valori reali su $[0, 1]$ non è un dominio.

Se per un elemento a di un anello R esiste $b \neq 0$ tale che $ab = 0$ (rispettivamente $ba = 0$), si dice che a è un *divisore sinistro* (rispettivamente *destro*) *dello zero*. È chiaro che 0 è un divisore sia sinistro sia destro dello zero se R ha più di un elemento. Se $a \neq 0$ è un divisore sinistro dello zero e $ab = 0$ con $b \neq 0$, allora b è un divisore destro non nullo dello zero. Da ciò e dalla definizione di dominio si vede che $R \neq 0$ è un dominio se e solo se non ha divisori non nulli dello zero (né destri né sinistri).

Osserviamo anche che un anello è un dominio se e solo se $R \neq 0$ e valgono le leggi di cancellazione ristrette: da $ab = ac$, $a \neq 0$, segue $b = c$, e da $ba = ca$, $a \neq 0$, segue $b = c$. Infatti, se R è un dominio e $ab = ac$, allora $a(b - c) = 0$; se $a \neq 0$, si ha $b - c = 0$ e quindi $b = c$. Analogamente $ba = ca$, $a \neq 0$, danno $b = c$. Viceversa sia R un anello non nullo in cui valgano queste leggi di cancellazione. Sia $ab = 0$ con $a \neq 0$. Allora $ab = a0$, e cancellando si ottiene $b = 0$. Dunque R è un dominio.

Un anello R si dice *anello con divisione* (anche *corpo*, in inglese *skew field*, *sfield* o *field*) se l'insieme R^* degli elementi non nulli è un sottogruppo di $(R, \cdot, 1)$. Equivalentemente: $1 \neq 0$ e per ogni $a \neq 0$ esiste b tale che $ab = 1 = ba$. Gli esempi 2, 3, 4, 6 e 9, nonché il secondo anello dell'esempio 7, sono anelli con divisione a moltiplicazione commutativa. Gli anelli con divisione che hanno questa proprietà sono detti *campi*. Nel § 2.4 daremo un esempio di anello con divisione non commutativo.

Ogni anello con divisione è un dominio; poiché i sottoanelli di un dominio sono domini, ogni sottoanello di un anello con divisione è un dominio. Il viceversa è falso: $\mathbb{Z}$ è un dominio che non è un anello con divisione, ed è un sottoanello del campo $\mathbb{Q}$. Chiameremo *sottoanello con divisione* un sottoanello che sia esso stesso un anello con divisione. Se $a \neq 0$ in un anello con divisione R, l'equazione $ax = b$ ha soluzione $x = a^{-1}b$. Per la legge di cancellazione ristretta, questa soluzione è unica. Analogamente $ya = b$ ha l'unica soluzione $y = ba^{-1}$.

Abbiamo visto che gli elementi invertibili di un monoide formano un sottogruppo. In particolare è un sottogruppo l'insieme U degli elementi invertibili di $(R, \cdot, 1)$. Chiameremo *unità* gli elementi di U, anche se ciò è leggermente in contrasto con il nome «l'unità» riservato a 1, e U si chiama *gruppo delle unità* (o *degli elementi invertibili*) dell'anello. Per esempio, il gruppo delle unità di $\mathbb{Z}$ è $\{1, -1\}$.

Esercizi

1. Mostrare che ogni dominio finito è un anello con divisione.

2. Mostrare che in un dominio i soli idempotenti ($e^2 = e$) sono $e = 0$ ed $e = 1$. Un elemento z si dice *nilpotente* se $z^n = 0$ per qualche $n \in \mathbb{Z}^+$. Mostrare che in un dominio il solo nilpotente è 0.

3. Sia z un elemento di un anello per cui esista $w \neq 0$ con $zwz = 0$. Mostrare che z è un divisore sinistro o destro dello zero.

4. Mostrare che, se $1 - ab$ è invertibile in un anello, lo è anche $1 - ba$.

5. Mostrare che una funzione f nell'esempio 8 del § 2.1 è un divisore dello zero se e solo se l'insieme dei punti x tali che $f(x) = 0$ contiene un intervallo aperto. Quali sono gli idempotenti di questo anello? I nilpotenti? Le unità?

6. Sia u un elemento di un anello che ammette un inverso destro. Dimostrare che le condizioni seguenti sono equivalenti: (1) u ha più di un inverso destro; (2) u non è un'unità; (3) u è un divisore sinistro dello zero.

7. (Kaplansky.) Dimostrare che se un elemento di un anello ha più di un inverso destro, ne ha infiniti. Costruire un controesempio per mostrare che ciò non vale per i monoidi.

8. Mostrare che un elemento u di un anello è un'unità con $v = u^{-1}$ se e solo se vale una delle due condizioni seguenti: (1) $uvu = u$, $vu^2v = 1$; (2) $uvu = u$ e v è l'unico elemento che soddisfa questa condizione.

9. (Hua.) Siano a, b elementi di un anello tali che a, b e $ab - 1$ siano unità. Mostrare che $a - b^{-1}$ e $(a - b^{-1})^{-1} - a^{-1}$ sono unità e che vale l'identità

$$\left((a - b^{-1})^{-1} - a^{-1}\right)^{-1} = aba - a.$$

10. (Cohn.) Sia G un gruppo, e un suo elemento e θ un'applicazione di $G_1 = \{x \in G \mid x \neq 1\}$ in sé con le seguenti proprietà:

$$\begin{aligned}
&\text{(i)} \quad \theta(yxy^{-1}) = y\theta(x)y^{-1} && (x \in G_1,\ y \in G),\\
&\text{(ii)} \quad \theta^2(x) = x,\\
&\text{(iii)} \quad \theta(x^{-1}) = e\theta(x)x^{-1},\\
&\text{(iv)} \quad \theta(xy^{-1}) = (\theta(\theta(x)\theta(y^{-1})))\theta(y^{-1}) && (x, y \in G_1,\ x \neq y).
\end{aligned}$$

Mostrare che esiste un unico anello con divisione D tale che $D^* = G$ e, in G, $\theta x = 1 - x$ per $x \in G_1$ ed $e = -1$.

2.3 Anelli di matrici

Il lettore conosce probabilmente già le matrici e i determinanti dallo studio dell'algebra lineare o dell'analisi in più variabili. Generalizzeremo ora queste nozioni nella misura richiesta dal lavoro successivo: matrici con elementi in un anello qualsiasi e determinanti di matrici con elementi in un anello commutativo. Per chi conosce già le matrici e i determinanti, il contenuto del paragrafo si può riassumere dicendo che i risultati consueti rimangono validi a questo livello di generalità.

Siano R un anello e n un intero positivo. Definiamo l'*anello* $M_n(R)$ *delle matrici* $n \times n$ *sull'anello* R. L'insieme sottostante è costituito dalle tabelle $n \times n$, o *matrici*,

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \tag{3}$$

di n righe e n colonne, con *elementi* (anche *coefficienti* o *coordinate*) $a_{ij} \in R$. L'elemento a_{ij} di R all'incrocio della riga i e della colonna j di A è detto elemento di posto (i, j) di A. Due matrici A e $B = (b_{ij})$ sono uguali se e solo se $a_{ij} = b_{ij}$ per ogni i, j; $M_n(R)$ è l'intero insieme delle matrici $n \times n$ con elementi in R. In breve, è il prodotto cartesiano di n^2 copie di R.

Definiamo l'addizione delle matrici con la formula

$$(a_{ij}) + (b_{ij}) = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & \cdots & a_{2n} + b_{2n} \\ \vdots & & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nn} + b_{nn} \end{pmatrix}.$$

Per ottenere la somma addizioniamo cioè gli elementi a_{ij} e b_{ij} che occupano lo stesso posto. Definiamo la matrice 0 come quella che ha tutti gli elementi uguali a 0. È facile verificare che con l'addizione data e con 0, $M_n(R)$ è un gruppo abeliano. La moltiplicazione di matrici è definita da

$$(a_{ij})(b_{ij}) = \begin{pmatrix} \sum_k a_{1k}b_{k1} & \sum_k a_{1k}b_{k2} & \cdots & \sum_k a_{1k}b_{kn} \\ \sum_k a_{2k}b_{k1} & \sum_k a_{2k}b_{k2} & \cdots & \sum_k a_{2k}b_{kn} \\ \vdots & \vdots & & \vdots \\ \sum_k a_{nk}b_{k1} & \sum_k a_{nk}b_{k2} & \cdots & \sum_k a_{nk}b_{kn} \end{pmatrix}.$$

Il prodotto $P = AB$ ha dunque nel posto (i, j) l'elemento $p_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$. Per esempio, nell'anello $M_3(\mathbb{Z})$ delle matrici 3×3 su $\mathbb{Z}$,

$$\begin{pmatrix} 1 & -2 & 3 \\ 0 & 1 & -1 \\ 2 & 5 & -2 \end{pmatrix} \begin{pmatrix} 0 & 3 & 4 \\ 2 & 5 & 1 \\ -1 & -6 & 2 \end{pmatrix} = \begin{pmatrix} -7 & -25 & 8 \\ 3 & 11 & -1 \\ 12 & 43 & 9 \end{pmatrix}.$$

Definiamo la matrice unità 1 mediante

$$1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix};$$

ha cioè l'unità di R sulla diagonale «principale», che va dall'angolo superiore sinistro a quello inferiore destro, e 0 altrove. Si vede subito che $A1 = A = 1A$ per $A \in M_n(R)$. La moltiplicazione è anche associativa: per $A = (a_{ij}), B = (b_{ij}), C = (c_{ij})$, l'elemento di posto (i,l) di $A(BC)$ è $\sum_{j,k} a_{ij}(b_{jk}c_{kl})$, e quello di $(AB)C$ è $\sum_{j,k}(a_{ij}b_{jk})c_{kl}$. L'uguaglianza segue dall'associatività della moltiplicazione in R. Valgono le leggi distributive: gli elementi di posto (i,j) di $A(B+C)$ e di $AB + AC$ sono rispettivamente

$$\sum_k a_{ik}(b_{kj} + c_{kj}) \quad \text{e} \quad \sum_k (a_{ik}b_{kj} + a_{ik}c_{kj}),$$

e sono uguali per una delle leggi distributive in R. Allo stesso modo si verifica l'altra legge distributiva in $M_n(R)$. Abbiamo così provato che $(M_n(R), +, \cdot, 0, 1)$ è un anello.

Definiamo e_{ij} come la matrice che ha un solo 1 nel posto (i,j) e ha 0 in tutti gli altri posti. Le n^2 matrici e_{ij}, $1 \le i, j \le n$, si chiamano abitualmente *unità matriciali*, benché, salvo che per $n = 1$, non siano unità (elementi invertibili) di $M_n(R)$. È facile verificarne la tavola di moltiplicazione:

$$e_{ij}e_{kl} = \delta_{jk}e_{il}, \tag{4}$$

dove δ_{jk} è il *delta di Kronecker*, definito da

$$\delta_{jj} = 1, \qquad \delta_{jk} = 0 \quad \text{se } j \neq k. \tag{5}$$

Inoltre

$$1 = e_{11} + e_{22} + \cdots + e_{nn}. \tag{6}$$

Le e_{ii} sono idempotenti: $e_{ii}^2 = e_{ii}$. Per $n > 1$ si ha $e_{11}e_{12} = e_{12}$ ed $e_{12}e_{11} = 0$, sicché $M_n(R)$ non è mai commutativo se $n > 1$ e $R \neq 0$.

Indicheremo con $\mathrm{diag}\{a_1, a_2, \ldots, a_n\}$ la matrice che ha $a_1, a_2, \ldots, a_n$, nell'ordine, sulla diagonale principale e zeri altrove:

$$\begin{pmatrix} a_1 & & & \\ & a_2 & & 0 \\ & & \ddots & \\ 0 & & & a_n \end{pmatrix}.$$

È chiaro che queste matrici diagonali formano un sottoanello di $M_n(R)$. Poniamo $a' = \mathrm{diag}\{a, a, \ldots, a\}$. Allora $a \mapsto a'$ è iniettiva e $(a+b)' = a' + b'$, $(ab)' = a'b'$, $0' = 0$, $1' = 1$. L'applicazione è quindi un monomorfismo sia di $(R, +, 0)$ in $(M_n(R), +, 0)$ sia di $(R, \cdot, 1)$ in $(M_n(R), \cdot, 1)$. Ne segue che $R' = \{a' \mid a \in R\}$ è un sottoanello di $M_n(R)$ e che $a \mapsto a'$, considerata come applicazione di R in R', è un *isomorfismo di anelli*: definiamo così un'applicazione che sia un isomorfismo per i gruppi additivi e per i monoidi moltiplicativi.

Identifichiamo ora R con il sottoanello isomorfo R' di $M_n(R)$, identificando $a \in R$ con la matrice diagonale $a' = \mathrm{diag}\{a, a, \ldots, a\}$. Ciò è analogo all'identificazione degli interi con i razionali aventi denominatore 1, e ha l'effetto di immergere R in $M_n(R)$. Osserviamo ora che moltiplicare una matrice A a sinistra (a destra) per $a \in R$ equivale a moltiplicare tutti i suoi elementi a sinistra (a destra) per a. Dunque $ae_{ij} = e_{ij}a$ è la matrice con a nel posto (i,j) e zeri altrove. Per la matrice A della (3) si ha così

$$A = \sum_{i,j} a_{ij}e_{ij}. \tag{7}$$

Ogni matrice è dunque una combinazione lineare delle e_{ij} con «coefficienti» $a_{ij} \in R$.

Il gruppo degli elementi invertibili di $M_n(R)$ è detto *gruppo lineare* $GL_n(R)$. Nel caso commutativo ricaveremo ora un criterio, espresso mediante il determinante, affinché una matrice A sia invertibile, ossia appartenga a $GL_n(R)$. Supponiamo che il lettore conosca la definizione di determinante e le sue proprietà elementari.[2] È facile convincersi che le principali formule sui determinanti, reperibili in ogni testo di algebra lineare, valgono per matrici su un anello commutativo qualsiasi. Se R è commutativo, per $A = (a_{ij})$ definiamo il *determinante*

$$\det A = \sum_{\pi} (\operatorname{sg} \pi) a_{1i_1} a_{2i_2} \cdots a_{ni_n}, \tag{8}$$

dove la somma è estesa a tutte le permutazioni π di $1, 2, \ldots, n$, con $i_j = \pi(j)$, e $\operatorname{sg} \pi$ vale 1 o -1 secondo che π sia pari o dispari. Il *cofattore* dell'elemento a_{ij} di A nella (3) è $(-1)^{i+j}$ volte il determinante della matrice $(n-1) \times (n-1)$ ottenuta sopprimendo la riga i e la colonna j di A. Ricordiamo che un determinante si può «sviluppare» secondo una riga o una colonna qualsiasi: si ottiene $\det A$ moltiplicando gli elementi di quella riga (o colonna) per i rispettivi cofattori e sommando i risultati. Dunque, indicato con A_{ij} il cofattore di a_{ij},

$$\begin{aligned} a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} &= \det A, \\ a_{1i}A_{1i} + a_{2i}A_{2i} + \cdots + a_{ni}A_{ni} &= \det A. \end{aligned} \tag{9}$$

Ricordiamo inoltre che la somma dei prodotti degli elementi di una riga (colonna) per i cofattori corrispondenti di un'altra riga (colonna) è zero:

$$\begin{aligned} a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{jn} &= 0, \quad i \neq j, \\ a_{1i}A_{1j} + a_{2i}A_{2j} + \cdots + a_{ni}A_{nj} &= 0, \quad i \neq j. \end{aligned} \tag{10}$$

Queste relazioni ci inducono a definire l'*aggiunta* della matrice $A = (a_{ij})$ come la matrice il cui elemento di posto (i, j) è $\alpha_{ij} = A_{ji}$. Con tale definizione è immediato che le (9) e (10) equivalgono alle equazioni matriciali

$$A(\operatorname{adj} A) = \det A = (\operatorname{adj} A)A, \tag{11}$$

dove $\det A$ al centro indica l'elemento corrispondente $\operatorname{diag}\{\det A, \ldots, \det A\}$ di $M_n(R)$. Ricordiamo anche la regola di moltiplicazione dei determinanti, che in forma matriciale è

$$\det(AB) = (\det A)(\det B). \tag{12}$$

Dalla (12) e dall'uguaglianza $\det 1 = 1$ segue che $A \mapsto \det A$ è un omomorfismo del monoide moltiplicativo di $M_n(R)$, con R commutativo, nel monoide moltiplicativo di R. Un tale omomorfismo manda $GL_n(R)$ in $U(R)$, gruppo delle unità di R: se $A \in GL_n(R)$, allora $\det A$ è un'unità di R. Viceversa, supponiamo che $\Delta = \det A$ sia un'unità. Poiché R è commutativo, $aB = Ba$ per ogni $a \in R$, $B \in M_n(R)$. In particolare $(\operatorname{adj} A)\Delta^{-1} = \Delta^{-1}(\operatorname{adj} A)$, sicché

$$A(\operatorname{adj} A)\Delta^{-1} = \Delta\Delta^{-1} = (\Delta^{-1} \operatorname{adj} A)A.$$

Dunque

$$(\operatorname{adj} A)\Delta^{-1} = A^{-1}. \tag{13}$$

Se $\det A$ è un'unità, A è perciò invertibile, e la (13) ne fornisce l'inversa. La parte principale di quanto abbiamo dimostrato si enuncia così.

Teorema 2.1. *Se R è un anello commutativo, una matrice $A \in M_n(R)$ è invertibile se e solo se il suo determinante è invertibile in R.*

Un caso particolare degno di nota è il seguente.

Corollario. *Se F è un campo, $A \in M_n(F)$ è invertibile se e solo se* $\det A \neq 0$.

[2] I principali teoremi sui determinanti saranno dimostrati più avanti, mediante le algebre esterne (§ 7.2, pp. 416–419 dell'originale).

Esercizi

1. Mostrare che la matrice $\begin{pmatrix} 1 & 4 & 1 \\ 0 & 1 & -1 \\ -3 & -6 & -8 \end{pmatrix}$ è invertibile in $M_3(\mathbb{Z})$ e trovarne l'inversa.

2. Dimostrare che, se R è un anello commutativo, $AB = 1$ in $M_n(R)$ implica $BA = 1$. (Per R non commutativo ciò non è sempre vero.)

3. Verificare che, per ogni $p \in R$ e per $i \neq j$, $1+pe_{ij}$ è invertibile in $M_n(R)$ con inversa $1-pe_{ij}$. Più in generale, mostrare che se z è un elemento nilpotente di un anello (cioè $z^n = 0$ per qualche intero positivo n), allora $1 - z$ è invertibile. Determinarne anche l'inverso.

4. Mostrare che $\mathrm{diag}\{a_1, \ldots, a_n\}$ è invertibile in $M_n(R)$ se e solo se ogni a_i è invertibile in R. Qual è l'inversa?

5. Verificare che $a+b\sqrt{-1} \mapsto \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, per $a, b \in \mathbb{R}$, è un isomorfismo di $\mathbb{C}$ con un sottoanello di $M_2(\mathbb{R})$.

6. Mostrare che in un anello qualsiasi l'insieme $C(S)$ degli elementi che commutano con ogni elemento di un sottoinsieme dato S forma un sottoanello. Se si prende per S l'intero anello, $C = C(S)$ si chiama *centro* dell'anello. Si osservi che tale sottoanello è commutativo. Determinare $C(S)$ in $M_n(R)$ per $S = \{e_{ij} \mid i, j = 1, \ldots, n\}$. Determinare anche il centro di $M_n(R)$.

7. Determinare $C(S)$ quando S è costituito dalla sola matrice $N = e_{12} + e_{23} + \cdots + e_{n-1,n}$.

8. Mostrare che, se R è commutativo e D è l'insieme delle matrici diagonali di $M_n(R)$, allora $C(D) = D$.

9. Sia S un anello contenente un sistema di unità matriciali, cioè elementi $\{e_{ij} \mid i, j = 1, \ldots, n\}$ tali che $e_{ij}e_{kl} = \delta_{jk}e_{il}$ e $\sum_{i=1}^{n} e_{ii} = 1$. Per $1 \leq i, j \leq n$ e $a \in S$ si ponga $a_{ij} = \sum_{k=1}^{n} e_{ki}ae_{jk}$. Mostrare che $a_{ij} \in R = C(\{e_{kl} \mid k, l = 1, \ldots, n\})$ e che $a = \sum_{i,j} a_{ij}e_{ij}$. Mostrare che, per $r_{ij} \in R$, l'uguaglianza $\sum r_{ij}e_{ij} = 0$ può valere soltanto se ogni $r_{ij} = 0$. Dedurne che $S \cong M_n(R)$ ($\cong$ indica isomorfismo).

10. Siano R un anello, R' un insieme e η un'applicazione biiettiva di R' su R. Mostrare che R' diventa un anello ponendo

$$\begin{aligned} a' + b' &= \eta^{-1}(\eta(a') + \eta(b')), \quad & 0' &= \eta^{-1}(0), \\ a'b' &= \eta^{-1}(\eta(a')\eta(b')), & 1' &= \eta^{-1}(1), \end{aligned}$$

    e che η è un isomorfismo di R' con R. Usare ciò per provare che, se u è un elemento invertibile di un anello R, allora $(R, +, \cdot_u, 0, u^{-1})$, dove $a \cdot_u b = aub$, è un anello isomorfo a R. Mostrare inoltre che $(R, \oplus, \circ, 1, 0)$, dove $a \oplus b = a + b - 1$ e $a \circ b = a + b - ab$, è un anello isomorfo a R.

11. Mostrare che gli anelli $M_{nm}(R)$ e $M_n(M_m(R))$ sono isomorfi. (*Suggerimento:* usare l'addizione e la moltiplicazione delle matrici «a blocchi».)

12. Mostrare che, se R è un campo, $A \in M_n(R)$ è un divisore dello zero in questo anello se e solo se A non è invertibile. Vale lo stesso per un anello commutativo R arbitrario? Spiegare.

2.4 Quaternioni

Nel 1843 W. R. Hamilton costruì il primo esempio di corpo in cui la moltiplicazione non è commutativa. Si trattava di un'estensione del campo dei numeri complessi i cui elementi erano quadruple di numeri reali $(\alpha, \beta, \gamma, \delta)$. Si definivano la consueta addizione e una moltiplicazione tale che $1 = (1,0,0,0)$ fosse l'unità e che $i = (0,1,0,0)$, $j = (0,0,1,0)$, $k = (0,0,0,1)$ soddisfacessero $i^2 = j^2 = k^2 = -1 = ijk$.[3] Hamilton chiamò *quaternioni* le sue quadruple. In precedenza aveva definito i numeri complessi come coppie di reali (α, β), con il prodotto $(\alpha,\beta)(\gamma,\delta) = (\alpha\gamma - \beta\delta, \alpha\delta + \beta\gamma)$. La scoperta dei quaternioni suscitò numerosi esperimenti con altri sistemi di numeri «ipercomplessi» e, col tempo, diede origine a una teoria della struttura volta a classificarli. Da questa scoperta prese così le mosse una parte considerevole dell'algebra.

Non seguiremo la via di Hamilton. Definiremo i quaternioni come un particolare sottoanello dell'anello $M_2(\mathbb{C})$ delle matrici complesse 2×2. Avremo così il vantaggio di ridurre i calcoli a un'unica, semplice verifica.

Consideriamo il sottoinsieme $\mathbb{H}$ di $M_2(\mathbb{C})$ costituito dalle matrici della forma

$$x = \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} = \begin{pmatrix} \alpha_0 + \alpha_1\sqrt{-1} & \alpha_2 + \alpha_3\sqrt{-1} \\ -\alpha_2 + \alpha_3\sqrt{-1} & \alpha_0 - \alpha_1\sqrt{-1} \end{pmatrix}. \tag{14}$$

Asseriamo che $\mathbb{H}$ è un sottoanello. Poiché per i numeri complessi $\overline{a_1 - a_2} = \bar{a}_1 - \bar{a}_2$, $\mathbb{H}$ è chiuso rispetto alla sottrazione ed è quindi un sottogruppo del gruppo additivo di $M_2(\mathbb{C})$. Ponendo $a = 1$, $b = 0$ nella (14) si ottiene la matrice identità, dunque $1 \in \mathbb{H}$. Si ha

$$\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}\begin{pmatrix} c & d \\ -\bar{d} & \bar{c} \end{pmatrix} = \begin{pmatrix} ac - b\bar{d} & ad + b\bar{c} \\ -\bar{b}c - \bar{a}\bar{d} & -\bar{b}d + \bar{a}\bar{c} \end{pmatrix}.$$

Poiché $\overline{a_1 a_2} = \bar{a}_1\bar{a}_2$, il secondo membro ha la forma $\left(\begin{smallmatrix} u & v \\ -\bar{v} & \bar{u} \end{smallmatrix}\right)$, con $u = ac - b\bar{d}$, $v = ad + b\bar{c}$. Dunque $\mathbb{H}$ è chiuso rispetto al prodotto ed è un sottoanello di $M_2(\mathbb{C})$.

Dimostriamo ora che $\mathbb{H}$ è un corpo. Osserviamo anzitutto che

$$\Delta \equiv \det \begin{pmatrix} \alpha_0 + \alpha_1\sqrt{-1} & \alpha_2 + \alpha_3\sqrt{-1} \\ -\alpha_2 + \alpha_3\sqrt{-1} & \alpha_0 - \alpha_1\sqrt{-1} \end{pmatrix} = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2.$$

Gli α_i sono reali, quindi Δ è reale e si annulla solo quando ogni $\alpha_i = 0$, ossia quando la matrice è nulla. Ogni elemento non nullo di $\mathbb{H}$ ha dunque un inverso in $M_2(\mathbb{C})$. Inoltre, per la definizione di aggiunta data nel § 2.3,

$$\operatorname{adj}\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} = \begin{pmatrix} \bar{a} & -b \\ \bar{b} & a \end{pmatrix}.$$

Poiché $\bar{\bar{a}} = a$, si ottiene questa matrice da x nella (14) sostituendo $\bar{a}$ ad a e $-b$ a b. Essa appartiene perciò a $\mathbb{H}$. Se $x \neq 0$, il suo inverso è $\left(\begin{smallmatrix} \bar{a}\Delta^{-1} & -b\Delta^{-1} \\ \bar{b}\Delta^{-1} & a\Delta^{-1} \end{smallmatrix}\right)$, che appartiene a $\mathbb{H}$. Pertanto $\mathbb{H}$ è un corpo.

Il centro di $\mathbb{H}$ contiene il campo $\mathbb{R}$ dei reali, identificato con l'insieme delle matrici diagonali $\operatorname{diag}(\alpha, \alpha)$, $\alpha \in \mathbb{R}$. L'anello $\mathbb{H}$ contiene anche le matrici

$$i = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \qquad j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad k = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}.$$

[3] Pare che Hamilton abbia impiegato dieci anni per arrivare a questa tavola di moltiplicazione. Prima di rendersi conto che occorreva passare alle quadruple e abbandonare la commutatività del prodotto, aveva speso considerevoli energie nel tentativo, impossibile, di costruire un campo di triple di numeri reali. Questo episodio potrà forse incoraggiare lo studente a cui capiti di imboccare una strada sbagliata nell'affrontare un problema. Si veda Carl A. Boyer, *A History of Mathematics*, New York, Wiley, 1968, p. 625.

Si verifica che

$$x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k. \tag{15}$$

Se $\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 k$, con $\beta_i \in \mathbb{R}$, le matrici

$$\begin{pmatrix} \alpha_0 + \alpha_1\sqrt{-1} & \alpha_2 + \alpha_3\sqrt{-1} \\ -\alpha_2 + \alpha_3\sqrt{-1} & \alpha_0 - \alpha_1\sqrt{-1} \end{pmatrix}, \quad \begin{pmatrix} \beta_0 + \beta_1\sqrt{-1} & \beta_2 + \beta_3\sqrt{-1} \\ -\beta_2 + \beta_3\sqrt{-1} & \beta_0 - \beta_1\sqrt{-1} \end{pmatrix}$$

sono uguali, quindi $\alpha_i = \beta_i$ per $0 \le i \le 3$. Pertanto ogni $x \in \mathbb{H}$ si scrive in uno e un solo modo nella forma (15). Il prodotto di due elementi,

$$(\alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k)(\beta_0 + \beta_1 i + \beta_2 j + \beta_3 k),$$

è determinato dalla somma e dal prodotto in $\mathbb{R}$, dalle leggi distributive e dalla tavola di moltiplicazione

$$\begin{gathered} i^2 = j^2 = k^2 = -1, \\ ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j. \end{gathered} \tag{16}$$

Queste relazioni mostrano fra l'altro che $\mathbb{H}$ non è commutativo. Abbiamo così costruito un corpo che non è un campo. Lo si chiama *corpo dei quaternioni reali*.

Esercizi

1. Per $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$, si ponga $\bar{x} = \alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k$. Dimostrare che $\overline{x+y} = \bar{x} + \bar{y}$, $\overline{xy} = \bar{y}\bar{x}$ e che $\bar{x} = x$ se $x \in \mathbb{R}$.

2. Dimostrare che $x\bar{x} = N(x)$, dove $N(x) = \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2$. Si ponga $T(x) = 2\alpha_0$. Dimostrare che x soddisfa l'equazione quadratica $x^2 - T(x)x + N(x) = 0$.

3. Provare che $N(xy) = N(x)N(y)$.

4. Dimostrare che i quaternioni $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k$ le cui «coordinate» α_i sono razionali formano un sottocorpo $\mathbb{H}_0$ di $\mathbb{H}$.

5. Verificare che i quaternioni x le cui coordinate α_i sono tutte intere oppure tutte metà di interi dispari formano un sottoanello I di $\mathbb{H}$. Si tratta di un sottocorpo? Provare che $T(x), N(x) \in \mathbb{Z}$ per ogni $x \in I$. Determinare il gruppo degli invertibili di I.

6. Provare che il sottoanello di $M_2(\mathbb{C})$ generato da $\mathbb{C}$ e da $\mathbb{H}$ è $M_2(\mathbb{C})$.

7. Siano m, n interi non nulli e sia R il sottoinsieme di $M_2(\mathbb{C})$ formato dalle matrici

$$\begin{pmatrix} a + b\sqrt{m} & c + d\sqrt{m} \\ n(c - d\sqrt{m}) & a - b\sqrt{m} \end{pmatrix}, \qquad a, b, c, d \in \mathbb{Q}.$$

   Provare che R è un sottoanello e che è un corpo se e solo se gli unici razionali x, y, z, t che soddisfano $x^2 - my^2 - nz^2 + mnt^2 = 0$ sono $x = y = z = t = 0$. Scegliere m, n in modo che R sia un corpo e poi in modo che non lo sia.

8. Determinare il centro di $\mathbb{H}$ e il sottoanello $C(i)$ degli elementi che commutano con i.

9. Sia S un sottocorpo di $\mathbb{H}$ stabile rispetto a ogni applicazione $x \mapsto dxd^{-1}$ con $0 \ne d \in \mathbb{H}$. Dimostrare che $S = \mathbb{H}$ oppure S è contenuto nel centro.

10. (Cartan–Brauer–Hua.) Siano D un corpo, C il suo centro e S un sottocorpo di D stabile rispetto a ogni applicazione $x \mapsto dxd^{-1}$ con $0 \ne d \in D$. Dimostrare che $S = D$ oppure $S \subseteq C$.

2.5 Ideali e anelli quoziente

Una *congruenza* in un anello R è una relazione che sia una congruenza sia per il gruppo additivo $(R, +, 0)$ sia per il monoide moltiplicativo $(R, \cdot, 1)$. Dunque $\equiv$ è una relazione di equivalenza tale che $a \equiv a'$ e $b \equiv b'$ implichino $a + b \equiv a' + b'$ e $ab \equiv a'b'$. Indichiamo con $\bar{a}$ la classe di $a \in R$ e con $\bar{R}$ l'insieme quoziente. Come si è visto nel § 1.5, in $\bar{R}$ si hanno le operazioni binarie $\bar{a} + \bar{b} = \overline{a + b}$ e $\bar{a}\bar{b} = \overline{ab}$. Esse definiscono il gruppo $(\bar{R}, +, \bar{0})$ e il monoide $(\bar{R}, \cdot, \bar{1})$. Inoltre

$$\bar{a}(\bar{b} + \bar{c}) = \bar{a}\,\overline{b + c} = \overline{a(b + c)} = \overline{ab + ac} = \overline{ab} + \overline{ac} = \bar{a}\bar{b} + \bar{a}\bar{c}.$$

Analogamente $(\bar{b} + \bar{c})\bar{a} = \bar{b}\bar{a} + \bar{c}\bar{a}$. Pertanto $(\bar{R}, +, \cdot, \bar{0}, \bar{1})$ è un anello, detto *anello quoziente* (o *anello delle differenze*) di R.

Ricordiamo anche che le congruenze in $(R, +, 0)$ si ottengono dai sottogruppi I, necessariamente normali perché il gruppo è commutativo, ponendo $a \equiv b$ se $a - b \in I$. La classe $\bar{a}$ è allora il laterale $a + I$. Se la relazione è anche una congruenza del monoide moltiplicativo, per $a \in R$ e $b \in I$ si ha $a \equiv a$, $b \equiv 0$, dunque $ab \equiv a0 = 0$ e $ba \equiv 0$. In altre parole, $a \in R$, $b \in I$ implicano $ab, ba \in I$. Viceversa, sia I un sottogruppo del gruppo additivo che soddisfa questa condizione. Se $a \equiv a'$ e $b \equiv b' \pmod{I}$, allora $a - a' \in I$ e $ab - a'b = (a - a')b \in I$. Inoltre $a'b - a'b' = a'(b - b') \in I$. Pertanto $ab - a'b' = (ab - a'b) + (a'b - a'b') \in I$ e $ab \equiv a'b' \pmod{I}$.

Definizione 2.2. Un *ideale* I di un anello R è un sottogruppo del gruppo additivo tale che, per ogni $a \in R$ e $b \in I$, entrambi i prodotti ab e ba appartengano a I.

Le congruenze di R si ottengono dunque dai suoi ideali I ponendo $a \equiv a'$ se $a - a' \in I$. Indicheremo il corrispondente anello quoziente $\bar{R}$ con R/I e lo chiameremo *anello quoziente di R rispetto all'ideale I*. I suoi elementi sono i laterali $a + I$, con le operazioni

$$(a + I) + (b + I) = (a + b) + I, \qquad (a + I)(b + I) = ab + I. \tag{17}$$

Lo zero di R/I è I e la sua unità è $1 + I$.

Esaminiamo l'«algebra» degli ideali di R. Dalla definizione segue subito che l'intersezione di una famiglia qualsiasi di ideali è un ideale. Se $S \subseteq R$, l'intersezione (S) di tutti gli ideali contenenti S (la famiglia non è vuota, poiché contiene R) è un ideale che contiene S ed è contenuto in ogni ideale che contiene S. Si chiama *ideale generato da S*. Se $S = \{a_1, \dots, a_n\}$ è finito, scriviamo $(a_1, \dots, a_n)$. Descrivere tutti i suoi elementi non è immediato. Esso contiene ogni somma finita di prodotti xa_iy, con $x, y \in R$, e non si possono in generale raccogliere $xa_iy + x'a_iy'$ in un solo termine. Occorre dunque considerare tutte le espressioni

$$\sum_{i_1} x_{1i_1} a_1 y_{1i_1} + \sum_{i_2} x_{2i_2} a_2 y_{2i_2} + \cdots + \sum_{i_n} x_{ni_n} a_n y_{ni_n}. \tag{18}$$

L'insieme I degli elementi di forma (18) è un ideale. Contiene inoltre ogni $a_i = 1a_i1$. Quindi $I = (a_1, \dots, a_n)$.

Dati due ideali I, J, indichiamo con $I + J$ l'ideale generato da $I \cup J$. Esso è l'insieme K degli elementi $a + b$, $a \in I$, $b \in J$. Infatti K è un ideale, contiene I e J ed è contenuto in ogni ideale che li contenga entrambi. Un altro ideale importante è il *prodotto* IJ, ossia l'ideale generato dai prodotti ab, $a \in I$, $b \in J$. Si vede facilmente che IJ è l'insieme delle somme $a_1b_1 + \cdots + a_mb_m$ con $a_i \in I$ e $b_i \in J$.

Talvolta occorre considerare una successione di ideali $I_1 \subseteq I_2 \subseteq \cdots$, detta *catena ascendente di ideali*. La sua unione $\bigcup I_j$ è un ideale. Basta provare che è chiusa rispetto alla sottrazione e alla moltiplicazione a sinistra e a destra per elementi di R. Siano $a, b \in \bigcup I_j$. Allora $a \in I_j$ per

qualche j e $b \in I_k$ per qualche k. Posto $l = \max(j,k)$, sia a sia b appartengono a I_l. Quindi $a - b \in I_l$, poiché I_l è un ideale. Inoltre $xa, ax \in I_j$ per ogni $x \in R$. Ne segue che $a - b, xa, ax$ appartengono all'unione, come si voleva.

Se R è commutativo, la descrizione di $(a_1, \ldots, a_n)$ si semplifica considerevolmente. Dalla (18) segue che esso è l'insieme delle somme $\sum_{i=1}^n x_i a_i = \sum_{i=1}^n a_i x_i$, con $x_i \in R$. In particolare, (a) è l'insieme dei prodotti xa, $x \in R$, e si chiama *ideale principale generato da* a.

Gli ideali forniscono una semplice caratterizzazione dei campi.

Teorema 2.2. *Sia $R \neq 0$ un anello commutativo. Allora R è un campo se e solo se i suoi soli ideali sono $R = (1)$ e $0 = (0)$.*

Dimostrazione. Sia R un corpo e sia I un suo ideale non nullo. Preso $0 \neq a \in I$, anche $1 = aa^{-1} \in I$. Il solo ideale che contiene 1 è R, poiché contiene ogni $x = x1$. Quindi $I = R$. In un corpo i soli ideali sono perciò 0 e R. Ciò vale in particolare per i campi. Viceversa, sia $R \neq 0$ commutativo con soli ideali $0, R$. Per $0 \neq a \in R$ si ha $(a) \neq 0$, quindi $(a) = R$. Ne segue $1 \in (a)$ ed esiste $x \in R$ tale che $ax = 1$. Ogni elemento non nullo è invertibile e R è un campo. □

Esercizi

1. Sia Γ l'anello delle funzioni reali continue su $[0,1]$ (esempio 8, p. 87 dell'originale). Per $S \subseteq [0,1]$, sia $Z_S = \{f \mid f(x) = 0 \text{ per ogni } x \in S\}$. Verificare che Z_S è un ideale. Posti $S_1 = [0, \frac{1}{2}]$, $S_2 = [\frac{1}{2}, 1]$, $I_1 = Z_{S_1}$, $I_2 = Z_{S_2}$, dimostrare che $I_1 I_2 = I_1 \cap I_2 = 0$.

2. Verificare l'associatività del prodotto degli ideali, $(IJ)K = I(JK)$ per ideali I, J, K.

3. Vale la legge distributiva $I(J + K) = IJ + IK$?

4. In un anello R, un *ideale destro* (rispettivamente *sinistro*) è un sottogruppo additivo I tale che $ba \in I$ (rispettivamente $ab \in I$) per ogni $a \in R$, $b \in I$. Verificare che in $M_2(R)$, per ogni R, le matrici $\left(\begin{smallmatrix} 0 & 0 \\ a & b \end{smallmatrix}\right)$ formano un ideale destro e le matrici $\left(\begin{smallmatrix} a & 0 \\ b & 0 \end{smallmatrix}\right)$ un ideale sinistro. Qualcuno dei due insiemi è un ideale bilatero?

5. Dimostrare la seguente estensione del teorema 2.2. Un anello $R \neq 0$ è un corpo se e solo se 0 e R sono i suoi soli ideali sinistri (destri).

6. Sia R commutativo e sia N l'insieme dei suoi elementi nilpotenti. Provare che N è un ideale e che R/N non ha elementi nilpotenti non nulli.

7. Siano I un ideale di R e U il gruppo degli invertibili. Sia U_1 il sottoinsieme degli $a \in U$ tali che $a \equiv 1 \pmod{I}$. Provare che U_1 è un sottogruppo normale di U.

8. Sia I un ideale di R e sia $M_n(I)$ l'insieme delle matrici $n \times n$ con elementi in I. Dimostrare che $M_n(I)$ è un ideale di $M_n(R)$, che ogni ideale di $M_n(R)$ è di questa forma per qualche ideale I di R, e che $I \mapsto M_n(I)$ è una biezione fra l'insieme degli ideali di R e quello degli ideali di $M_n(R)$.

2.6 Ideali e anelli quoziente di $\mathbb{Z}$

Dopo le considerazioni generali della sezione precedente, esaminiamo gli ideali di $\mathbb{Z}$ e i corrispondenti quozienti $\mathbb{Z}/I$. Ne ricaveremo alcuni risultati di teoria dei numeri.

Come abbiamo visto nei §§ 1.5 e 1.10, i sottogruppi di $(\mathbb{Z}, +, 0)$ sono i gruppi ciclici $\langle k \rangle$ con k intero non negativo. Poiché $\langle k \rangle = \{xk \mid x \in \mathbb{Z}\}$, questo gruppo coincide con l'ideale principale

(k) dei multipli di k. Ogni ideale è un sottogruppo, quindi ogni ideale di $\mathbb{Z}$ è principale. Ora $(l) \supseteq (k)$ se e solo se $k \in (l)$, cioè $k = lm$ con $m \in \mathbb{Z}$. L'inclusione $(l) \supseteq (k)$ equivale pertanto alla divisibilità $l \mid k$. Di conseguenza, per $m, n \in \mathbb{Z}$, l'ideale (m, n) da essi generato è (d), dove d è un massimo comune divisore di m e n. Infatti (m, n) contiene (m) e (n), onde $d \mid m$ e $d \mid n$. D'altra parte, se $e \mid m$ ed $e \mid n$, allora (e) contiene (m) e (n) e perciò anche $(m, n) = (d)$, ossia $e \mid d$. Analogamente $(m) \cap (n) = ([m, n])$, dove $[m, n]$ è un minimo comune multiplo di m, n.

Passiamo al quoziente $\mathbb{Z}/(k)$, detto *anello delle classi di resto modulo k*. Siccome $(k) = (-k)$, possiamo supporre $k \geq 0$. Per $k = 0$ il quoziente si identifica con $\mathbb{Z}$. Per $k > 0$ i suoi elementi sono i k laterali

$$\bar{0} = (k), \quad \bar{1} = 1 + (k), \quad \bar{2} = 2 + (k), \quad \ldots, \quad \overline{k-1} = k - 1 + (k).$$

Supponiamo dapprima k composto, $k = lm$ con $l, m > 1$. Nel quoziente $\bar{l} \neq \bar{0}$ e $\bar{m} \neq \bar{0}$, mentre $\bar{l}\bar{m} = \bar{k} = \bar{0}$. Per k composto vi sono dunque divisori dello zero non nulli. Sia ora $k = p$ primo. Ogni $\bar{a} \neq \bar{0}$ in $\mathbb{Z}/(p)$ è invertibile. Poiché il quoziente è commutativo ($\bar{a}\bar{b} = \overline{ab} = \overline{ba} = \bar{b}\bar{a}$), ne segue che è un campo. Infatti, dato $\bar{a} \neq \bar{0}$, si ha $p \nmid a$, dunque 1 è un massimo comune divisore di p, a. Esistono interi x, y con $ax + py = 1$, onde $\bar{1} = \overline{ax + py} = \overline{ax} + \overline{py} = \bar{a}\bar{x}$ e $\bar{x}$ è l'inverso di $\bar{a}$. Questi semplici risultati meritano un enunciato a sé.

Teorema 2.3. *Per k composto, l'anello $\mathbb{Z}/(k)$ non è un dominio. Per p primo, $\mathbb{Z}/(p)$ è un campo.*

Determiniamo ora il gruppo $U(\mathbb{Z}/(k))$ degli elementi invertibili. Per $k = 0$ essi sono 1 e -1. Per $k > 0$ vale quanto segue.

Teorema 2.4. *Il gruppo $U(\mathbb{Z}/(k))$, con $k > 0$, è costituito dalle classi $\bar{a} = a + (k)$ per cui a e k sono primi fra loro, ossia hanno 1 come massimo comune divisore.*

Dimostrazione. Se $(a, k) = 1$ (in termini di ideali, $(a, k) = (1)$), esistono interi x, y tali che $ax + ky = 1$. Allora $\bar{a}\bar{x} = \bar{1}$ e $\bar{a}$ è invertibile. Viceversa, se $\bar{a}\bar{b} = \bar{1}$, allora $\overline{ab} = \bar{1}$ e $ab = 1 + mk$ con $m \in \mathbb{Z}$. Ogni divisore comune di a, k divide perciò 1, quindi a, k sono primi fra loro. □

Il risultato mostra che $|U(\mathbb{Z}/(k))|$ è il numero $\varphi(k)$ degli interi positivi minori di k e primi con k. La funzione così definita sugli interi positivi si chiama *funzione φ di Eulero* (si veda l'esercizio 4, p. 47 dell'originale). Ad esempio, gli invertibili di $\mathbb{Z}/(12)$ sono $\bar{1}, \bar{5}, \bar{7}, \overline{11}$, dunque $\varphi(12) = 4$. Nella sezione successiva daremo in un esercizio una formula per calcolare $\varphi(k)$ dalla scomposizione di k in primi. Osserviamo intanto che per p primo la definizione dà $\varphi(p) = p - 1$. È poi facile vedere che $\varphi(p^e) = p^e - p^{e-1} = p^e(1 - 1/p)$.

Ricordiamo che in un gruppo finito G si ha $a^{|G|} = 1$ per ogni $a \in G$. Da ciò e dal teorema 2.4 segue che se $(a, k) = 1$ allora $\bar{a}^{\varphi(k)} = \bar{1}$. La formulazione abituale è la seguente.

Teorema 2.5 (Eulero). *Se l'intero a è primo con l'intero positivo k, allora $a^{\varphi(k)} \equiv 1 \pmod{k}$.*

Per $k = p$ primo si ritrova un risultato precedente, dovuto a Fermat.

Corollario. *Se p è primo e l'intero a non è divisibile per p, allora $a^{p-1} \equiv 1 \pmod{p}$.*

Una formulazione leggermente diversa è $a^p \equiv a \pmod{p}$. Essa vale per ogni a, essendo evidente se $p \mid a$. D'altra parte, se $a^p \equiv a \pmod{p}$ e $a \not\equiv 0 \pmod{p}$, cancellando a si ottiene $a^{p-1} \equiv 1 \pmod{p}$. I due enunciati sono dunque equivalenti.

Esercizi

1. Scrivere le tavole di addizione e moltiplicazione di $\mathbb{Z}/(5)$ e $\mathbb{Z}/(6)$.
2. Dimostrare che $\mathbb{Z}/(k)$ contiene elementi nilpotenti non nulli ($z^n = 0$, $z \neq 0$) se e solo se k è divisibile per il quadrato di un primo. Determinare gli elementi nilpotenti di $\mathbb{Z}/(180)$.
3. Dimostrare che in un corpo finito D si ha $a^{|D|} = a$ per ogni $a \in D$.
4. Sia $A \in GL_2(\mathbb{Z}/(p))$, ossia una matrice invertibile 2×2 sul campo $\mathbb{Z}/(p)$. Dimostrare che $A^q = 1$ per $q = (p^2-1)(p^2-p)$. Dimostrare inoltre che $A^{q+2} = A^2$ per ogni $A \in M_2(\mathbb{Z}/(p))$.
5. Sia T l'insieme delle matrici triangolari $\left(\begin{smallmatrix} a & 0 \\ b & c \end{smallmatrix}\right)$ con $a, b, c \in \mathbb{Z}$. Verificare che T è un sottoanello di $M_2(\mathbb{Z})$. Determinare i suoi ideali.

2.7 Omomorfismi di anelli. Teoremi fondamentali

Definiamo ora gli omomorfismi di anelli e deduciamone le proprietà fondamentali. Tutto seguirà dai risultati sugli omomorfismi di monoidi e di gruppi dei §§ 1.9 e 1.10, poiché il punto di partenza è il seguente.

Definizione 2.3. Un *omomorfismo* di un anello R in un anello R' è un'applicazione di R in R' che è un omomorfismo sia del gruppo additivo sia del monoide moltiplicativo di R nelle rispettive strutture di R'.

Ricordando che η è un omomorfismo del gruppo G in G' se $\eta(ab) = \eta(a)\eta(b)$, vediamo che le condizioni affinché un'applicazione η dell'anello R nell'anello R' sia un omomorfismo sono

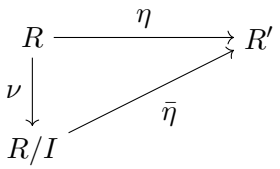
$$\eta(a+b) = \eta(a) + \eta(b), \qquad \eta(ab) = \eta(a)\eta(b), \qquad \eta(1) = 1',$$

dove $1'$ è l'unità di R'. A un ideale I di R corrispondono una congruenza e il quoziente $\bar{R} = R/I$. L'applicazione naturale $\nu : a \mapsto \bar{a}$ è un epimorfismo per i gruppi additivi e per i monoidi moltiplicativi, quindi è un epimorfismo (omomorfismo suriettivo) dell'anello R su $\bar{R}$. Come per i gruppi, chiamiamo *nucleo* di un omomorfismo η di R l'insieme $K = \eta^{-1}(0')$, con $0'$ lo zero di R'. La relazione $a \equiv b \pmod{K}$, ossia $a - b \in K$, è una congruenza. Il risultato del § 2.5 mostra quindi che K è un ideale di R, fatto verificabile anche direttamente. L'omomorfismo η è un monomorfismo (omomorfismo iniettivo) se e solo se il suo nucleo è 0. L'immagine di un omomorfismo di R in R' è un sottoanello di R', essendo un sottogruppo additivo e un sottomonoide moltiplicativo.

Sia ora $\eta : R \to R'$ un omomorfismo e sia I un ideale contenuto nel suo nucleo. Sappiamo allora che

$$\bar{\eta} : \bar{a} = a + I \longmapsto \eta(a)$$

è un omomorfismo di gruppi e di monoidi e dunque di anelli. Si chiama *omomorfismo (di anelli) indotto* da R/I in R'. Si ha il diagramma commutativo



e $\bar{\eta}$ è il solo omomorfismo da R/I a R' che lo renda commutativo. Inoltre $\bar{\eta}$ è un monomorfismo se e solo se I coincide con il nucleo di η. In questo caso otteniamo il seguente teorema.

Teorema fondamentale degli omomorfismi di anelli. *Sia $\eta : R \to R'$ un omomorfismo di anelli e sia $K = \eta^{-1}(0')$ il suo nucleo. Allora K è un ideale di R ed esiste un unico omomorfismo $\bar{\eta} : R/K \to R'$ tale che $\eta = \bar{\eta}\nu$, dove ν è l'omomorfismo naturale di R in R/K. Inoltre ν è un epimorfismo e $\bar{\eta}$ un monomorfismo.*

Ne segue immediatamente il corollario.

Corollario. *Ogni immagine omomorfa di un anello R è isomorfa a un suo quoziente R/K rispetto a un ideale K.*

La corrispondenza fra sottogruppi di un gruppo e della sua immagine omomorfa, data nel teorema 1.8′, si applica agli anelli attraverso i loro gruppi additivi.

Teorema 2.6. *Sia η un epimorfismo di R su R' con nucleo K. Nella corrispondenza biunivoca fra i sottogruppi H di $(R, +, 0)$ contenenti K e i sottogruppi additivi di R', data da $H \mapsto \eta(H)$, H è un sottoanello (un ideale) se e solo se $\eta(H)$ è un sottoanello (un ideale) di R'. Inoltre, se I è un ideale di R contenente K,*

$$a + I \longmapsto \eta(a) + I', \qquad I' = \eta(I), \tag{19}$$

è un isomorfismo di R/I su R'/I'.

Dimostrazione. L'immagine di un omomorfismo è un sottoanello, quindi se H è un sottoanello di R, $\eta(H)$ è un sottoanello di R'. Se H è un ideale di R, $\eta(H)$ è un sottogruppo additivo di R'. Dati $h \in H$, $x' \in R'$, esiste x con $\eta(x) = x'$. Allora $\eta(h)x' = \eta(h)\eta(x) = \eta(hx) \in \eta(H)$ e analogamente $x'\eta(h) \in \eta(H)$. Quindi $\eta(H)$ è un ideale. Se H' è un sottoanello (un ideale) di R', il sottogruppo additivo $\eta^{-1}(H')$ è evidentemente un sottoanello (un ideale) di R. La corrispondenza biunivoca fra i sottogruppi additivi di R che contengono K e i sottogruppi additivi di R' induce dunque corrispondenze biunivoche fra i rispettivi sottoanelli e fra i rispettivi ideali. Dai risultati sui gruppi sappiamo inoltre che (19) è un isomorfismo dei gruppi additivi di R/I e di R'/I' quando $I \supseteq K$, $I' = \eta(I)$. Poiché

$$(a + I)(b + I) = ab + I \longmapsto \eta(ab) + I' = \eta(a)\eta(b) + I' = (\eta(a) + I')(\eta(b) + I'),$$

la (19) è un isomorfismo di anelli. □

L'isomorfismo fra R/I e R'/I' del teorema precedente viene talvolta chiamato *primo teorema di isomorfismo per gli anelli*. Come nel caso dei gruppi, si ha anche il seguente risultato.

Secondo teorema di isomorfismo per gli anelli. *Siano R un anello, S un suo sottoanello e I un suo ideale. Allora $S + I = \{s + i \mid s \in S,\ i \in I\}$ è un sottoanello di R che contiene I come ideale, $S \cap I$ è un ideale di S e si ha l'isomorfismo*

$$s + I \longmapsto s + (S \cap I), \qquad s \in S, \tag{20}$$

di $(S + I)/I$ su $S/(S \cap I)$.

Dimostrazione. Si verifica direttamente che $S + I$ è un sottoanello. Evidentemente I è un ideale di $S + I$. L'omomorfismo $s \mapsto s + I$ di S in R/I è la restrizione a S dell'omomorfismo naturale di R in R/I. La sua immagine è $(S + I)/I$ e il suo nucleo è l'insieme degli s tali che $s + I = I$, ossia $S \cap I$. Ne segue l'isomorfismo $s + (S \cap I) \mapsto s + I$ di $S/(S \cap I)$ su $(S + I)/I$. L'isomorfismo (20) è la sua inversa. □

Applichiamo ora il teorema fondamentale degli omomorfismi per individuare il più piccolo sottoanello di R, quello generato da 1. Lo chiameremo *sottoanello fondamentale* o *anello primo*

di R (benché possa non avere nulla a che fare con i numeri primi). Dovendo usare $\mathbb{Z}$ con la sua unità 1, per il momento indichiamo con e l'unità di R. Consideriamo $n \mapsto ne$ da $\mathbb{Z}$ in R. In R valgono (§ 2.1)

$$(n+m)e = ne + me, \qquad (nm)e = (nm)e^2 = (ne)(me).$$

Poiché $1 \mapsto e$, l'applicazione è un omomorfismo. L'immagine $\mathbb{Z}e = \{ne \mid n \in \mathbb{Z}\}$ è dunque un sottoanello di R. Se S è un sottoanello qualsiasi di R, allora $e \in S$ e $\mathbb{Z}e \subseteq S$. Pertanto $\mathbb{Z}e$ è l'anello primo. Considerando $\mathbb{Z}e$ come codominio si ottiene un epimorfismo, dunque $\mathbb{Z}e \cong \mathbb{Z}/K$ per un ideale $K = (k)$ di $\mathbb{Z}$, $k \geq 0$. Per $k = 0$ si ha $\mathbb{Z}e \cong \mathbb{Z}$, per $k > 0$ si ha l'anello dei resti modulo k. Possiamo tornare a indicare con 1 l'unità di R e identificare l'anello primo con l'anello $\mathbb{Z}$ o $\mathbb{Z}/(k)$ a cui è isomorfo. Con questa convenzione si ha il teorema che segue.

Teorema 2.7. *L'anello primo di un anello R è $\mathbb{Z}$ oppure l'anello $\mathbb{Z}/(k)$ dei resti modulo un intero $k > 0$.*

Per k composto, $\mathbb{Z}/(k)$ ha divisori dello zero non nulli. Se R è un dominio, il suo anello primo è dunque $\mathbb{Z}$ oppure $\mathbb{Z}/(p)$ con p primo. Diremo che R ha *caratteristica* k se il suo anello primo è $\mathbb{Z}/(k)$, $k \geq 0$, con $\mathbb{Z}/(0) = \mathbb{Z}$. La caratteristica di un dominio è quindi 0 o un primo p. Se un anello ha caratteristica $k > 0$, per ogni a dell'anello si ha $ka = (k1)a = 0$. Chiaramente k è il più piccolo intero positivo con questa proprietà.

Esercizi

1. Se $\eta : R \to R'$ e $\zeta : R' \to R''$ sono omomorfismi di anelli, provare che $\zeta\eta : R \to R''$ è un omomorfismo.

2. Provare che, se u è invertibile in R e $\eta : R \to R'$ è un omomorfismo, $\eta(u)$ è invertibile in R'. Se η è suriettivo, l'omomorfismo del gruppo degli invertibili di R in quello di R' è anch'esso suriettivo?

3. Siano I un ideale di R e n un intero positivo. Applicare il teorema fondamentale degli omomorfismi per provare che $M_n(R)/M_n(I) \cong M_n(R/I)$.

4. Se R è un anello commutativo di caratteristica prima p, provare che $a \mapsto a^p$ è un endomorfismo di R, ossia un omomorfismo di R in sé. È anche un automorfismo?

5. Sia F un campo finito di caratteristica p prima. Provare che $p - 1 \mid |F| - 1$. Dedurne che se $|F|$ è pari la caratteristica è due. (Vedremo più avanti che $|F|$ è una potenza di p.)

6. Un anello R è *semplice* se $R \neq 0$ e i suoi soli ideali sono R e 0. Dimostrare che la caratteristica di un anello semplice è 0 o un primo p.

7. Se S è un sottoinsieme di un anello (campo) R, il *sottoanello* (*sottocampo*) *generato da* S è l'intersezione di tutti i sottoanelli (sottocampi) che contengono S. Se esso coincide con R, si dice che S è un *insieme di generatori* dell'anello (campo). Provare che due omomorfismi η_1, η_2 da R in un secondo anello (campo) coincidono se $\eta_1(s) = \eta_2(s)$ per ogni s di un insieme di generatori di R.

8. Dimostrare che ogni omomorfismo di un corpo in un anello $R \neq 0$ è iniettivo.

9. Per anelli $R_1, \ldots, R_n$, la *somma diretta* $R_1 \oplus \cdots \oplus R_n$ si definisce come per i monoidi e i gruppi. L'insieme sottostante è $R = R_1 \times \cdots \times R_n$. Le operazioni e gli elementi distinti

sono

$$(a_1,\dots,a_n)+(b_1,\dots,b_n)=(a_1+b_1,\dots,a_n+b_n),$$
$$(a_1,\dots,a_n)(b_1,\dots,b_n)=(a_1b_1,\dots,a_nb_n),$$
$$0=(0_1,\dots,0_n),\qquad 1=(1_1,\dots,1_n),$$

dove $0_i, 1_i$ sono lo zero e l'unità di R_i. Verificare che R è un anello. Provare che i suoi invertibili sono le n-uple $(u_1,\dots,u_n)$ con u_i invertibile in R_i. Dedurre che, posti $U=U(R)$, $U_i=U(R_i)$, si ha $U=U_1\times\cdots\times U_n$ e, per U_i finiti, $|U|=\prod|U_i|$.

10. (Teorema cinese del resto.) Due ideali I_1, I_2 di R siano primi fra loro nel senso che $I_1+I_2=R$. Dati $a_1,a_2\in R$, provare che esiste $a\in R$ con $a\equiv a_i \pmod{I_i}$ per $i=1,2$. Più in generale, se $I_1,\dots,I_m$ sono ideali tali che $I_j+\bigcap_{k\neq j}I_k=R$ per $1\le j\le m$, provare che per ogni $(a_1,\dots,a_m)$, $a_i\in R$, esiste $a\in R$ con $a\equiv a_k \pmod{I_k}$ per ogni k.

11. Usare il teorema cinese del resto e il teorema fondamentale degli omomorfismi per provare che, per ideali I_1, I_2 primi fra loro e $I=I_1\cap I_2$, si ha $R/I\cong R/I_1\oplus R/I_2$.

12. Usando l'esercizio 11, provare che $\varphi(mn)=\varphi(m)\varphi(n)$ per interi m,n primi fra loro, dove φ è la funzione di Eulero (p. 105 dell'originale). Verificare inoltre che per p primo $\varphi(p^e)=p^e-p^{e-1}$. Dedurre che, se $n=p_1^{e_1}\cdots p_r^{e_r}$ con p_i primi distinti, allora

$$\varphi(n)=\prod_{i=1}^{r}(p_i^{e_i}-p_i^{e_i-1})=n\prod_{i=1}^{r}\left(1-\frac{1}{p_i}\right).$$

13. Provare che l'unico omomorfismo di anelli di $\mathbb{R}$ in sé è l'identità.

14. Sia R l'anello delle funzioni reali continue su $[0,1]$ (esempio 8, p. 87 dell'originale). Osservare che per $0\le t\le 1$ la valutazione $\eta_t : f\mapsto f(t)$ è un omomorfismo di R in $\mathbb{R}$. Provare che ogni omomorfismo $\eta : R\to\mathbb{R}$ ha questa forma. (Suggerimento. Se $\eta\neq\eta_t$, esiste $f_t\in R$ con $\eta(f_t)\neq\eta_t(f_t)=f_t(t)$. Allora $g_t=f_t-\eta(f_t)1\in R$, $g_t(t)\neq 0$, ma $\eta(g_t)=0$. Mostrare che esistono finiti t_i tali che $g(x)=\sum g_{t_i}^2(x)\neq 0$ per ogni x. Allora $g^{-1}\in R$, ma $\eta(g)=0$.)

15. Un *ideale massimale* di R è un ideale proprio I tale che non esista un ideale proprio $I'\supsetneq I$. Dimostrare che un ideale I di un anello commutativo R è massimale se e solo se R/I è un campo.

16. Un *ideale primo* I di un anello commutativo R è caratterizzato dalle condizioni $I\neq R$ e, se $ab\in I$, allora $a\in I$ oppure $b\in I$. Provare che ogni ideale massimale è primo.

17. Determinare gli ideali di $\mathbb{Z}/(60)$ e individuare fra essi gli ideali massimali e quelli primi.

2.8 Anti-isomorfismi

Sia R un anello commutativo e sia $M_n(R)$ l'anello delle matrici $n\times n$ a coefficienti in R. Per $A=(a_{ij})\in M_n(R)$ chiamiamo *trasposta* di A la matrice ${}^tA$ che ha a_{ji} al posto (i,j). Essa si ottiene riflettendo gli elementi di A rispetto alla diagonale principale. Ad esempio,

$$A=\begin{pmatrix}1&2&3\\2&-1&4\\5&-1&6\end{pmatrix},\qquad {}^tA=\begin{pmatrix}1&2&5\\2&-1&-1\\3&4&6\end{pmatrix}.$$

È chiaro che ${}^t({}^tA)=A$, quindi $A\mapsto {}^tA$ è biiettiva. Inoltre, se $A=(a_{ij})$ e $B=(b_{ij})$, allora $A+B=(a_{ij}+b_{ij})$, quindi l'elemento (i,j) di ${}^t(A+B)$ è $a_{ji}+b_{ji}$. Pertanto ${}^t(A+B)={}^tA+{}^tB$,

e la trasposizione $t : A \mapsto {}^tA$ è un automorfismo del gruppo additivo di $M_n(R)$. Ovviamente ${}^t1 = 1$. Per $P = AB$ si ha $p_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$, dunque l'elemento (i, j) di ${}^tP$ è $\sum_{k=1}^n a_{jk}b_{ki}$. Quello di ${}^tB\,{}^tA$ è $\sum_{k=1}^n b_{ki}a_{jk} = \sum_{k=1}^n a_{jk}b_{ki}$ per la commutatività di R. Abbiamo provato che

$$ {}^t(AB) = ({}^tB)({}^tA). \tag{21}$$

Un'applicazione $x \mapsto x^*$ di R in sé che sia un automorfismo del gruppo additivo, mandi 1 in 1 e inverta l'ordine dei fattori, $(xy)^* = y^*x^*$, è detta *anti-automorfismo* di R. Se inoltre $x^{**} = x$ per ogni $x \in R$, la si chiama *involuzione*. I nostri calcoli mostrano che la trasposizione in $M_n(R)$, per R commutativo, è un'involuzione.

Un altro importante esempio è l'applicazione

$$ x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \longmapsto \bar{x} = \alpha_0 - \alpha_1 i - \alpha_2 j - \alpha_3 k, \qquad \alpha_i \in \mathbb{R}, \tag{22}$$

nell'algebra dei quaternioni $\mathbb{H}$ di Hamilton. Lo si può verificare direttamente, oppure dedurlo dalla proprietà di anti-automorfismo della trasposizione, come ora mostreremo. Anzitutto, se u è un elemento invertibile di un anello, $x \mapsto uxu^{-1}$ è un automorfismo. Come per i gruppi, questi automorfismi si chiamano *interni*. Osserviamo poi che componendo un automorfismo con un anti-automorfismo, in qualunque ordine, si ottiene un anti-automorfismo. Ne segue che

$$ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \equiv \operatorname{adj} \begin{pmatrix} a & b \\ c & d \end{pmatrix} $$

è un anti-automorfismo di $M_2(R)$. La formula mostra inoltre che $\operatorname{adj}(\operatorname{adj} A) = A$, quindi l'applicazione «aggiunta» è un'involuzione. Poniamo ora $R = \mathbb{C}$ e richiamiamo la definizione di $\mathbb{H}$ come sottoanello di $M_2(\mathbb{C})$ delle matrici $\left(\begin{smallmatrix} a & b \\ -\bar{b} & \bar{a} \end{smallmatrix}\right)$. Ricordiamo anche che

$$ i = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}, \quad j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad k = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}. $$

Si ha $\operatorname{adj} i = -i$, $\operatorname{adj} j = -j$, $\operatorname{adj} k = -k$. Quindi l'involuzione $x \mapsto \operatorname{adj} x$ di $M_2(\mathbb{C})$ lascia stabile $\mathbb{H}$ e vi induce l'involuzione (22).

Un'applicazione $x \mapsto x'$ da un anello R a un anello R' si dice *anti-isomorfismo* se è un isomorfismo dei gruppi additivi e soddisfa

$$ (xy)' = y'x', \qquad 1 \longmapsto 1', \text{ unità di } R'. \tag{23}$$

Se una tale applicazione esiste, R e R' si dicono *anti-isomorfi*. Talvolta è utile disporre di un anello anti-isomorfo a un anello dato R. Lo si costruisce facilmente mantenendo lo stesso insieme R, la stessa somma, 1 e 0, e definendo un nuovo prodotto con l'invertire l'ordine dei fattori prima di moltiplicare in R. Indicandolo con $\times$, poniamo

$$ a \times b = ba. \tag{24}$$

Allora

$$ (a \times b) \times c = ba \times c = c(ba), \qquad a \times (b \times c) = a \times cb = (cb)a, $$

e

$$ \begin{aligned} a \times (b + c) &= (b + c)a = ba + ca = a \times b + a \times c, \\ (b + c) \times a &= a(b + c) = ab + ac = b \times a + c \times a. \end{aligned} $$

Inoltre $a \times 1 = 1a = a = a1 = 1 \times a$. Dunque $(R, +, \times, 0, 1)$ è un anello. Per distinguerlo da $R = (R, +, \cdot, 0, 1)$, lo indicheremo con R^0 e lo chiameremo *anello opposto* di R. L'identità è un anti-isomorfismo di R su R^0. Ogni anti-isomorfismo di R equivale inoltre a un isomorfismo di R^0.

Esercizi

1. Provare che l'identità di R è un anti-automorfismo se e solo se R è commutativo.

2. Dimostrare che $x = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 k \mapsto x^* = \alpha_0 - \alpha_1 i + \alpha_2 j + \alpha_3 k$ è un'involuzione di $\mathbb{H}$.

3. Sia $x \mapsto x'$ un anti-isomorfismo di R su R'. Per $A = (a_{ij})$, si ponga $A^* = {}^t(a'_{ij})$. Verificare che $A \mapsto A^*$ è un anti-isomorfismo di $M_n(R)$ su $M_n(R')$.

4. Sia $a \mapsto a^*$ un anti-automorfismo di R. Siano $H = \{h \mid h^* = h\}$ gli elementi *simmetrici* o $*$-*simmetrici* e $K = \{k \mid k^* = -k\}$ gli elementi *antisimmetrici* o $*$-*antisimmetrici*. Verificare che H, K sono sottogruppi additivi di R. Posti $\{ab\} = ab + ba$, $[ab] = ab - ba$, dimostrare che, se $a, b, c \in H$, anche aba, a^n per $n \in \mathbb{N}$, $\{ab\}$, $abc+cba$, $[[ab]c]$ appartengono a H e che $[ab] \in K$. Se $a, b \in K$, provare che $[ab] \in K$ e, se $a \in H$, $b \in K$, che $[ab] \in H$.

5. Sia $u = \left(\begin{smallmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{smallmatrix}\right) \in M_3(\mathbb{Q})$ e siano

$$x = \begin{pmatrix} u & 0 \\ 0 & u^2 \end{pmatrix}, \qquad y = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

   dove $0, 1$ sono le matrici nulla e identità di $M_3(\mathbb{Q})$. Dunque $x, y \in M_6(\mathbb{Q})$. Verificare $x^3 = 0 = y^2$ e $yx = x^2 y$. Sia R il sottoanello di $M_6(\mathbb{Q})$ generato da $\mathbb{Q}, x, y$. Provare che ogni suo elemento ha forma $f(x) + g(x)y$ con $f(x) = a + bx + cx^2$, $g(x) = a' + b'x + c'x^2$, $a, b, c, a', b', c' \in \mathbb{Q}$, e che $(1, x, x^2, y, yx, yx^2)$ è una base di R come spazio vettoriale su $\mathbb{Q}$.[4] Dimostrare che, se x' è un elemento nilpotente di R e $y' \in R$ soddisfa $(y')^2 = 0$, allora $y'(x')^2 = 0$. Dedurre che R non ammette anti-automorfismi.

6. Si chiami *anti-omomorfismo* di R in R' un omomorfismo η dei gruppi additivi che manda 1 nell'unità $1'$ e soddisfa $\eta(ab) = \eta(b)\eta(a)$. Verificare che la composizione di un omomorfismo (un anti-omomorfismo) con un anti-omomorfismo (un omomorfismo) è un anti-omomorfismo e che la composizione di due anti-omomorfismi è un omomorfismo.

7. Si definisca *omomorfismo di Jordan* da R in R' un omomorfismo additivo η tale che $\eta(1) = 1$ e $\eta(aba) = \eta(a)\eta(b)\eta(a)$. Dimostrare che ogni omomorfismo e ogni anti-omomorfismo è un omomorfismo di Jordan. Provare che gli omomorfismi di Jordan soddisfano

$$\begin{aligned} \eta(a^k) &= \eta(a)^k, \quad k \in \mathbb{N}, \\ \eta(abc + cba) &= \eta(a)\eta(b)\eta(c) + \eta(c)\eta(b)\eta(a), \\ \eta(ab + ba) &= \eta(a)\eta(b) + \eta(b)\eta(a). \end{aligned}$$

8. (Jacobson e Rickart.) Sia η un omomorfismo di Jordan di un anello R in un dominio D. Provare che per ogni $a, b \in R$ vale $\eta(ab) = \eta(a)\eta(b)$ oppure $\eta(ab) = \eta(b)\eta(a)$.

9. (Hua.) Sia $\eta : R \to R'$ tale che $\eta(a + b) = \eta(a) + \eta(b)$, $\eta(1) = 1$ e, per ogni $a, b \in R$, $\eta(ab) = \eta(a)\eta(b)$ oppure $\eta(ab) = \eta(b)\eta(a)$. Dimostrare che η è un omomorfismo oppure un anti-omomorfismo.

10. (Jacobson e Rickart.) Provare che ogni omomorfismo di Jordan di un anello in un dominio è un omomorfismo oppure un anti-omomorfismo.

[4] La base è riportata come nell'originale. Dalle relazioni date risulta $yx^2 = 0$. La lista coerente con la forma $f(x) + g(x)y$ è $(1, x, x^2, y, xy, x^2y)$.

11. (Hua.) Sia η un'applicazione di un corpo D in un corpo D' tale che (i) η è un omomorfismo additivo, (ii) $\eta(1) = 1'$, (iii) $a \neq 0$ implica $\eta(a) \neq 0$ e $\eta(a)^{-1} = \eta(a^{-1})$. Provare che η è un omomorfismo oppure un anti-omomorfismo. (Suggerimento. Usare l'identità di Hua, esercizio 9, p. 92 dell'originale.)

2.9 Campo delle frazioni di un dominio commutativo

Abbiamo visto che ogni sottoanello di un corpo è un dominio. È naturale domandarsi se valga il viceversa. Ogni dominio D si può immergere in un corpo, ossia ammette un monomorfismo in un qualche corpo F? In tal caso D sarebbe isomorfo a un sottoanello D' di F e, identificandolo con D', potremmo considerarlo come un sottoanello di F. La questione rimase aperta per qualche tempo, finché A. Malcev diede una risposta negativa costruendo il primo esempio di dominio che non si può immergere in alcun corpo. Alcuni esercizi alla fine della sezione indicano l'esempio di Malcev. Qui ci interessa soprattutto il principale risultato positivo in questa direzione. Ogni dominio *commutativo* si può immergere in un campo. Il procedimento è precisamente quello, ben noto, con cui si costruiscono i numeri razionali a partire dagli interi. Per capire perché funzioni conviene prima esaminare la relazione fra un sottoanello D di un campo e il sottocampo F generato da D.

Sia dunque D un sottoanello di un campo e sia F il sottocampo da esso generato. Quali sono gli elementi di F? Se $a, b \in D$ e $b \neq 0$, certamente $ab^{-1} \in F$. Osserviamo ora che F è precisamente l'insieme degli elementi di questa forma. Infatti le uguaglianze

$$\begin{aligned} ab^{-1} + cd^{-1} &= adb^{-1}d^{-1} + cbb^{-1}d^{-1} = (ad + bc)(bd)^{-1}, \\ 0 = 0b^{-1}, &\qquad -ab^{-1} = (-a)b^{-1}, \\ (ab^{-1})(cd^{-1}) &= acb^{-1}d^{-1} = ac(bd)^{-1}, \\ 1 = aa^{-1} \quad (a \neq 0), &\qquad (ab^{-1})^{-1} = ba^{-1} \quad (a \neq 0) \end{aligned}$$

mostrano che $\{ab^{-1} \mid a, b \in D,\ b \neq 0\}$ è un sottocampo del campo dato. Si osservi che la commutatività del prodotto interviene più volte nei calcoli. Poiché F è generato da D, nessun suo sottocampo proprio contiene D. D'altra parte, l'insieme degli ab^{-1} contiene D attraverso gli elementi $a1^{-1} = a$. Pertanto

$$F = \{ab^{-1} \mid a, b \in D,\ b \neq 0\}. \tag{25}$$

Resta da chiedersi quando due elementi ab^{-1} e cd^{-1} siano uguali. Ciò accade se e solo se $ad = bc$, come si vede moltiplicando la prima uguaglianza per bd, o la seconda per $(bd)^{-1}$.

Sia ora dato un dominio commutativo D. Vogliamo immergerlo in un campo. Le osservazioni precedenti suggeriscono di costruire gli elementi di una minima estensione di D che sia un campo a partire dalle coppie (a, b) con $a, b \in D$, $b \neq 0$. La coppia (a, b) dovrà fare le veci di ab^{-1}. Procediamo dunque come segue.

Sia D^* l'insieme degli elementi non nulli di D. Poiché $D \neq 0$, si ha $D^* \neq \varnothing$. Sul prodotto $D \times D^*$ introduciamo la relazione $(a, b) \sim (c, d)$ se e solo se $ad = bc$. Si ha $(a, b) \sim (a, b)$ perché $ab = ba$. Se $(a, b) \sim (c, d)$, allora $ad = bc$, ossia $cb = da$, e $(c, d) \sim (a, b)$. Infine, se $(a, b) \sim (c, d)$ e $(c, d) \sim (e, f)$, allora $ad = bc$, $cf = de$, onde $adf = bcf = bde$. Poiché $d \neq 0$ e D è commutativo, si può cancellare d e si ottiene $af = be$, ossia $(a, b) \sim (e, f)$. La relazione è quindi un'equivalenza. Chiameremo *frazione* (o *quoziente*) a/b la classe di (a, b). Dunque $a/b = c/d$ se e solo se $ad = bc$. Sia $F = \{a/b\}$ l'insieme quoziente così ottenuto.

Introduciamo in F un'addizione, una moltiplicazione, uno zero e un'unità che ne facciano un campo. Per due frazioni $a/b, c/d$ si ha $bd \neq 0$, poiché $b, d \neq 0$, e si può dunque formare la

frazione $(ad+bc)/bd$. Se inoltre $a/b = a'/b'$ e $c/d = c'/d'$, allora

$$(ad+bc)/bd = (a'd'+b'c')/b'd'. \tag{26}$$

Infatti $ab' = ba'$, $cd' = dc'$, quindi $ab'dd' = ba'dd'$, $cd'bb' = dc'bb'$ e, sommando, $ab'dd' + cd'bb' = ba'dd' + dc'bb'$, ossia $(ad+bc)b'd' = (a'd'+b'c')bd$. Ciò prova la (26). Pertanto

$$a/b + c/d = (ad+bc)/bd \tag{27}$$

definisce un'operazione univoca in F. Analogamente ac/bd è una frazione. Se $a/b = a'/b'$ e $c/d = c'/d'$, da $ab' = ba'$ e $cd' = c'd$ segue $ab'cd' = ba'c'd$, dunque $ac/bd = a'c'/b'd'$. Perciò

$$(a/b)(c/d) = ac/bd \tag{28}$$

definisce una moltiplicazione univoca in F. Ponendo $0 = 0/1$ e $1 = 1/1$ si ottiene $a/b + 0 = a/b + 0/1 = (a1 + b0)/b1 = a/b$ e analogamente $0 + a/b = a/b$. Inoltre $(a/b)1 = a/b = 1(a/b)$. Una verifica diretta, che lasciamo al lettore, mostra che $(F, +, \cdot, 0, 1)$ è un anello commutativo. Se $a/b \neq 0$, allora $a \neq 0$, poiché $0/b = 0/1$ in virtù di $01 = 0b$. Quindi b/a è una frazione e $(a/b)(b/a) = ab/ab = 1/1 = 1$. Ogni frazione non nulla ha dunque inversa $(a/b)^{-1} = b/a$, e F è un campo.

Consideriamo ora l'applicazione di D in F

$$a \longmapsto a/1. \tag{29}$$

Essa manda 0 in 0 e 1 in 1. Inoltre $a + b \mapsto (a+b)/1 = a/1 + b/1$ e $ab \mapsto ab/1 = (a/1)(b/1)$. Quindi (29) è un omomorfismo. Se $a/1 = 0 = 0/1$, si ha $a1 = 10 = 0$, dunque $a = 0$. Il nucleo è nullo e (29) è un monomorfismo. Abbiamo così provato il teorema seguente.

Teorema 2.8. *Ogni dominio commutativo si può immergere in un campo.*

Identifichiamo ora a con $a/1$, come si identifica l'intero a con il razionale $a/1$. Allora D si identifica con un sottoanello di F e, per ogni $a/b \in F$, si ha $a/b = (a/1)(1/b) = (a/1)(b/1)^{-1} = ab^{-1}$. È dunque chiaro che D genera il campo F. Chiameremo F il *campo delle frazioni* di D. La sua proprietà fondamentale rispetto agli omomorfismi è data dal teorema che segue.

Teorema 2.9. *Siano D un dominio commutativo e F il suo campo delle frazioni. Ogni monomorfismo η_D di D in un campo F' si estende in uno e un solo modo a un monomorfismo η_F di F in F'.*

Dimostrazione. Indichiamo η_D con $a \mapsto a'$. Proviamo dapprima l'unicità di un eventuale prolungamento a un omomorfismo $\eta_F : F \to F'$. Per $b \neq 0$ esso deve mandare b^{-1} in $(b')^{-1}$ e dunque ab^{-1} in $a'(b')^{-1}$. Poiché ogni elemento di F ha forma ab^{-1}, il prolungamento deve essere l'applicazione $ab^{-1} \mapsto a'(b')^{-1}$. Resta da mostrare che questa è ben definita, è un monomorfismo ed estende η_D. Se $ab^{-1} = cd^{-1}$, allora $ad = bc$, dunque $a'd' = b'c'$ in F' e $a'(b')^{-1} = c'(d')^{-1}$. L'applicazione è quindi univoca. Per $a, b, c, d \in D$, $b, d \neq 0$, i calcoli

$$\begin{aligned} ab^{-1} + cd^{-1} &= (ad+bc)(bd)^{-1} \longmapsto (ad+bc)'((bd)')^{-1} \\ &= (a'd'+b'c')(b')^{-1}(d')^{-1} = a'(b')^{-1} + c'(d')^{-1}, \\ (ab^{-1})(cd^{-1}) &= ac(bd)^{-1} \longmapsto a'c'(b')^{-1}(d')^{-1} = (a'(b')^{-1})(c'(d')^{-1}) \end{aligned}$$

verificano la compatibilità con le operazioni. Inoltre $1 \mapsto 1'$, l'unità di F', perché D, F hanno la stessa unità e η_D è un omomorfismo. L'applicazione estende η_D, poiché $a = a1^{-1} \mapsto a'(1')^{-1} = a'$. Ogni omomorfismo di un campo in un campo è iniettivo (esercizio 8, p. 110 dell'originale). Abbiamo così il monomorfismo richiesto e ne abbiamo già provato l'unicità. □

Esercizi

1. Qual è il campo delle frazioni di un campo?

2. Siano D un dominio e F_1, F_2 campi che lo contengono come sottoanello e sono entrambi generati da D. Dimostrare che esiste un unico isomorfismo di F_1 su F_2 che sia l'identità su D.

3. Dimostrare che ogni monoide commutativo in cui vale la cancellazione ($ab = ac \Rightarrow b = c$) si può immergere in un gruppo abeliano.

4. Siano m, n interi positivi primi fra loro e a, b elementi di un dominio commutativo. Se $a^m = b^m$ e $a^n = b^n$, provare che $a = b$.

5. Sia R un anello commutativo e sia S un sottomonoide del suo monoide moltiplicativo. In $R \times S$ si ponga $(a, s) \sim (b, t)$ se esiste $u \in S$ con $u(at - bs) = 0$. Dimostrare che si tratta di una relazione di equivalenza. Indicare con a/s la classe di (a, s) e con RS^{-1} l'insieme quoziente. Provare che RS^{-1} è un anello rispetto a

$$a/s + b/t = (at + bs)/st, \qquad (a/s)(b/t) = ab/st, \qquad 0 = 0/1, \quad 1 = 1/1.$$

   Provare che $a \mapsto a/1$ è un omomorfismo di R in RS^{-1} e che è iniettivo se e solo se nessun elemento di S è un divisore dello zero di R. Mostrare che gli elementi $s/1$, $s \in S$, sono invertibili in RS^{-1}.

6. (Ore.) Sia D un dominio, anche non commutativo, con la proprietà del *multiplo comune destro*: ogni coppia di elementi non nulli $a, b \in D$ ha un multiplo comune destro non nullo $m = ab_1 = ba_1$. In $D \times D^*$, con D^* l'insieme degli elementi non nulli, si ponga $(a, b) \sim (c, d)$ se, per $b_1, d_1 \neq 0$ tali che $bd_1 = db_1$, si ha $ad_1 = cb_1$. Provare che la condizione è indipendente dalla scelta di b_1, d_1 e che $\sim$ è una relazione di equivalenza. Sia F l'insieme delle classi a/b. Provare che F è un corpo rispetto a

$$\begin{aligned} &a/b + c/d = (ad_1 + cb_1)/m, \quad m = bd_1 = db_1, \\ &0 = 0/1, \quad 1 = 1/1, \qquad (a/b)(c/d) = ac_1/db_1, \end{aligned}$$

   dove nell'ultima formula $b_1 \neq 0$ e $cb_1 = bc_1$. Mostrare che $a \mapsto a/1$ è un monomorfismo di D in F e che F è l'insieme degli $(a/1)(b/1)^{-1}$, $a, b \in D$, $b \neq 0$.

7. (Malcev.) Siano a_i, b_i, $1 \leq i \leq 4$, elementi di un gruppo tali che $a_1a_2 = a_3b_4$, $a_1b_2 = a_3b_4$, $b_1a_2 = b_3a_4$. Provare che $b_1b_2 = b_3b_4$.[5] Sia W il monoide libero sui generatori a_i, b_i, $1 \leq i \leq 4$ (p. 68 dell'originale), e sia $\equiv$ la minima congruenza, ossia l'intersezione di tutte le congruenze, che contenga le coppie (a_1a_2, a_3a_4), (a_1b_2, a_3b_4), (b_1a_2, b_3a_4). Posto $S = W/\equiv$, dimostrare che in S valgono le leggi di cancellazione ma che S non si può immergere in un gruppo.

8. (Malcev.) Sia $\mathbb{Z}[S]$ l'insieme delle combinazioni lineari a coefficienti interi degli elementi del monoide S dell'esercizio 7, con le consuete definizioni di uguaglianza, addizione, moltiplicazione, 0 e 1 (esercizio 8, p. 127 dell'originale). Dimostrare che $\mathbb{Z}[S]$ è un dominio che non si può immergere in un corpo.

[5] La prima relazione è stampata $a_1a_2 = a_3b_4$ nell'originale. La successiva presentazione del monoide e la proprietà di Malcev richiedono $a_1a_2 = a_3a_4$.

2.10 Anelli di polinomi

Per il resto del capitolo, salvo il § 2.17 e qualche esercizio, *tutti gli anelli saranno commutativi* e la parola «anello» sarà sinonimo di «anello commutativo».

Spesso interessa studiare un anello R' rispetto a un suo sottoanello R. A tal fine consideriamo i sottoanelli di R' generati da R e da sottoinsiemi U di R'. Indicheremo un tale sottoanello con $R[U]$ e diremo che si ottiene «aggiungendo» U a R. Dato un altro sottoinsieme V, $R[U][V]$ si ottiene aggiungendo V a $R[U]$. Esso coincide con $R[U \cup V]$, il sottoanello ottenuto aggiungendo $U \cup V$ a R. Infatti $R[U \cup V]$ contiene $R[U]$ e V, quindi contiene il sottoanello $R[U][V]$ da essi generato. Viceversa $R[U][V]$ contiene R e $U \cup V$, perciò contiene $R[U \cup V]$. Dunque $R[U][V] = R[U \cup V]$.

Ci interessano soprattutto i sottoanelli ottenuti aggiungendo insiemi finiti all'anello «di base» R. Per $U = \{u_1, \dots, u_n\}$ scriveremo $R[u_1, \dots, u_n]$. Per induzione sulla precedente osservazione si ottiene

$$R[u_1, \dots, u_n] = R[u_1][u_2] \cdots [u_n]. \tag{30}$$

Si procede cioè per aggiunzioni successive di singoli elementi ai sottoanelli già costruiti. È naturale cominciare da $R[u]$. Possiamo indicarne subito tutti gli elementi. Sono i *polinomi in u a coefficienti in R*, ossia gli elementi di forma

$$a_0 + a_1 u + a_2 u^2 + \cdots + a_n u^n, \qquad a_i \in R. \tag{31}$$

È chiaro che $R[u]$ li contiene tutti. Inoltre per due polinomi $\sum_{i=0}^{n} a_i u^i$, $\sum_{j=0}^{m} b_j u^j$ con $n \geq m$, si ha

$$\begin{aligned}(a_0 + a_1 u + \cdots + a_n u^n) + (b_0 + b_1 u + \cdots + b_m u^m) \\ = (a_0 + b_0) + (a_1 + b_1)u + \cdots + (a_m + b_m)u^m + a_{m+1}u^{m+1} + \cdots + a_n u^n.\end{aligned} \tag{32}$$

Poiché $(a_i u^i)(b_j u^j) = a_i b_j u^{i+j}$, la distributività dà

$$(a_0 + a_1 u + \cdots + a_n u^n)(b_0 + b_1 u + \cdots + b_m u^m) = p_0 + p_1 u + \cdots + p_{n+m} u^{n+m}, \tag{33}$$

dove

$$p_i = \sum_{j=0}^{i} a_j b_{i-j} = \sum_{j+k=i} a_j b_k. \tag{34}$$

Anche $0, 1$ sono polinomi in u e $-\sum a_i u^i = \sum(-a_i)u^i$. Questi polinomi formano dunque un sottoanello di R', che coincide con $R[u]$.

Le formule (32)–(34) spiegano come calcolare somme e prodotti di polinomi dati. Rimane però una difficoltà. Quando due espressioni polinomiali in u rappresentano lo stesso elemento? Espressioni diverse possono rappresentarlo. Ad esempio, se $u \in R$, eventualità non esclusa, l'elemento $u \in R[u]$ si rappresenta sia con $a_0 = u$ sia con $a_1 u$ per $a_1 = 1$. Un esempio meno banale si ha per $R' = \mathbb{C}$, $R = \mathbb{R}$, $u = \sqrt{-1}$, dove $u^2 = -1$.

Costruiremo ora un anello $R[x]$ in cui le sole relazioni $a_0 + a_1 x + \cdots = b_0 + b_1 x + \cdots$ siano quelle ovvie, con $a_i = b_i$ per ogni i. Intuitivamente cerchiamo l'insieme delle espressioni $a_0 + a_1 x + \cdots + a_n x^n$, $a_i \in R$, dove l'uguaglianza è *definita* dall'uguaglianza di tutti i coefficienti. Somma e prodotto saranno dati dalle (32)–(34) sostituendo x a u. Vogliamo dunque che un polinomio in x determini la successione dei propri coefficienti, tutti nulli da un certo indice in poi. Siamo condotti a *identificarlo* con una successione $(a_0, a_1, \dots, a_n, 0, 0, \dots)$ e a definire per queste successioni somma e prodotto secondo le (32)–(34).

Eseguiamo la costruzione nei dettagli. Dato un anello R, sia $R[x]$ l'insieme delle successioni infinite $(a_0, a_1, a_2, \ldots)$ con solo un numero finito di termini non nulli. Due successioni $(a_0, a_1, \ldots)$ e $(b_0, b_1, \ldots)$ sono uguali precisamente quando $a_i = b_i$ per ogni i. In altri termini, gli elementi sono le applicazioni $i \mapsto a_i$ dall'insieme $\mathbb{N}$ degli interi non negativi in R che si annullano per i sufficientemente grande. Per il momento la x nella notazione $R[x]$ non ha significato, ma tra breve comparirà una vera x a giustificarla. Introduciamo l'operazione

$$(a_0, a_1, \ldots) + (b_0, b_1, \ldots) = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \ldots),$$

il cui risultato appartiene evidentemente a $R[x]$, e lo zero $0 = (0, 0, 0, \ldots)$. È immediato che $(R[x], +, 0)$ è un gruppo abeliano. Introduciamo poi il prodotto

$$(a_0, a_1, a_2, \ldots)(b_0, b_1, b_2, \ldots) = (p_0, p_1, p_2, \ldots), \tag{35}$$

con p_i dato dalla (34). Se $a_i = 0$ per $i > n$ e $b_j = 0$ per $j > m$, allora $p_k = 0$ per $k > m + n$, quindi il secondo membro è in $R[x]$. Poniamo anche $1 = (1, 0, 0, \ldots)$. Allora $(a_0, a_1, \ldots)1 = (a_0, a_1, \ldots) = 1(a_0, a_1, \ldots)$. Se $A = (a_0, a_1, \ldots)$, $B = (b_0, b_1, \ldots)$ e $C = (c_0, c_1, \ldots)$, il termine di posto $i + 1$ di $(AB)C$ è

$$\sum_{m+l=i} \left(\sum_{j+k=m} a_j b_k \right) c_l = \sum_{j+k+l=i} (a_j b_k) c_l.$$

Il termine corrispondente di $A(BC)$ è

$$\sum_{m+j=i} a_j \left(\sum_{k+l=m} b_k c_l \right) = \sum_{j+k+l=i} a_j (b_k c_l).$$

L'associatività in R dà dunque $(AB)C = A(BC)$. In modo analogo si verificano le leggi distributive. La commutatività del prodotto segue dalla (34) e da quella di R. Pertanto $(R[x], +, \cdot, 0, 1)$ è un anello commutativo.

L'applicazione $a \mapsto a' = (a, 0, 0, \ldots)$ di R in $R[x]$ è chiaramente un monomorfismo. Identifichiamo a con a' e dunque R con la sua immagine, sottoanello di $R[x]$. Indichiamo ora con x l'elemento $(0, 1, 0, 0, \ldots)$. Dalla formula del prodotto, per induzione su $k \geq 0$, si vede che x^k è la successione con 1 al posto $k + 1$ e 0 altrove. Per $a \in R$, identificato con a', ax^k ha a al posto $k + 1$ e 0 altrove. Pertanto

$$(a_0, a_1, \ldots, a_n, 0, 0, \ldots) = a_0 + a_1 x + \cdots + a_n x^n,$$

e $R[x]$ è l'anello ottenuto aggiungendo x a R. Lo chiameremo *anello dei polinomi su R nell'indeterminata x*. L'ultima formula e la definizione di uguaglianza mostrano che $\sum a_i x^i = \sum b_i x^i$ implica $a_i = b_i$ per ogni i. In particolare $\sum a_i x^i = 0$ implica che ogni $a_i = 0$.

Costruito $R[x]$, possiamo servircene per studiare qualsiasi $R[u]$. Vedremo infatti che ogni $R[u]$ è un'immagine omomorfa di $R[x]$. Sarà dunque $R[u] \cong R[x]/I$ per un ideale I di $R[x]$. Il problema delle relazioni in $R[u]$ si risolve osservando che $a_0 + a_1 u + \cdots = b_0 + b_1 u + \cdots$ se e solo se $\sum a_i x^i \equiv \sum b_i x^i \pmod{I}$. Conoscere I equivarrà dunque a conoscere le relazioni. La proprietà fondamentale di $R[x]$ rispetto agli omomorfismi è la seguente.

Teorema 2.10. *Siano R, S anelli commutativi, $\eta : R \to S$ un omomorfismo e $u \in S$. Sia $R[x]$ l'anello dei polinomi su R nell'indeterminata x. Esiste uno e un solo prolungamento di η a un omomorfismo $\eta_u : R[x] \to S$ che mandi x in u.*

Dimostrazione. Per $A = a_0 + a_1x + \cdots + a_nx^n$ poniamo $\eta_u(A) = a'_0 + a'_1u + \cdots + a'_nu^n$, dove $a' = \eta(a)$. Se $B = b_0 + b_1x + \cdots + b_mx^m$, allora $AB = p_0 + p_1x + \cdots + p_{n+m}x^{n+m}$ con $p_i = \sum_{j+k=i} a_jb_k$. Quindi $\eta_u(AB) = p'_0 + p'_1u + \cdots + p'_{n+m}u^{n+m}$, dove $p'_i = \sum_{j+k=i} a'_jb'_k$ perché η è un omomorfismo. D'altra parte

$$\eta_u(A)\eta_u(B) = (a'_0+a'_1u+\cdots+a'_nu^n)(b'_0+b'_1u+\cdots+b'_mu^m) = p'_0+p'_1u+\cdots+p'_{n+m}u^{n+m} = \eta_u(AB).$$

È ancor più semplice verificare $\eta_u(A + B) = \eta_u(A) + \eta_u(B)$, e lasciamo il calcolo al lettore. Per $a \in R$ si ha $\eta_u(a) = a' = \eta(a)$, dunque η_u prolunga η. Inoltre $\eta_u(1) = \eta(1) = 1$, unità di S, e $\eta_u(x) = u$. Abbiamo quindi l'omomorfismo richiesto. Poiché $R[x]$ è generato da R e x, è l'unico con queste proprietà (esercizio 7, p. 110 dell'originale). □

Sia S un sovranello di R, cioè un anello che lo contiene come sottoanello, e sia $u \in S$. Il teorema dà un unico omomorfismo che coincide con l'identità su R e manda x in u. Scriveremo $A(x)$ per $A = a_0 + a_1x + \cdots + a_nx^n$ e $A(u)$ per la sua immagine. Useremo dunque la consueta notazione funzionale benché non si tratti propriamente di funzioni. Sarà anche comodo parlare di «sostituzione di u a x in $A(x)$» per indicare l'applicazione dell'omomorfismo di $R[x]$ in S che prolunga l'identità su R e manda x in u. Se I è il nucleo, $R[u] \cong R[x]/I$. Poiché su R l'omomorfismo è l'identità, $R \cap I = 0$. Otteniamo così una descrizione precisa degli anelli costruiti aggiungendo un singolo elemento.

Corollario. *$R[u] \cong R[x]/I$, dove x è un'indeterminata e I un ideale di $R[x]$ tale che $I \cap R = 0$.*

Viceversa, dato un ideale I con $I \cap R = 0$, la restrizione a R dell'omomorfismo naturale $\nu : R[x] \to R[x]/I$ è iniettiva. Possiamo identificare R con la sua immagine, ossia a con $a + I$. Allora R è un sottoanello di $R[x]/I$. Poiché $R[x]$ è generato da R e x, la sua immagine è generata da R e $u = x + I$. Pertanto $R[x]/I \cong R[u]$. □

L'omomorfismo $A(x) \mapsto A(u)$ è iniettivo se e solo se $A(u) = 0$ implica $A(x) = 0$, cioè $a_0 + a_1u + \cdots + a_nu^n = 0$ implica che tutti gli a_i siano nulli. In questo caso u si dice *trascendente* su R, altrimenti *algebrico* su R. Nel caso classico $S = \mathbb{R}$ (o $\mathbb{C}$) e $R = \mathbb{Q}$. Un numero reale o complesso si dice dunque algebrico o trascendente secondo che lo sia su $\mathbb{Q}$.

Passiamo da un elemento a un numero finito di elementi. Invertendo in parte l'ordine precedente, cominciamo direttamente dalla generalizzazione del teorema 2.10.

Teorema 2.11. *Per ogni anello R e ogni intero positivo r esiste un anello $R[x_1, \ldots, x_r]$ con la seguente proprietà «universale». Per ogni anello S, omomorfismo $\eta : R \to S$ e applicazione $i \mapsto u_i$ di $\{1, \ldots, r\}$ in S, esiste un unico omomorfismo $\eta_{u_1,\ldots,u_r} : R[x_1, \ldots, x_r] \to S$ che prolunga η e manda x_i in u_i, $1 \le i \le r$.*

Dimostrazione. Definiamo $R[x_1, \ldots, x_r]$ per induzione. $R[x_1]$ è l'anello dei polinomi nell'indeterminata x_1 su R e, in generale, $R[x_1, \ldots, x_i]$ è l'anello dei polinomi nell'indeterminata x_i su $R[x_1, \ldots, x_{i-1}]$. Per il teorema 2.10 vi è un omomorfismo $\eta_{u_1} : R[x_1] \to S$ che prolunga η e manda x_1 in u_1. Supponiamo per induzione di avere un omomorfismo di $R[x_1, \ldots, x_{r-1}]$ che prolunghi η e mandi x_i in u_i per $1 \le i < r$. Per il teorema 2.10 esso si prolunga a un omomorfismo $\eta_{u_1,\ldots,u_r}$ di $R[x_1, \ldots, x_r] = R[x_1, \ldots, x_{r-1}][x_r]$ in S che manda x_r in u_r. Si ottiene così il prolungamento desiderato. L'unicità segue poiché $R[x_1, \ldots, x_r]$ è generato da R e dalle x_i. □

Esiste essenzialmente un solo anello con la proprietà del teorema 2.11. Se $R[y_1, \ldots, y_r]$ è un altro tale anello, esiste un omomorfismo $\zeta : R[x_1, \ldots, x_r] \to R[y_1, \ldots, y_r]$ identico su R che manda x_i in y_i. Esiste anche un omomorfismo ζ' in senso opposto identico su R che manda y_i in x_i. Allora $\zeta'\zeta$ è un endomorfismo identico su R e sulle x_i, quindi è l'identità di $R[x_1, \ldots, x_r]$. Analogamente $\zeta\zeta'$ è l'identità su $R[y_1, \ldots, y_r]$. Dunque ζ e ζ' sono isomorfismi.

Chiameremo $R[x_1, \dots, x_r]$ *anello dei polinomi su* R *nelle* r *indeterminate* $x_1, \dots, x_r$. L'unicità essenziale appena provata rende la scelta della costruzione una questione puramente estetica (un'altra costruzione sarà indicata nell'esercizio 9). Benché il procedimento per aggiunzioni *successive* non tratti le indeterminate in modo simmetrico, il risultato finale è simmetrico. Precisamente vale il seguente teorema.

Teorema 2.12. *Sia* $R[x_1, \dots, x_r]$ *l'anello dei polinomi in* r *indeterminate su* R *e sia* π *una permutazione di* $1, \dots, r$. *Esiste un unico automorfismo* $\zeta(\pi)$ *di questo anello identico su* R *che mandi* x_i *in* $x_{\pi(i)}$ *per ogni* i.

Dimostrazione. Il teorema 2.11 dà un unico endomorfismo con queste proprietà. Mostriamo che è un automorfismo. Confrontando le immagini dei generatori $R \cup \{x_1, \dots, x_r\}$, per due permutazioni π_1, π_2 si ha $\zeta(\pi_1\pi_2) = \zeta(\pi_1)\zeta(\pi_2)$ e $\zeta(1) = 1$. Pertanto $\zeta(\pi)\zeta(\pi^{-1}) = 1 = \zeta(\pi^{-1})\zeta(\pi)$. □

A ogni $(i_1, \dots, i_r) \in \mathbb{N}^{(r)}$, ossia ogni r-upla di interi non negativi, associamo il *monomio* $x_1^{i_1} \cdots x_r^{i_r}$. Si ha

$$(x_1^{i_1} \cdots x_r^{i_r})(x_1^{j_1} \cdots x_r^{j_r}) = x_1^{i_1+j_1} \cdots x_r^{i_r+j_r}.$$

Come per $r = 1$, ne segue che $R[x_1, \dots, x_r]$ è l'insieme dei *polinomi* $\sum a_{i_1 \cdots i_r} x_1^{i_1} \cdots x_r^{i_r}$, somme finite con *coefficienti* in R. Ad esempio gli elementi di $R[x, y]$ sono i polinomi $a_{00} + a_{10}x + a_{01}y + a_{20}x^2 + a_{11}xy + a_{02}y^2 + \cdots$, $a_{ij} \in R$.

Mostriamo che a r-uple distinte corrispondono monomi distinti e che le sole relazioni fra monomi distinti sono quelle banali, con tutti i coefficienti nulli. Basta provare che, se $\sum_{(i)} a_{i_1 \cdots i_r} x_1^{i_1} \cdots x_r^{i_r} = 0$ per una somma su un numero finito di indici distinti $(i) = (i_1, \dots, i_r)$, ogni coefficiente è nullo. Ne seguirà infatti anche la prima affermazione, poiché due monomi uguali con indici diversi darebbero una relazione non banale con coefficienti $1, -1$. Per $r = 1$ lo abbiamo già dimostrato. Supponiamo dunque il risultato vero per $r - 1$. Possiamo scrivere

$$\sum_{(i)} a_{i_1 \cdots i_r} x_1^{i_1} \cdots x_r^{i_r} = \sum_{i_r} A_{i_r} x_r^{i_r}, \qquad A_{i_r} = \sum_{(i')} a_{i_1 \cdots i_{r-1} i_r} x_1^{i_1} \cdots x_{r-1}^{i_{r-1}},$$

dove i_r varia su un sottoinsieme finito di $\mathbb{N}$ e $(i') = (i_1, \dots, i_{r-1})$ su un insieme finito di indici distinti. Se la prima somma è nulla, anche $\sum A_{i_r} x_r^{i_r} = 0$, quindi ogni $A_{i_r} = 0$. Per induzione $a_{i_1 \cdots i_{r-1} i_r} = 0$ per ogni i_r fissato e ogni (i'), dunque per ogni (i).

Come per $r = 1$, l'omomorfismo di $R[x_1, \dots, x_r]$ su $R[u_1, \dots, u_r]$ che manda a in a per $a \in R$ e x_i in u_i è un isomorfismo se e solo se vale la seguente condizione di indipendenza: $\sum_{(i)} a_{i_1 \cdots i_r} u_1^{i_1} \cdots u_r^{i_r} = 0$ implica che tutti i coefficienti siano nulli. Si dice allora che $u_1, \dots, u_r$ sono *algebricamente indipendenti su* R. Questa proprietà delle x_i dà una nuova caratterizzazione di $R[x_1, \dots, x_r]$ come estensione di R.

Esercizi

1. Provare che il numero complesso $\omega = -\frac{1}{2} + \frac{1}{2}\sqrt{3}\,i$, $i = \sqrt{-1}$, è algebrico su $\mathbb{Q}$ e che $\mathbb{Q}[\omega] \cong \mathbb{Q}[x]/I$ con $I = (x^2 + x + 1)$.

2. Provare che $\sqrt{3} \notin \mathbb{Q}[\sqrt{2}]$ e che $1, \sqrt{2}, \sqrt{3}, \sqrt{6}$ sono linearmente indipendenti su $\mathbb{Q}$. Dimostrare che $u = \sqrt{2} + \sqrt{3}$ è algebrico e determinare un ideale I tale che $\mathbb{Q}[x]/I \cong \mathbb{Q}[u]$.

3. Sia I un ideale di R e sia $I[x_1, \dots, x_r]$ l'insieme dei polinomi di $R[x_1, \dots, x_r]$ a coefficienti in I. Provare che è un ideale e che $R[x_1, \dots, x_r]/I[x_1, \dots, x_r] \cong (R/I)[y_1, \dots, y_r]$ con y_i indeterminate su R/I.

4. Sia $\Delta = \prod_{i>j}(x_i - x_j) \in \mathbb{Z}[x_1, \ldots, x_r]$ e sia $\zeta(\pi)$ l'automorfismo che manda x_i in $x_{\pi(i)}$. (Ogni automorfismo di questo anello è l'identità su $\mathbb{Z}$. Perché?) Verificare che per una trasposizione τ si ha $\zeta(\tau)(\Delta) = -\Delta$. Dedurne il risultato del § 1.6: se π è prodotto di un numero pari di trasposizioni, ogni sua fattorizzazione in trasposizioni ha lunghezza pari. Provare che ogni $\zeta(\pi)$ fissa Δ^2.

5. Verificare che le costruzioni date per $R[x]$ e $R[x_1, \ldots, x_r]$ valgono anche se R non è commutativo. Mostrare che in tal caso le x_i sono nel centro di $R[x_1, \ldots, x_r]$. Enunciare e provare gli analoghi dei teoremi 2.10 e 2.11.

6. Provare che $M_n(R[x_1, \ldots, x_r]) \cong M_n(R)[x_1, \ldots, x_r]$, con le x_i indeterminate in entrambi i membri.

7. Sia $R[[x]]$ l'insieme di tutte le successioni $(a_0, a_1, a_2, \ldots)$, $a_i \in R$, senza restrizioni. Provare che con $+, \cdot, 0, 1$ definiti come nell'anello dei polinomi si ottiene un anello, detto *anello delle serie formali in un'indeterminata.*

8. Siano M un monoide, R un anello commutativo e $R[M]$ l'insieme delle applicazioni $m \mapsto f(m)$ da M in R che si annullano tranne che in un numero finito di punti. Si ponga

$$(f+g)(m) = f(m) + g(m), \quad (fg)(m) = \sum_{pq=m} f(p)g(q),$$
$$0(m) = 0, \quad 1(1) = 1, \quad 1(m) = 0 \ (m \neq 1).$$

   Provare che $R[M]$ è un anello. Le applicazioni a' con $a'(1) = a$ e $a'(m) = 0$ per $m \neq 1$ costituiscono un sottoanello isomorfo a R; le applicazioni m' con $m'(m) = 1$ e $m'(n) = 0$ per $n \neq m$ costituiscono un sottomonoide del monoide moltiplicativo di $R[M]$ isomorfo a M. Identificare queste strutture. Provare che R è nel centro di $R[M]$ e che ogni elemento di $R[M]$ si scrive come combinazione lineare $\sum r_i m_i$ di elementi di M a coefficienti in R. Provare che, per m_i distinti, $\sum r_i m_i = 0$ se e solo se ogni $r_i = 0$. Se $\sigma : R \to S$ è un omomorfismo di anelli con $\sigma(R)$ nel centro di S e τ è un omomorfismo di M nel monoide moltiplicativo di S, provare che c'è un unico omomorfismo di $R[M]$ in S che coincide con σ su R e con τ su M. Quando M è un gruppo, $R[M]$ si chiama *algebra del gruppo M su R.*

9. Sia R commutativo e sia $\mathbb{N}^{(r)}$ il monoide commutativo libero con r generatori x_i (p. 68 dell'originale). Mostrare che $R[\mathbb{N}^{(r)}]$, definito come nell'esercizio 8, coincide con $R[x_1, \ldots, x_r]$, con le x_i indeterminate.

10. Sia $M = FM^{(r)}$ il monoide libero con r generatori $x_1, \ldots, x_r$ (p. 68 dell'originale), e si costruisca $R[M]$ come nell'esercizio 8. Questa è l'*algebra libera su R generata dalle x_i.* Enunciare la sua proprietà fondamentale rispetto agli omomorfismi.

2.11 Alcune proprietà degli anelli di polinomi e loro applicazioni

Sia $R[x]$ l'anello dei polinomi in un'indeterminata x sull'anello commutativo R. Se $0 \neq f(x) \in R[x]$, possiamo scrivere

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n, \qquad a_n \neq 0. \tag{36}$$

Si dice che a_n è il *coefficiente direttivo* e n il *grado*, $\deg f$, di $f(x)$. Conviene assegnare a 0 il grado simbolico $-\infty$, con le convenzioni $-\infty < n$ per ogni $n \in \mathbb{N}$, $-\infty + (-\infty) = -\infty$ e

$-\infty + n = -\infty$. Si ha $f(x) \in R$ se e solo se $\deg f = 0$ o $-\infty$, e $f(x) \in R^*$, l'insieme degli elementi non nulli di R, se e solo se $\deg f = 0$. È inoltre chiaro che

$$\deg(f(x) + g(x)) \leq \max(\deg f(x), \deg g(x)), \tag{37}$$

con uguaglianza salvo eventualmente quando $\deg f = \deg g$. Se $g(x) = b_0 + b_1 x + \cdots + b_m x^m$ con $b_m \neq 0$ e $f(x)$ è dato da (36), allora

$$f(x)g(x) = a_0 b_0 + (a_0 b_1 + a_1 b_0)x + \cdots + a_n b_m x^{n+m}. \tag{38}$$

Se almeno uno fra a_n, b_m non è un divisore dello zero, $a_n b_m \neq 0$ e

$$\deg(f(x)g(x)) = \deg f(x) + \deg g(x). \tag{39}$$

Tenendo conto della convenzione su $-\infty$, la (39) vale per tutti i polinomi se $R = D$ è un dominio. In questo caso le proprietà del grado hanno la conseguenza seguente.

Teorema 2.13. *Se D è un dominio, lo è anche l'anello $D[x_1, \ldots, x_r]$ dei polinomi in r indeterminate su D. I suoi elementi invertibili sono precisamente quelli di D.*

Dimostrazione. Consideriamo prima $D[x]$. Se $f(x)g(x) = 0$, il grado del prodotto è $-\infty$. Per la (39) ciò è possibile solo se $\deg f = -\infty$ oppure $\deg g = -\infty$, ossia se $f = 0$ o $g = 0$. Se $fg = 1$, la (39) dà $\deg f = 0 = \deg g$. Un elemento invertibile di $D[x]$ e il suo inverso appartengono dunque a D. Gli invertibili di $D[x]$ sono quindi quelli di D. Entrambe le affermazioni si estendono a $D[x_1, \ldots, x_r]$ per induzione su r. □

Passiamo alla generalizzazione del familiare algoritmo di divisione dei polinomi. Di solito interesserà il caso in cui i coefficienti appartengono a un campo, ma talvolta occorre la seguente situazione più generale.

Teorema 2.14. *Siano $f(x), g(x) \in R[x]$ con $g(x) \neq 0$, e siano m il grado e b_m il coefficiente direttivo di g. Esistono $k \in \mathbb{N}$ e $q(x), r(x) \in R[x]$, con $\deg r < \deg g$, tali che*

$$b_m^k f(x) = q(x)g(x) + r(x). \tag{40}$$

Dimostrazione. Se $\deg f < \deg g$, basta scrivere $f = 0g + f$. Supponiamo dunque $n = \deg f \geq m = \deg g$ e poniamo

$$b_m f(x) - a_n x^{n-m} g(x) = f_1(x). \tag{41}$$

Nei due termini a sinistra il coefficiente di x^n è $a_n b_m$, quindi $\deg f_1 < \deg f$. L'induzione sul grado di f dà $k_1 \in \mathbb{N}$ e $q_1, r \in R[x]$ con $\deg r < \deg g$ tali che

$$b_m^{k_1} f_1(x) = g(x)q_1(x) + r(x). \tag{42}$$

Dalle (41) e (42) segue

$$b_m^{k_1+1} f(x) = b_m^{k_1} a_n x^{n-m} g(x) + g(x)q_1(x) + r(x) = g(x)q(x) + r(x),$$

con $q(x) = b_m^{k_1} a_n x^{n-m} + q_1(x)$. □

A proposito del teorema 2.14 osserviamo anzitutto che la dimostrazione dà un algoritmo per determinare $k, q(x), r(x)$ in un numero finito di passi. È la consueta divisione «in colonna» dei polinomi, come il lettore potrà verificare su qualche esempio. Si può sempre prendere k uguale al

maggiore fra 0 e $\deg f - \deg g + 1$. Inoltre, se b_m è invertibile, si può dividere per b_m^k e ottenere, con nuovi q, r,

$$f(x) = q(x)g(x) + r(x), \qquad \deg r < \deg g. \tag{40'}$$

Ciò accade sempre se $R = F$ è un campo. In questo caso il «quoziente» $q(x)$ e il «resto» $r(x)$ sono unici. Infatti da $f = qg + r = q_1 g + r_1$, con $\deg r, \deg r_1 < \deg g$, segue $(q - q_1)g = r_1 - r$. Se $q \neq q_1$, il grado del primo membro è almeno m, mentre il secondo ha grado minore di m, assurdo. Dunque $q = q_1$ e $r = r_1$. Pertanto g è un *divisore* o *fattore* di f, ossia esiste q con $f = gq$, se e solo se $r = 0$. Lo si accerta in un numero finito di passi con l'algoritmo di divisione. Infine, passando al campo delle frazioni, la (40′) equivale alla forse più familiare $f(x)/g(x) = q(x) + r(x)/g(x)$.

Corollario 1 (teorema del resto). *Dati $f(x) \in R[x]$ e $a \in R$, esiste un unico $q(x) \in R[x]$ tale che*

$$f(x) = (x - a)q(x) + f(a). \tag{43}$$

Dimostrazione. Per quanto precede ci sono un unico $q(x) \in R[x]$ e un $r \in R$ tali che $f(x) = (x - a)q(x) + r$. Sostituendo $x = a$, cioè applicando l'omomorfismo $R[x] \to R$ identico su R che manda x in a, si ottiene $f(a) = (a - a)q(a) + r = r$. Ne segue (43) con q unico. □

Dal corollario 1 segue immediatamente il prossimo.

Corollario 2 (teorema del fattore). *$(x - a) \mid f(x)$, cioè $x - a$ è un fattore di $f(x)$, se e solo se $f(a) = 0$.*

Applicheremo questi risultati a $F[x]$, dove F è un campo, e più generalmente a $F[u]$, anello generato da F e da un solo elemento u. Diremo che D è un *dominio a ideali principali* se ogni suo ideale è principale. Ricordiamo che questa proprietà vale per $\mathbb{Z}$ (§ 2.6). Dimostriamo ora il seguente risultato.

Teorema 2.15. *L'anello $F[x]$ dei polinomi in un'indeterminata su un campo F è un dominio a ideali principali.*

Dimostrazione. Sia I un ideale di $F[x]$. Per $I = 0$ si ha $I = (0)$. Se $I \neq 0$, i gradi dei suoi elementi non nulli sono interi non negativi, quindi esiste $g \neq 0$ in I di grado minimo. Per $f \in I$ l'algoritmo di divisione dà $f = qg + r$ con $\deg r < \deg g$. Poiché $f, g \in I$ e I è un ideale, $r = f - qg \in I$. Se $r \neq 0$ si contraddice la minimalità del grado di g, dunque $r = 0$ e $f = qg$. Ogni elemento di I è multiplo di g, e tutti questi multipli appartengono a I. Quindi $I = (g)$. Ciò vale per ogni ideale e $F[x]$ non ha divisori dello zero non nulli, come richiesto. □

Il risultato non si estende a più indeterminate. Per $r > 1$, ad esempio, i polinomi di $F[x_1, \ldots, x_r]$ privi di termine costante, ossia $\sum a_{i_1 \cdots i_r} x_1^{i_1} \cdots x_r^{i_r}$ con $a_{0 \cdots 0} = 0$, formano l'ideale $I = (x_1, \ldots, x_r)$. Se $I = (a)$, allora $a \mid x_i$ per ogni i. Poiché x_i è irriducibile, a è un invertibile o è associato a x_i. Siccome $r > 1$ e $I \neq (1)$, entrambe le possibilità sono escluse. Quindi I non è principale.

In $F[x]$ si ha $(f) \supseteq (g)$ se e solo se $g = fh$, ossia $f \mid g$. Se $f \mid g$ e $g \mid f$, si ha $g = fh$ e $f = gk$, dunque $g = gkh$. Per $g \neq 0$ segue $kh = 1$ e k, h sono elementi non nulli di F. Pertanto il generatore g di un ideale $(g) \neq 0$ è determinato a meno di un fattore invertibile. Possiamo normalizzarlo imponendo che il suo coefficiente direttivo sia 1. Questa proprietà lo determina univocamente. I polinomi con coefficiente direttivo 1 si diranno *monici*.

Consideriamo un anello $F[u]$, dove F è un campo. L'epimorfismo $f(x) \mapsto f(u)$ di $F[x]$ su $F[u]$ ha per nucleo un ideale I con $I \cap F = 0$ (§ 2.10). Ora $I = (g)$, e g non è invertibile

perché $I \cap F = 0$. Pertanto $g = 0$ oppure $\deg g > 0$. Nel primo caso $I = 0$, l'epimorfismo è un isomorfismo e u è trascendente su F. Nel secondo possiamo scegliere g come generatore monico di I. Lo chiameremo *polinomio minimo su* F dell'elemento algebrico u. È il polinomio monico di grado minimo avente u come radice, nel senso che $g(u) = 0$. Se $f(u) = 0$, allora $f \in I = (g)$ e quindi $g \mid f$. La struttura di $F[u]$ dipende da come g si fattorizza in $F[x]$. Ad esempio si ha il seguente teorema.

Teorema 2.16. *Sia u algebrico su F con polinomio minimo $g(x)$. Se g è irriducibile in $F[x]$, ossia non si può scrivere $g = fh$ con entrambi i fattori di grado positivo, allora $F[u]$ è un campo. Se invece g è riducibile, $F[u]$ non è un dominio.*

Dimostrazione. Ogni ideale di $F[x]/I$ ha forma J/I con J ideale di $F[x]$ contenente $I = (g)$ (teorema 2.6). Si ha $J = (f)$ e $g = fh$. Se g è irriducibile, uno fra f, h è invertibile. Nel primo caso $J = F[x]$, nel secondo $J = I$. Dunque $F[u] \cong F[x]/I$ ha solo gli ideali 0 e l'intero anello, perciò è un campo per il teorema 2.2. Se invece $g = fh$ con $\deg f, \deg h > 0$, entrambi i gradi sono minori di $\deg g$. Quindi $f(u), h(u) \neq 0$, ma $f(u)h(u) = g(u) = 0$. In $F[u]$ vi sono dunque divisori dello zero non nulli. □

Applichiamo il teorema del fattore al seguente risultato sulle radici.

Teorema 2.17. *Un polinomio $f(x)$ di grado $n > 0$ in $F[x]$, con F campo, ha al più n radici distinte in F.*

Dimostrazione. Siano $a_1, \ldots, a_r$ radici distinte. Dimostriamo per induzione su r che $\prod_{i=1}^{r}(x - a_i)$ divide $f(x)$. Per $r = 1$ è il teorema del fattore. Supponiamo il risultato per $r - 1$, cosicché $f(x) = \prod_{j=1}^{r-1}(x - a_j)h(x)$ in $F[x]$. Allora $0 = f(a_r) = \prod_{j=1}^{r-1}(a_r - a_j)h(a_r)$. Tutte le differenze sono non nulle, dunque $h(a_r) = 0$ e, per il caso $r = 1$, $h(x) = (x - a_r)k(x)$. Pertanto $f(x) = \prod_{i=1}^{r}(x - a_i)k(x)$. Confrontando i gradi si ottiene $r \leq n$. □

Applicando questo risultato insieme al criterio di ciclicità di un gruppo abeliano finito del teorema 1.4 (p. 46 dell'originale), otteniamo il seguente elegante teorema sui campi.

Teorema 2.18. *Ogni sottogruppo finito del gruppo moltiplicativo di un campo è ciclico.*

Dimostrazione. Sia G un sottogruppo finito di F^*, il gruppo degli elementi non nulli del campo F. Esso è abeliano. Per il criterio ricordato, G è ciclico se e solo se $|G| = \exp G$, il più piccolo intero positivo m tale che $a^m = 1$ per ogni $a \in G$. Poiché $a^{|G|} = 1$, si ha sempre $\exp G \leq |G|$. D'altra parte il polinomio $f(x) = x^{\exp G} - 1$ ha al più $\exp G$ radici in F per il teorema 2.17, quindi anche in G. Dunque $|G| \leq \exp G$, da cui l'uguaglianza e la ciclicità di G. □

Il risultato non vale per i corpi non commutativi. Ad esempio in $\mathbb{H}$ i quaternioni $\pm 1, \pm i, \pm j, \pm k$ formano un sottogruppo finito non ciclico del gruppo moltiplicativo. In particolare, per un campo finito F il gruppo F^* è ciclico. Gli elementi non nulli di $\mathbb{Z}/(p)$, con p primo, formano dunque sotto moltiplicazione un gruppo ciclico di ordine $p - 1$. Alcune conseguenze aritmetiche sono indicate negli esercizi.

Esercizi

1. Sia $f(x) = x^n + a_1 x^{n-1} + \cdots + a_n$, $a_i \in F$, con F campo e $n > 0$. Posto $u = x + (f(x))$ in $F[x]/(f(x))$, provare che ogni elemento di $F[u]$ si scrive in modo unico $b_0 + b_1 u + \cdots + b_{n-1}u^{n-1}$, $b_j \in F$.

2. Nell'esercizio 1 si prendano $F = \mathbb{Q}$, $f(x) = x^3 + 3x - 2$. Provare che $F[u]$ è un campo ed esprimere $(2u^2 + u - 3)(3u^2 - 4u + 1)$ e $(u^2 - u + 4)^{-1}$ come polinomi in u di grado al più due.

3. (a) Dimostrare che $\mathbb{Q}[\sqrt{2}]$ e $\mathbb{Q}[\sqrt{3}]$ non sono isomorfi. (b) Siano $\mathbb{F}_p = \mathbb{Z}/(p)$ con p primo, $R_1 = \mathbb{F}_p[x]/(x^2 - 2)$, $R_2 = \mathbb{F}_p[x]/(x^2 - 3)$. Stabilire se $R_1 \cong R_2$ per $p = 2, 5, 11$.

4. Dimostrare che $x^3 + x^2 + 1$ è irriducibile in $(\mathbb{Z}/(2))[x]$ e che il quoziente per $(x^3 + x^2 + 1)$ è un campo di otto elementi.

5. Costruire campi di 25 e di 125 elementi.

6. Mostrare che $x^3 - x$ ha sei radici in $\mathbb{Z}/(6)$.

7. Usando il teorema cinese del resto (esercizi 10 e 11, p. 110 dell'originale), provare che, se $f \in F[x]$ è monico e $f = gh$ con $(g, h) = 1$, allora $F[x]/(f) \cong F[x]/(g) \oplus F[x]/(h)$. Dimostrare inoltre che se $f = \prod_{i=1}^{n}(x - a_i)$ con a_i distinti, allora $F[x]/(f) \cong F \oplus \cdots \oplus F$ con n addendi.

8. Provare che nel corpo $\mathbb{H}$ vi sono infiniti elementi u tali che $u^2 = -1$.

9. Dimostrare che l'ideale $(3, x^3 - x^2 + 2x - 1)$ di $\mathbb{Z}[x]$ non è principale.

10. Sia I l'ideale dell'esercizio 9. Il quoziente $\mathbb{Z}[x]/I$ è un dominio? (Suggerimento. Provare che $\mathbb{Z}[x]/I \cong \bar{\mathbb{Z}}[\bar{x}]/\bar{I}$, dove $\bar{\mathbb{Z}} = \mathbb{Z}/(3)$, $\bar{I} = (\bar{x}^3 - \bar{x}^2 + \bar{2}\bar{x} - \bar{1})$, $\bar{x} = x + (3)$.)

11. Sia R un anello privo di elementi nilpotenti non nulli ($z^n = 0 \Rightarrow z = 0$). Se $f(x) \in R[x]$ è un divisore dello zero, provare che esiste $0 \neq a \in R$ con $af(x) = 0$. (Nota. Il risultato vale per R qualsiasi.)

12. Sia F un campo di q elementi, $F^* = \{a_1, \ldots, a_{q-1}\}$. Provare che $a_1 \cdots a_{q-1} = -1$. (Suggerimento. Usare la dimostrazione del teorema 2.18 e, per q pari, l'esercizio 5, p. 110 dell'originale.)

13. Provare il teorema di Wilson: se p è un primo di $\mathbb{Z}$, allora $(p-1)! \equiv -1 \pmod{p}$.

14. Trovare generatori per i gruppi ciclici $\mathbb{Z}_p^*$ degli elementi non nulli di $\mathbb{Z}/(p)$ per $p = 3, 5, 7, 11$.

15. Un intero a si chiama *residuo quadratico* o *non residuo quadratico* modulo il primo p secondo che $x^2 \equiv a \pmod{p}$ abbia o non abbia soluzione. Il *simbolo di Legendre* $\left(\frac{a}{p}\right)$ vale 0 se $a \equiv 0 \pmod{p}$, 1 se $a \not\equiv 0$ ed è un residuo quadratico, -1 se non è un residuo quadratico. Osservare che vale 1 se e solo se $a+(p)$ è un quadrato nel gruppo moltiplicativo di $\mathbb{Z}/(p)$. Dedurne che per $p \neq 2$ vale 1 se e solo se $a^{(p-1)/2} \equiv 1 \pmod{p}$. Per interi a, b provare $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$.

16. Siano $f(x), g(x) \neq 0$ in $F[x]$, con $\deg g = m$. Provare che $f(x)$ si scrive in modo unico $a_0(x) + a_1(x)g(x) + a_2(x)g(x)^2 + \cdots + a_r(x)g(x)^r$, con $\deg a_i < m$.

Il prossimo esercizio dà un'altra dimostrazione del teorema del resto con vari vantaggi rispetto a quella del testo. Fornisce in particolare una formula esplicita per il quoziente e vale per anelli non commutativi.

17. Sia $f(x) = a_0 + a_1 x + \cdots + a_n x^n$. Si ha $x^i - a^i = (x^{i-1} + ax^{i-2} + \cdots + a^{i-1})(x - a)$ per $i \geq 1$. Moltiplicando a sinistra per a_i e sommando su i si ottiene $\sum_{i=0}^{n} a_i x^i - \sum_{i=0}^{n} a_i a^i = \sum_{i=1}^{n} a_i(x^{i-1} + ax^{i-2} + \cdots + a^{i-1})(x - a)$. Pertanto $f(x) = q(x)(x - a) + f(a)$, dove $f(a) = \sum_{i=0}^{n} a_i a^i$ e $q(x) = \sum_{j=1}^{n} q_j x^{j-1}$ con $q_j = a_j + a_{j+1}a + \cdots + a_n a^{n-j}$.

2.12 Funzioni polinomiali

Il lettore conosce dal calcolo infinitesimale le funzioni polinomiali di una variabile reale. Estendiamo questa nozione a un campo qualsiasi F e determiniamo la relazione fra l'anello delle funzioni polinomiali e quello dei polinomi in indeterminate su F.

Per un insieme non vuoto S, indichiamo con F^S l'insieme delle applicazioni $s \mapsto f(s)$ di S in F. Come al solito, $f = g$ significa $f(s) = g(s)$ per ogni s, e somma e prodotto sono definiti da $(f+g)(s) = f(s)+g(s)$, $(fg)(s) = f(s)g(s)$. Ogni $a \in F$ definisce la *funzione costante* $a(s) = a$. Abbiamo in particolare le funzioni costanti $0, 1$. Si verifica direttamente che $(F^S, +, \cdot, 0, 1)$ è un anello commutativo. Ad esempio

$$((f+g)h)(s) = (f(s)+g(s))h(s) = f(s)h(s)+g(s)h(s) = (fh+gh)(s),$$

quindi $(f+g)h = fh+gh$. Definendo $(-f)(s) = -f(s)$ si ha $f + (-f) = 0$. È inoltre immediato che l'applicazione che manda ogni $a \in F$ nella corrispondente funzione costante è un monomorfismo. Identificheremo F con la sua immagine, cosicché F^S diventa un'estensione del campo F.

Poniamo $S = F$ e consideriamo l'anello delle applicazioni di F in sé. Oltre alle costanti ha particolare importanza l'identità $s \mapsto s$, indicata finora con 1 o 1_F. Qui seguiremo l'uso del calcolo infinitesimale e indicheremo con s tanto la funzione identica quanto la variabile, sperando che questo doppio significato non crei più della consueta confusione. Consideriamo il sottoanello $F[s]$ generato da F, campo delle funzioni costanti, e da s, funzione identica. I suoi elementi si chiamano *funzioni polinomiali in una variabile su* F. Poiché $F[s]$ è generato da F e da s, si ha l'epimorfismo di $F[x]$, con x indeterminata, su $F[s]$ che coincide con l'identità su F e manda x in s. Si ha dunque $f(x) \mapsto f(s)$, dove, per $f(x) = a_0 + a_1x + \cdots + a_nx^n$, $f(s)$ è la funzione $s \mapsto a_0 + a_1s + \cdots + a_ns^n$.

L'omomorfismo $f(x) \mapsto f(s)$ è un isomorfismo se e solo se F è infinito. Infatti $f(s) = 0$ nell'anello delle funzioni significa che $f(a) = 0$ per ogni $a \in F$. Abbiamo già visto che un polinomio non nullo di grado n ha al più n radici distinte. Per F infinito, se $f(a) = 0$ per ogni a, necessariamente $f = 0$. Il nucleo dell'epimorfismo è quindi nullo e otteniamo un isomorfismo di $F[x]$ con l'anello delle funzioni polinomiali. Se invece $F = \{a_1, \ldots, a_q\}$ è finito, il polinomio $h(x) = (x-a_1)(x-a_2)\cdots(x-a_q)$ è non nullo, mentre la funzione $h(s) = (s-a_1)\cdots(s-a_q)$ è nulla, poiché $h(a_i) = 0$ per ogni i. L'omomorfismo non è quindi un isomorfismo. Lo si vede anche contando: l'insieme di tutte le applicazioni $F \to F$ è finito, e dunque $F[s]$ è finito, mentre $F[x]$ è infinito. Tra questi anelli non esiste quindi alcun isomorfismo.

La definizione per più variabili è una generalizzazione immediata. Prendiamo $S = F^{(r)} = F \times \cdots \times F$, prodotto di r copie di F, i cui elementi sono le successioni finite $(s_1, \ldots, s_r)$. Come prima, l'anello delle funzioni $F^S = F^{F^{(r)}}$ è un'estensione di F. Scegliamo le r funzioni particolari dette «proiezioni sugli r assi», $(s_1, \ldots, s_r) \mapsto s_i$, $1 \le i \le r$. Indichiamo la i-esima proiezione con s_i, secondo l'uso, e consideriamo l'anello $F[s_1, \ldots, s_r]$ che si ottiene aggiungendo queste funzioni al campo delle costanti. I suoi elementi si chiamano *funzioni polinomiali in r variabili su F*. Si ha l'epimorfismo da $F[x_1, \ldots, x_r]$ che manda $a \mapsto a$ per $a \in F$ e $x_i \mapsto s_i$. Indichiamo l'immagine di $f(x_1, \ldots, x_r)$ con $f(s_1, \ldots, s_r)$. Per un campo finito di q elementi, come per $r = 1$, non si ha un isomorfismo. Per un campo infinito, invece, l'applicazione è un isomorfismo, come ora proveremo. L'asserzione equivale al teorema seguente.

Teorema 2.19. *Se F è un campo infinito e $0 \neq f(x_1, \ldots, x_r) \in F[x_1, \ldots, x_r]$, con x_i indeterminate, esistono $a_1, \ldots, a_r \in F$ tali che $f(a_1, \ldots, a_r) \neq 0$.*

Dimostrazione. Abbiamo già trattato il caso $r = 1$. Sia $r > 1$ e supponiamo il risultato vero per $r - 1$. Scriviamo $f = B_0 + B_1 x_r + B_2 x_r^2 + \cdots + B_n x_r^n$, con $B_i \in F[x_1, \ldots, x_{r-1}]$ e $B_n = B_n(x_1, \ldots, x_{r-1}) \neq 0$. Per induzione esistono $a_i \in F$ tali che $B_n(a_1, \ldots, a_{r-1}) \neq 0$. Allora

$$f(a_1, \ldots, a_{r-1}, x_r) = B_0(a_1, \ldots, a_{r-1}) + B_1(a_1, \ldots, a_{r-1})x_r + \cdots + B_n(a_1, \ldots, a_{r-1})x_r^n$$

è un polinomio non nullo di $F[x_r]$. Si può quindi scegliere $x_r = a_r$ in modo che $f(a_1, \ldots, a_r) \neq 0$. □

Per F finito si può determinare facilmente anche il nucleo K dell'epimorfismo di $F[x_1, \ldots, x_r]$ sull'anello delle funzioni polinomiali. Accenniamo alla dimostrazione, lasciando al lettore i dettagli. Se $|F| = q$, l'argomento precedente mostra che un polinomio $f \neq 0$ di grado minore di q in ciascuna x_i dà una funzione $f(s_1, \ldots, s_r) \neq 0$. Inoltre $x_i^q - x_i \in K$, poiché $a^q = a$ per $a \in F$ (esercizio 3, p. 105 dell'originale). Ogni polinomio si può scrivere

$$f(x_1, \ldots, x_r) = \sum_{i=1}^{r} f_i(x_1, \ldots, x_r)(x_i^q - x_i) + f_0(x_1, \ldots, x_r), \tag{44}$$

dove il grado di f_0 è minore di q in ciascuna x_i. Infatti si scriva ogni potenza $x_i^k = (x_i^q - x_i)q_k(x_i) + r_k(x_i)$ con $q_k, r_k \in F[x_i]$ e $\deg r_k < q$. Sostituendo in ogni monomio $x_1^{k_1} \cdots x_r^{k_r}$ di f si ottiene la (44). Ora $f \in K$ se e solo se $f_0 = 0$. Dunque K è l'ideale generato dagli $x_i^q - x_i$ e l'anello delle funzioni polinomiali in r variabili su un campo di q elementi è isomorfo a

$$F[x_1, \ldots, x_r]/(x_1^q - x_1, x_2^q - x_2, \ldots, x_r^q - x_r).$$

Esercizi

1. Dimostrare la seguente estensione del teorema 2.19. Sia F infinito e siano $f, g \in F[x_1, \ldots, x_r]$ con $g \neq 0$. Se $f(a_1, \ldots, a_r) = 0$ in ogni punto in cui $g(a_1, \ldots, a_r) \neq 0$, allora $f = 0$.

Nei restanti esercizi F è un campo finito con $|F| = q$.

2. Provare che ogni funzione in r variabili su F, cioè ogni elemento di $F^{F^{(r)}}$, è polinomiale. (Suggerimento. Contare gli elementi dei due insiemi.)

3. Definire il *grado* di $x_1^{i_1} \cdots x_r^{i_r}$ come $\sum i_j$ e il *grado totale* di f come il massimo dei gradi dei monomi che vi compaiono, cioè con coefficiente $a_{i_1 \cdots i_r} \neq 0$. Provare che il metodo usato per la (44), sostituendo ogni x_i^k con $(x_i^q - x_i)q_k(x_i) + r_k(x_i)$, $\deg r_k < q$, produce un f_0 con $\deg f_0 \leq \deg f$ e di grado minore di q in ciascuna x_i.

4. Se f_0, g_0 hanno grado minore di q in ciascuna indeterminata e definiscono la stessa funzione, dimostrare che $f_0 = g_0$.

5. Sia $f(0, \ldots, 0) = 0$ e $f(a_1, \ldots, a_r) \neq 0$ per ogni $(a_1, \ldots, a_r) \neq (0, \ldots, 0)$. Dimostrare che se $g = 1 - f^{q-1}$, allora $g(a_1, \ldots, a_r) = 1$ nell'origine e 0 altrove.

6. Mostrare che g dell'esercizio 5 determina la medesima funzione polinomiale di $f_0 = (1 - x_1^{q-1})(1 - x_2^{q-1}) \cdots (1 - x_r^{q-1})$. Dedurre che $\deg g \geq r(q - 1)$.

7. (Artin–Chevalley.) Sia $f(x_1, \ldots, x_r)$ un polinomio di grado $n < r$, dove r è il numero delle indeterminate. Se $f(0, \ldots, 0) = 0$, provare che esiste $(a_1, \ldots, a_r) \neq (0, \ldots, 0)$ per cui $f(a_1, \ldots, a_r) = 0$.

2.13 Polinomi simmetrici

Sia R un anello e sia $R[x_1, \dots, x_r]$ l'anello dei polinomi su R in r indeterminate. Come abbiamo visto nel teorema 2.12, ogni permutazione $\pi : i \mapsto i'$ di $\{1, \dots, r\}$ determina un automorfismo $\zeta(\pi)$ che manda $a \mapsto a$ per $a \in R$ e $x_i \mapsto x_{i'}$. Un polinomio $f(x_1, \dots, x_r)$ è *simmetrico* nelle x_i se ogni $\zeta(\pi)$ lo lascia fisso. L'insieme dei polinomi simmetrici è un sottoanello Σ di $R[x_1, \dots, x_r]$ contenente R. I coefficienti delle potenze di x nel polinomio

$$g(x) = (x - x_1)(x - x_2) \cdots (x - x_r) \tag{45}$$

sono simmetrici. Infatti possiamo estendere $\zeta(\pi)$ a un automorfismo $\zeta'(\pi)$ di $R[x_1, \dots, x_r, x]$ che fissa x. Allora $\zeta'(\pi)(g(x)) = (x - x_{1'})(x - x_{2'}) \cdots (x - x_{r'}) = g(x)$. Pertanto, scrivendo

$$g(x) = x^r - p_1 x^{r-1} + p_2 x^{r-2} - \cdots + (-1)^r p_r, \tag{46}$$

con $p_i \in R[x_1, \dots, x_r]$, si ha $\zeta(\pi)(p_i) = p_i$ per ogni π, ossia $p_i \in \Sigma$. Confrontando (45) e (46) otteniamo

$$p_1 = \sum_{i=1}^{r} x_i, \quad p_2 = \sum_{i<j} x_i x_j, \quad p_3 = \sum_{i<j<k} x_i x_j x_k, \quad \dots, \quad p_r = x_1 x_2 \cdots x_r. \tag{47}$$

I p_i si chiamano *polinomi simmetrici elementari* in $x_1, \dots, x_r$. Dimostreremo che $\Sigma = R[p_1, \dots, p_r]$ e che i p_i sono algebricamente indipendenti su R.

L'uguaglianza $\Sigma = R[p_1, \dots, p_r]$ significa che ogni polinomio simmetrico si esprime come polinomio nei p_i a coefficienti in R. Basta dimostrarlo per i polinomi omogenei. Si dice *omogeneo* un polinomio i cui termini $a x_1^{k_1} \cdots x_r^{k_r}$ hanno tutti lo stesso grado totale $k_1 + \cdots + k_r$. Ogni polinomio si scrive in modo unico come somma di polinomi omogenei di grado diverso. Poiché $\zeta(\pi)$ manda polinomi omogenei di grado k in polinomi omogenei dello stesso grado, se f è simmetrico lo sono anche le sue componenti omogenee.

Sia quindi f simmetrico e omogeneo di grado m. Ordiniamo lessicograficamente i monomi di grado m. Diremo che $x_1^{k_1} \cdots x_r^{k_r}$ è *maggiore* di $x_1^{l_1} \cdots x_r^{l_r}$ se $k_1 = l_1, \dots, k_s = l_s$ e $k_{s+1} > l_{s+1}$, con $s \geq 0$. Ad esempio $x_1^2 x_2 x_3 > x_1 x_2^3 > x_1 x_2^2 x_3$. Sia $x_1^{k_1} \cdots x_r^{k_r}$ il maggiore monomio di f con coefficiente non nullo. Poiché f è simmetrico, contiene tutti i monomi che se ne ottengono permutando le indeterminate. Dunque $k_1 \geq k_2 \geq \cdots \geq k_r$.

Consideriamo ora il maggiore monomio del polinomio omogeneo e simmetrico $p_1^{d_1} \cdots p_r^{d_r}$, $d_i \geq 0$. Se M_1, M_2 sono monomi di grado m e N è un monomio di grado qualsiasi, allora $M_1 > M_2$ implica $NM_1 > NM_2$. In particolare, se anche $N_1 > N_2$, si ha $M_1 N_1 > M_2 N_2$. Il maggior monomio di p_i è $x_1 x_2 \cdots x_i$. Ne segue che il maggior monomio di $p_1^{d_1} \cdots p_r^{d_r}$ è

$$x_1^{d_1 + d_2 + \cdots + d_r} x_2^{d_2 + \cdots + d_r} \cdots x_r^{d_r}.$$

Quindi $p_1^{k_1 - k_2} p_2^{k_2 - k_3} \cdots p_r^{k_r}$ ha lo stesso maggior monomio di f. Se a è il suo coefficiente in f, allora $f_1 = f - a p_1^{k_1 - k_2} p_2^{k_2 - k_3} \cdots p_r^{k_r}$ ha maggior monomio minore di quello di f, se non è già nullo. Ripetiamo il procedimento con f_1. Essendovi solo un numero finito di monomi di grado m, dopo un numero finito di passi si ottiene un'espressione di f in funzione di $p_1, \dots, p_r$.

Proviamo che i p_i sono algebricamente indipendenti. Supponiamo $\sum_{(d)} a_{d_1 \cdots d_r} p_1^{d_1} \cdots p_r^{d_r} = 0$, dove la somma è estesa a un insieme finito di indici distinti $(d) = (d_1, \dots, d_r)$, con $d_i \geq 0$. Se la relazione non è banale, almeno un coefficiente è non nullo. A ogni (d) associamo $(k) = (k_1, \dots, k_r)$ ponendo $k_i = d_i + d_{i+1} + \cdots + d_r$. Il grado di $p_1^{d_1} \cdots p_r^{d_r}$ nelle x_i è allora $m = \sum_{i=1}^r k_i = \sum_{i=1}^r i d_i$ e il suo maggior monomio è $x_1^{k_1} \cdots x_r^{k_r}$. Se per un secondo (d') si ha $k_i' = d_i' + \cdots + d_r' = k_i$ per ogni i, allora $d_i' = d_i$ per ogni i. Monomi distinti nei p_i hanno dunque

maggiori monomi distinti nelle x_i. Fra gli indici (d) con coefficiente non nullo scegliamone uno per cui m sia massimo e il maggior monomio $x_1^{k_1}\cdots x_r^{k_r}$ sia massimo. Espandendo nelle x_i, questo monomio compare una sola volta con coefficiente $a_{d_1\cdots d_r}\neq 0$. Ciò contraddice l'indipendenza algebrica delle x_i. Abbiamo provato le prime due affermazioni del teorema che segue.

Teorema 2.20. *Ogni polinomio simmetrico si esprime come polinomio nei polinomi simmetrici elementari p_i. Questi sono algebricamente indipendenti su R. Ogni x_i è algebrica su $R[p_1,\dots,p_r]$.*

L'ultima affermazione segue da (45) e (46), poiché $g(x_i)=x_i^r-p_1x_i^{r-1}+p_2x_i^{r-2}-\cdots+(-1)^rp_r=0$.

Esercizi

1. Esprimere $\sum_{i,j,k \text{ distinti}} x_i^2x_j^2x_k$, $r\geq 5$, in funzione dei p_i.
2. Sia $\Delta=\prod_{i<j}(x_i-x_j)$. Provare che Δ^2 è simmetrico e, per $r=3$, esprimerlo in funzione dei polinomi simmetrici elementari.
3. (Identità di Newton.) Posto $s_k=\sum_{i=1}^n x_i^k$, dimostrare le relazioni fra i polinomi simmetrici s_k e i polinomi simmetrici elementari p_i:

$$\begin{aligned} s_k-p_1s_{k-1}+p_2s_{k-2}-\cdots+(-1)^{k-1}p_{k-1}s_1+(-1)^k kp_k&=0, \quad 1\leq k\leq n,\\ s_{n+j}-p_1s_{n+j-1}+\cdots+(-1)^kp_ks_{n+j-k}+\cdots+(-1)^np_ns_j&=0, \quad j>0. \end{aligned}$$

Queste formule ricorsive esprimono le somme di potenze s_k come polinomi nei p_i. Mostrano anche che $k!\,p_k$ è un polinomio in $s_1,\dots,s_k$ a coefficienti interi. (Traccia. Si scriva $f(x)=x^n-p_1x^{n-1}+\cdots+(-1)^np_n=\prod_{i=1}^n(x-x_i)=(x-x_i)q_i(x)$. Per l'esercizio 17, p. 134 dell'originale,

$$q_i(x)=x^{n-1}-(p_1-x_i)x^{n-2}+\cdots+(-1)^k(p_k-p_{k-1}x_i+p_{k-2}x_i^2-\cdots+(-1)^kx_i^k)x^{n-k-1}+\cdots.$$

La derivazione formale (pp. 230–231 dell'originale) dà

$$\begin{aligned} nx^{n-1}-(n-1)p_1x^{n-2}+\cdots=\sum_{i=1}^n q_i(x)&=nx^{n-1}-(np_1-s_1)x^{n-2}+\cdots\\ &\quad+(-1)^k(np_k-p_{k-1}s_1+\cdots+(-1)^ks_k)x^{n-k-1}+\cdots. \end{aligned}$$

Confrontando i coefficienti di x^{n-k-1} si ottiene il primo gruppo di identità per gli indici fino a $n-1$. Le restanti si ottengono sommando su i le relazioni $x_i^{n+j}-p_1x_i^{n+j-1}+\cdots+(-1)^np_nx_i^j=0$ per $j\geq 0$.)

2.14 Monoidi e anelli fattoriali

Nella parte restante del capitolo studieremo la teoria elementare della divisibilità nei domini commutativi. In molti domini importanti ogni $a\neq 0$ non invertibile si può scrivere $a=p_1p_2\cdots p_s$ con i p_i irriducibili, e la fattorizzazione è unica a meno di fattori invertibili e dell'ordine dei fattori. Si possono allora determinare tutti i divisori di a, a meno di fattori invertibili, e dare una semplice condizione affinché $a\mid b$, ossia l'equazione $ax=b$ sia risolubile. Poiché la teoria della fattorizzazione è puramente moltiplicativa e riguarda soprattutto il monoide moltiplicativo del dominio, conviene cominciare dalla divisibilità nei monoidi.

Sia M un monoide commutativo con la legge di cancellazione ($ab = ac$ implica $b = c$), e sia U il sottogruppo degli elementi invertibili. Per $a, b \in M$ diciamo che b è un *fattore* o *divisore* di a, e scriviamo $b \mid a$, se esiste $c \in M$ tale che $a = bc$. In tal caso a è un *multiplo* di b. La divisibilità è transitiva e riflessiva ($b \mid a$ e $c \mid b$ implicano $c \mid a$, e $a \mid a$), ma non è simmetrica. Un elemento u è invertibile se e solo se $u \mid 1$. Gli invertibili sono fattori banali poiché dividono ogni elemento, $a = u(u^{-1}a)$. Se $a \mid b$ e $b \mid a$, diciamo che a, b sono *associati* e scriviamo $a \sim b$. Allora $b = au$, $a = bv$, dunque $b = bvu$. Per cancellazione $vu = 1$, e v, u sono invertibili. Il viceversa è immediato, quindi due elementi sono associati se e solo se differiscono per un fattore invertibile. Poiché gli invertibili costituiscono un sottogruppo, essere associati è una relazione di equivalenza.

Se $b \mid a$ ma $a \nmid b$, b si dice *fattore proprio* di a. Se u è invertibile e $u = vw$, anche v, w sono invertibili. Quindi un invertibile non ha fattori propri. Un elemento $a \in M$ si dice *irriducibile*[6] se non è invertibile e i suoi soli fattori propri sono invertibili. Se a non è né invertibile né irriducibile, si scrive $a = bc$ con b, c fattori propri di a. Ogni associato di un irriducibile è irriducibile.

Se $a = p_1 \cdots p_s$ con i p_i irriducibili, si ha anche $a = p'_1 \cdots p'_s$ con $p'_i = u_i p_i$ per qualsiasi scelta di invertibili u_i tali che $u_1 \cdots u_s = 1$. Se M ha invertibili diversi da 1 e $s > 1$, ciò permette sempre di modificare la fattorizzazione. Per commutatività si può anche cambiare l'ordine dei fattori. Diremo che una fattorizzazione in irriducibili è *essenzialmente unica* se queste sono le sole modifiche possibili. Più precisamente, $a = p_1 \cdots p_s$ è essenzialmente unica se per ogni altra fattorizzazione $a = p'_1 \cdots p'_t$, con p'_j irriducibili, si ha $t = s$ e, per un'opportuna permutazione $i \mapsto i'$ di $\{1, \ldots, s\}$, $p'_{i'} \sim p_i$.

Definizione 2.4. Un monoide commutativo M con la legge di cancellazione si dice *fattoriale* (talvolta *gaussiano* o *a fattorizzazione unica*) se ogni suo elemento non invertibile ha una fattorizzazione essenzialmente unica in irriducibili. Un dominio D si dice *fattoriale* se lo è il monoide D^* dei suoi elementi non nulli.

Il nostro obiettivo principale nel resto del capitolo sarà provare che varie classi importanti di domini sono fattoriali. Il seguente esempio mostra che non tutti lo sono.

Esempio. Sia $D = \mathbb{Z}[\sqrt{-5}]$, l'insieme dei complessi $a + b\sqrt{-5}$ con $a, b \in \mathbb{Z}$. Si verifica facilmente che è un sottoanello di $\mathbb{C}$, dunque un dominio. Per studiarne l'aritmetica introduciamo la *norma* $N(r) = r\bar{r} = a^2 + 5b^2$ per $r = a + b\sqrt{-5}$. Poiché il valore assoluto complesso è moltiplicativo, lo è anche la norma, $N(rs) = N(r)N(s)$. Se $r \neq 0$, $N(r)$ è un intero positivo. Usiamola anzitutto per determinare gli invertibili di D. Se $rs = 1$, allora $N(r)N(s) = 1$, dunque $N(r) = a^2 + 5b^2 = 1$. Per a, b interi ciò è possibile solo se $a = \pm 1$, $b = 0$. Quindi $U = \{1, -1\}$ e i soli associati di r sono $r, -r$. Mostriamo che 9 ha due fattorizzazioni in irriducibili che non differiscono solo per fattori invertibili:

$$9 = 3 \cdot 3 = (2 + \sqrt{-5})(2 - \sqrt{-5}).$$

Tutti i fattori indicati sono irriducibili. Infatti se $3 = rs$ con $r, s \in D$, allora $9 = N(3) = N(r)N(s)$. Se entrambi i fattori non sono invertibili, deve essere $N(r) = N(s) = 3$, ma $a^2 + 5b^2 = 3$ non ha soluzioni intere. Dunque 3 è irriducibile, e analogamente lo sono $2 + \sqrt{-5}$ e $2 - \sqrt{-5}$. Inoltre 3 non è associato a nessuno dei due. Perciò 9 ammette fattorizzazioni in irriducibili ma non una fattorizzazione essenzialmente unica. Il dominio $\mathbb{Z}[\sqrt{-5}]$ non è fattoriale.

[6] Usiamo questo termine anziché «primo», adoperato fin qui per l'aritmetica di $\mathbb{Z}$. Nel caso generale gli elementi primi avranno una diversa definizione, data poco oltre.

In un monoide fattoriale M basta conoscere una fattorizzazione di un non invertibile a per individuarne tutti i fattori a meno di invertibili. Se $a = p_1 \cdots p_s = bc$, dove $b = p'_1 \cdots p'_l$, $c = p''_1 \cdots p''_n$ con fattori irriducibili, allora $p_1 \cdots p_s = p'_1 \cdots p'_l p''_1 \cdots p''_n$. Per unicità $p'_j \sim p_{i_j}$ con gli i_j distinti, dunque $b \sim p_{i_1} \cdots p_{i_l}$. Ogni fattore di a è così associato a un prodotto di fattori scelti nella fattorizzazione data. Chiamando *lunghezza* di a il numero s dei suoi fattori irriducibili, ogni suo fattore proprio ha lunghezza minore. Ne segue la condizione seguente.

Condizione delle catene di divisori. In M non esiste una successione infinita $a_1, a_2, \ldots$ in cui ogni a_{i+1} sia un fattore proprio di a_i.

Equivalentemente, in ogni successione tale che $a_{i+1} \mid a_i$ esiste N con $a_N \sim a_{N+1} \sim a_{N+2} \sim \cdots$.

Ricaviamo una seconda condizione necessaria. Un elemento p si dice *primo* se non è invertibile e $p \mid ab$ implica $p \mid a$ oppure $p \mid b$. In altri termini, $p \nmid a$ e $p \nmid b$ implicano $p \nmid ab$. Sia p irriducibile in un monoide fattoriale e sia $p \mid ab$. Se a è invertibile, $ab \sim b$, quindi $p \mid b$. Analogamente, se b è invertibile, $p \mid a$. Altrimenti scriviamo $a = p_1 \cdots p_s$, $b = p'_1 \cdots p'_t$ in irriducibili. Poiché $p \mid p_1 \cdots p_s p'_1 \cdots p'_t$, si ha $p \sim p_i$ per qualche i oppure $p \sim p'_j$ per qualche j. Dunque $p \mid a$ o $p \mid b$. Abbiamo mostrato che ogni monoide fattoriale soddisfa la seguente condizione.

Condizione di primalità. In M ogni irriducibile è primo.

Dimostriamo che le due condizioni sono sufficienti per la fattorialità. La condizione delle catene assicura l'esistenza di una fattorizzazione in irriducibili per ogni non invertibile a. Mostriamo prima che a ha un fattore irriducibile. Se a è irriducibile non vi è nulla da provare. Altrimenti scriviamo $a = a_1 b_1$ con a_1 fattore proprio non invertibile. Se a_1 non è irriducibile, scriviamo $a_1 = a_2 b_2$ con a_2 fattore proprio non invertibile di a_1. In tal modo otteniamo una successione $a, a_1, a_2, \ldots$ di fattori propri successivi. Per la condizione delle catene il procedimento deve terminare dopo un numero finito di passi con un fattore irriducibile a_n di a. Poniamo $p_1 = a_n$ e scriviamo $a = p_1 a'$. Se a' è invertibile, a è irriducibile e abbiamo finito. Altrimenti $a' = p_2 a''$ con p_2 irriducibile. Proseguendo si ottiene $a, a', a'', \ldots$, dove ogni elemento è fattore proprio del precedente e $a^{(i-1)} = p_i a^{(i)}$. Il procedimento si arresta con un irriducibile $a^{(s-1)} = p_s$, onde $a = p_1 a' = p_1 p_2 a'' = \cdots = p_1 \cdots p_s$.

Mostriamo ora che la primalità assicura l'unicità essenziale. Siano

$$a = p_1 p_2 \cdots p_s = p'_1 p'_2 \cdots p'_t \tag{48}$$

due fattorizzazioni in irriducibili. Se $s = 1$, $a = p_1$ è irriducibile, dunque $t = 1$ e $p'_1 = p_1$. Supponiamo per induzione che un prodotto di $s-1$ irriducibili abbia una sola fattorizzazione a meno di associati e dell'ordine. Per la condizione di primalità p_1 è primo. Per induzione, se un primo divide $a_1 \cdots a_r$, divide un qualche a_i. Pertanto $p_1 \mid p'_j$ per qualche j e, riordinando i fattori, possiamo supporre $p_1 \mid p'_1$. Per irriducibilità $p'_1 \sim p_1$, cioè $p'_1 = p_1 u_1$ con u_1 invertibile. Sostituendo nella (48) e cancellando p_1 si ottiene

$$b = p_2 \cdots p_s = u_1 p'_2 \cdots p'_t = p''_2 \cdots p''_t,$$

con $p''_2 = u_1 p'_2$ e $p''_i = p'_i$ per $i > 2$, tutti irriducibili. Per induzione $s - 1 = t - 1$ e dopo un opportuno riordinamento dei p''_j si ha $p_j \sim p''_j$ per $j = 2, \ldots, s$. Quindi $s = t$ e, riordinando, $p_i \sim p'_i$ per ogni i. Abbiamo dimostrato il criterio seguente.

Teorema 2.21. *Un monoide commutativo con la legge di cancellazione è fattoriale se e solo se soddisfa la condizione delle catene di divisori e la condizione di primalità.*

Possiamo sostituire la seconda condizione richiedendo che ogni coppia di elementi di M abbia un massimo comune divisore. Un elemento d è un *massimo comune divisore* di a, b se $d \mid a, d \mid b$ e ogni c che divide entrambi divide d. Siano d, d' due massimi comuni divisori. Allora $d \mid d'$ e $d' \mid d$, quindi $d \sim d'$. Il massimo comune divisore, se esiste, è determinato a meno di un fattore invertibile. Indicheremo con (a, b) una sua qualunque determinazione. La nozione duale è quella di *minimo comune multiplo*: m lo è per a, b se $a \mid m$, $b \mid m$ e $m \mid n$ per ogni n tale che $a \mid n$, $b \mid n$. Lo indicheremo con $[a, b]$.

In un monoide fattoriale ogni coppia a, b ha un massimo comune divisore e un minimo comune multiplo. Se a è invertibile, a e b sono rispettivamente un massimo comune divisore e un minimo comune multiplo della coppia. Supponiamo dunque a non invertibile. Raggruppando i suoi fattori irriducibili associati, scegliendo per ognuno un rappresentante e raccogliendo i fattori invertibili, scriviamo

$$a = up_1^{e_1} p_2^{e_2} \cdots p_r^{e_r}, \tag{49}$$

dove u è invertibile, i p_i sono irriducibili a due a due non associati e gli e_i sono interi positivi. I fattori di a hanno la forma $u' p_1^{e_1'} \cdots p_r^{e_r'}$ con u' invertibile e $0 \leq e_i' \leq e_i$. Per due elementi non invertibili a, b si può scegliere la stessa famiglia di irriducibili non associati, scrivendo

$$a = up_1^{e_1} \cdots p_t^{e_t}, \qquad b = vp_1^{f_1} \cdots p_t^{f_t}, \tag{50}$$

con u, v invertibili ed esponenti non negativi. Consideriamo

$$d = p_1^{g_1} \cdots p_t^{g_t}, \qquad g_i = \min(e_i, f_i). \tag{51}$$

È chiaro che $d \mid a$ e $d \mid b$. Se c divide entrambi, si ha $c = wp_1^{k_1} \cdots p_t^{k_t}$ con w invertibile e $0 \leq k_i \leq e_i, f_i$, dunque $k_i \leq g_i$ e $c \mid d$. Pertanto d è un massimo comune divisore di a, b. Analogamente, per $h_i = \max(e_i, f_i)$,

$$m = p_1^{h_1} \cdots p_t^{h_t} \tag{52}$$

è un minimo comune multiplo di a, b.

Se un massimo comune divisore di a, b è invertibile, possiamo prendere $(a, b) = 1$ e diciamo che a, b sono *primi fra loro*. Ciò avviene se e solo se uno dei due è invertibile o nessun irriducibile divide entrambi.

Sia ora M un monoide commutativo con cancellazione che soddisfa la seguente ipotesi.

Condizione del massimo comune divisore. Due elementi qualsiasi di M hanno un massimo comune divisore.

Mostreremo che ciò implica la primalità degli irriducibili. Suddividiamo l'argomento in semplici lemmi.

Lemma 1. *Ogni famiglia finita $a_1, \ldots, a_r$ ammette un massimo comune divisore, ossia un $d \in M$ che divida tutti gli a_i e sia divisibile per ogni loro divisore comune e.*

Dimostrazione. Basta prendere successivamente $d_1 = (a_1, a_2)$, $d_2 = (d_1, a_3)$ e così via fino ad aver incluso a_r. La definizione mostra che l'ultimo elemento ottenuto è un massimo comune divisore degli a_i. □

Lo indichiamo con $(a_1, \ldots, a_r)$.

Lemma 2. $((a,b),c) \sim (a,(b,c))$.

Dimostrazione. Entrambi i membri sono massimi comuni divisori di a, b, c. □

Lemma 3. $c(a,b) \sim (ca, cb)$.

Dimostrazione. Poniamo $d = (a,b)$, $e = (ca, cb)$. Poiché $cd \mid ca$ e $cd \mid cb$, si ha $cd \mid e$, dunque $e = cdu$. Ora $ca = ex = cdux$, quindi $a = dux$ e $du \mid a$. Analogamente $du \mid b$, onde $du \mid d$. Pertanto u è invertibile e $(ca, cb) \sim cd \sim c(a,b)$. □

Lemma 4. *Se* $(a,b) \sim 1$ *e* $(a,c) \sim 1$, *allora* $(a, bc) \sim 1$.

Dimostrazione. Il lemma 3 dà $(ac, bc) \sim c$ e chiaramente $(a, ac) \sim a$. Quindi

$$1 \sim (a,c) \sim (a,(ac,bc)) \sim ((a,ac),bc) \sim (a,bc). \qquad \square$$

Lemma 5. *La condizione del massimo comune divisore implica quella di primalità.*

Dimostrazione. Sia p irriducibile e supponiamo $p \nmid a$, $p \nmid b$. Per irriducibilità $(p,a) \sim 1$, $(p,b) \sim 1$. Il lemma 4 dà $(p, ab) \sim 1$, quindi $p \nmid ab$. Pertanto $p \mid ab$ implica $p \mid a$ oppure $p \mid b$. □

Si ottiene così un secondo criterio.

Teorema 2.22. *Un monoide commutativo con cancellazione è fattoriale se e solo se soddisfa la condizione delle catene di divisori e la condizione del massimo comune divisore.*

Dimostrazione. Per il lemma 5 le due ipotesi comportano la condizione delle catene e quella di primalità, quindi la fattorialità per il teorema 2.21. Viceversa, se M è fattoriale soddisfa la condizione delle catene e, come abbiamo visto, ogni coppia dei suoi elementi ha un massimo comune divisore. □

Esercizi

1. Provare che in un monoide fattoriale $ab \sim [a,b](a,b)$.
2. In un monoide commutativo M con cancellazione provare che essere associati è una congruenza. Sia $\bar{M}$ il monoide quoziente. Provare che in $\bar{M}$ vale la cancellazione, che $\bar{1}$ è il solo invertibile e che M è fattoriale se e solo se lo è $\bar{M}$.
3. Provare che $\mathbb{Z}[\sqrt{-5}]$ soddisfa la condizione delle catene di divisori (per gli elementi non nulli).
4. Provare che $\mathbb{Z}[x]$ soddisfa la condizione delle catene di divisori.
5. Sia D l'insieme delle espressioni $a_1 x^{\alpha_1} + \cdots + a_n x^{\alpha_n}$, con a_i in un campo F ed α_i razionali non negativi. Definire uguaglianza e somma nel modo consueto e il prodotto per distributività a partire da $(a_i x^{\alpha_i})(a_j x^{\alpha_j}) = a_i a_j x^{\alpha_i + \alpha_j}$. (Si può procedere rigorosamente come nell'esercizio 8, p. 127 dell'originale.) Provare che D è un dominio e che la condizione delle catene di divisori non vi è soddisfatta.
6. Dimostrare che ogni primo è irriducibile.
7. Sia $\mathbb{Z}[\sqrt{10}]$ l'insieme dei reali $a + b\sqrt{10}$ con $a, b \in \mathbb{Z}$. Provare che non è fattoriale.
8. Siano p un primo della forma $4n+1$ e q un primo tale che $\left(\frac{q}{p}\right) = -1$ (per il simbolo di Legendre si veda l'esercizio 15 del § 2.11). Provare che $\mathbb{Z}[\sqrt{pq}]$ non è fattoriale.

2.15 Domini a ideali principali e domini euclidei

Applichiamo ai domini i risultati sulla fattorizzazione nei monoidi. Essi valgono per ogni dominio commutativo D, poiché i suoi elementi non nulli formano un sottomonoide D^* del monoide moltiplicativo di D e vi vale la legge di cancellazione. Concetti e risultati si trasferiscono dunque a D^*. È importante osservare, come già per $\mathbb{Z}$, che $b \mid a$ equivale all'inclusione $(b) \supseteq (a)$ degli ideali principali. Infatti quest'ultima equivale ad $a \in (b)$, ossia ad $a = bc$. Poiché in D^* due elementi sono associati se e solo se si dividono a vicenda, si ha $a \sim b$ se e solo se $(a) \supseteq (b)$ e $(b) \supseteq (a)$, ossia $(a) = (b)$. Pertanto a è fattore proprio di b se e solo se $(a) \supsetneq (b)$. La condizione delle catene di divisori per $M = D^*$ equivale così alla seguente.

Condizione delle catene ascendenti di ideali principali. In D non esistono catene infinite strettamente ascendenti di ideali principali $(a_1) \subsetneq (a_2) \subsetneq (a_3) \subsetneq \cdots$.

Abbiamo chiamato *dominio a ideali principali* un dominio in cui ogni ideale è principale. Sappiamo che $\mathbb{Z}$ e $F[x]$ per ogni campo F ne sono esempi e ne daremo altri fra poco. Mostriamo che ogni dominio a ideali principali D è fattoriale. Dimostriamo anzitutto la condizione delle catene per gli ideali principali, e dunque per tutti gli ideali. In ogni anello l'unione di una catena ascendente di ideali è un ideale (§ 2.5). Per una catena $(a_1) \subseteq (a_2) \subseteq (a_3) \subseteq \cdots$, dunque, $I = \bigcup(a_i)$ è un ideale di D. Pertanto $I = (d)$ per qualche $d \in I$. Si ha $d \in (a_n)$ per qualche n e quindi $I = (d) \subseteq (a_n)$. Per $m \geq n$, $(a_m) \supseteq (a_n) \supseteq I \supseteq (a_m)$, onde $(a_n) = (a_{n+1}) = \cdots$. Una catena strettamente ascendente infinita non può esistere.

Per concludere basta dimostrare che D^* soddisfa la condizione di primalità oppure quella del massimo comune divisore. Le proveremo entrambe, ottenendo due diverse dimostrazioni della fattorialità. Siano $a, b \in D$ e si consideri l'ideale (a, b).[7] Esattamente come in $\mathbb{Z}$ (§ 2.6), se $(a, b) = (d)$ allora d è un massimo comune divisore di a, b. Poiché ogni ideale è principale, ogni coppia ha dunque un massimo comune divisore.

Mostriamo direttamente che in un dominio a ideali principali ogni irriducibile è primo. Si avrà così una dimostrazione indipendente dalle considerazioni sui massimi comuni divisori che hanno condotto al teorema 2.22. Sia p irriducibile in D^* e sia $p \mid ab$ ma $p \nmid a$, con $a, b \in D^*$. L'irriducibilità significa che non esistono ideali I con $D \supsetneq I \supsetneq (p)$. Poiché $a \notin (p)$, si ha $(p, a) \supsetneq (p)$, quindi $(p, a) = (1)$. Esistono perciò $u, v \in D$ con $up + va = 1$. Allora $upb + vab = b$ e, poiché $p \mid ab$, segue $p \mid b$. Dunque p è primo. Abbiamo così provato due volte il teorema seguente.

Teorema 2.23. *Ogni dominio a ideali principali è fattoriale.*

In particolare $F[x]$ è fattoriale per ogni campo F. Il risultato dà anche un'altra dimostrazione della fattorialità di $\mathbb{Z}$ (p. 22 dell'originale). La nozione di dominio a ideali principali è un bel concetto astratto. Occorre però un criterio pratico per provare che un anello è a ideali principali. Ce lo offre la nozione di dominio euclideo.

Definizione 2.5. Un dominio D si dice *euclideo* se esiste un'applicazione $\delta : D \to \mathbb{N}$ negli interi non negativi tale che, per $a, b \in D$, $b \neq 0$, esistano $q, r \in D$ con $a = bq + r$ e $\delta(r) < \delta(b)$.

L'anello $\mathbb{Z}$ è euclideo per $\delta(a) = |a|$. L'algoritmo di divisione dei polinomi mostra che $F[x]$ è euclideo per $\delta(f) = 2^{\deg f}$, intendendo $2^{-\infty} = 0$. Un altro esempio importante è il seguente.

[7] In queste considerazioni si può ammettere anche $a = 0$ o $b = 0$.

Anello degli interi gaussiani $\mathbb{Z}[\sqrt{-1}]$**.** È il sottoinsieme di $\mathbb{C}$ dei numeri $m+ni$ con $m, n \in \mathbb{Z}$ e $i = \sqrt{-1}$. I suoi elementi si identificano con i punti del «reticolo» di coordinate intere nel piano complesso. Si verifica subito che $\mathbb{Z}[i]$ è un sottoanello di $\mathbb{C}$, e dunque un dominio d'integrità. Per $a = m + ni$ poniamo $\delta(a) = a\bar{a} = |a|^2 = m^2 + n^2$. Allora $\delta(a) \in \mathbb{N}$ e $\delta(ab) = \delta(a)\delta(b)$. Per verificare la condizione euclidea, sia $b \neq 0$ e sia $ab^{-1} = \mu + \nu i$ con μ, ν razionali. Scegliamo interi u, v con $|u - \mu| \leq \frac{1}{2}$, $|v - \nu| \leq \frac{1}{2}$. Posti $\varepsilon = \mu - u$, $\eta = \nu - v$, si hanno $|\varepsilon|, |\eta| \leq \frac{1}{2}$. Allora

$$a = b[(u + \varepsilon) + (v + \eta)i] = bq + r,$$

dove $q = u + vi \in \mathbb{Z}[i]$, $r = b(\varepsilon + \eta i)$. Poiché $r = a - bq$, anche $r \in \mathbb{Z}[i]$. Inoltre

$$\delta(r) = |r|^2 = |b|^2(\varepsilon^2 + \eta^2) \leq |b|^2(\tfrac{1}{4} + \tfrac{1}{4}) = \tfrac{1}{2}\delta(b) < \delta(b).$$

Dunque $\mathbb{Z}[\sqrt{-1}]$ è euclideo. Il risultato principale sui domini euclidei è il seguente.

Teorema 2.24. *Ogni dominio euclideo è a ideali principali.*

Dimostrazione. La dimostrazione è identica a quella per $D = F[x]$. Sia I un ideale nel dominio euclideo D. Se $I = 0$, si ha $I = (0)$. Altrimenti scegliamo $b \neq 0$ in I con $\delta(b)$ minimo fra i valori assunti sugli elementi non nulli di I. Per $a \in I$ esistono $q, r \in D$ tali che $a = bq + r$ e $\delta(r) < \delta(b)$. Poiché $r = a - bq \in I$, la scelta di b impone $r = 0$. Allora $a = bq$ e $I = (b)$. □

Poiché ogni dominio a ideali principali è fattoriale, otteniamo il corollario.

Corollario. *I domini euclidei sono fattoriali.*

Esercizi

1. Sia F un campo. È un dominio a ideali principali?
2. Provare che l'insieme $\mathbb{Z}[\sqrt{2}]$ dei reali $m + n\sqrt{2}$ con $m, n \in \mathbb{Z}$ è un dominio euclideo per $\delta(m + n\sqrt{2}) = |m^2 - 2n^2|$.
3. Sia D l'insieme dei complessi $m + n\sqrt{-3}$ con m, n entrambi interi oppure entrambi metà di interi dispari (esercizio 4, p. 89 dell'originale). Provare che è euclideo per $\delta(m + n\sqrt{-3}) = m^2 + 3n^2$.
4. Siano D un dominio a ideali principali ed E un dominio che lo contiene come sottoanello. Se d è un massimo comune divisore di a, b in D, provare che lo è anche in E.
5. Per $0 \neq a$ in un dominio a ideali principali D, provare che $D/(a)$ è un campo se a è primo e non è un dominio se a non è primo.
6. Sia D un dominio euclideo la cui funzione δ soddisfa (i) $\delta(ab) = \delta(a)\delta(b)$ e (ii) $\delta(a + b) \leq \max(\delta(a), \delta(b))$. Dimostrare che D è un campo oppure $D = F[x]$, con F campo e x indeterminata.
7. Sia p un primo della forma $4n+1$, $n \in \mathbb{Z}$. Applicare il criterio dell'esercizio 15 del § 2.11 per mostrare che $\left(\frac{-1}{p}\right) = 1$. Dedurne che p non è primo in $\mathbb{Z}[i]$, l'anello degli interi gaussiani.
8. Usando l'esercizio 7, provare che ogni primo p della forma $4n + 1$ è somma di due quadrati $a^2 + b^2$, $a, b \in \mathbb{Z}$.
9. Determinare i primi, cioè gli irriducibili, di $\mathbb{Z}[i]$.

10. Mostrare che un intero positivo m è somma di due quadrati di interi se e solo se ogni primo della forma $4n+3$ compare con molteplicità pari nella fattorizzazione di m.

11. (Algoritmo di Euclide per il massimo comune divisore.) Siano a_1, a_2 elementi non nulli di un dominio euclideo. Definire per ricorrenza a_i, q_i mediante $a_1 = q_1 a_2 + a_3$ e $a_i = q_i a_{i+1} + a_{i+2}$, con $\delta(a_{i+2}) < \delta(a_{i+1})$. Provare che esiste n tale che $a_n \neq 0$ ma $a_{n+1} = 0$ e che $d = a_n = (a_1, a_2)$. Usare inoltre le uguaglianze per esprimere $d = xa_1 + ya_2$.

12. Applicare quanto precede ai polinomi $x^3 + x^2 + x - 3$ e $x^4 - x^3 + 3x^2 + x - 4$ in $\mathbb{Q}[x]$.

I tre esercizi seguenti intendono spiegare uno dei misteri del calcolo integrale, la decomposizione delle funzioni razionali in fratti semplici.

13. Sia F un campo e $0 \neq f(x) = f_1(x)f_2(x) \in F[x]$, con $\deg f_i > 0$ e $(f_1, f_2) = 1$. Se $\deg g < \deg f$, provare che esistono $u_i \in F[x]$ con $g = u_2 f_1 + u_1 f_2$ e $\deg u_i < \deg f_i$. (Suggerimento. È chiara l'esistenza di v_1, v_2 con $v_2 f_1 + v_1 f_2 = g$. Dividere v_i per f_i ottenendo un resto u_i di grado inferiore a $\deg f_i$. Ragionare sui gradi.) Nel campo delle frazioni $F(x)$ di $F[x]$ si ha così $g/f = u_1/f_1 + u_2/f_2$. Provare per induzione che per $f = p_1^{e_1} \cdots p_r^{e_r}$, con i p_i primi distinti, $g/f = \sum_{i=1}^{r} g_i/p_i^{e_i}$ con $\deg g_i < \deg(p_i^{e_i})$.

14. Se $0 \neq g, p \in F[x]$, provare che vi sono $a_i \in F[x]$ con $\deg a_i < \deg p$ tali che $g = a_0 + a_1 p + \cdots + a_{e-1}p^{e-1}$ e

$$\frac{g(x)}{p(x)^e} = \frac{a_0(x)}{p(x)^e} + \frac{a_1(x)}{p(x)^{e-1}} + \cdots + \frac{a_{e-1}(x)}{p(x)}.$$

15. Assumendo che gli irriducibili di $\mathbb{R}[x]$ siano lineari o quadratici (lo proveremo nel capitolo 5), dimostrare che per $f, g \in \mathbb{R}[x]$, $\deg g < \deg f$, si può decomporre g/f in fratti semplici delle forme $a/(x-r)^e$ o $(bx+c)/(x^2+sx+t)^e$, con x^2+sx+t irriducibile. Precisamente, se $f = \prod_{i=1}^{m}(x-r_i)^{e_i} \prod_{j=1}^{n}(x^2+s_jx+t_j)^{f_j}$ con i quadratici irriducibili, mostrare che g/f si può scrivere

$$\sum_{i=1}^{m}\sum_{k=1}^{e_i} \frac{a_{ik}}{(x-r_i)^k} + \sum_{j=1}^{n}\sum_{\ell=1}^{f_j} \frac{b_{j\ell}x + c_{j\ell}}{(x^2+s_jx+t_j)^\ell}.$$

16. Studiare l'unicità delle decomposizioni degli esercizi 13–15.

17. Definire la funzione di Möbius $\mu(n)$ sugli interi positivi mediante (a) $\mu(1) = 1$, (b) $\mu(n) = 0$ se n ha un fattore quadratico maggiore di 1, (c) $\mu(n) = (-1)^s$ se $n = p_1 \cdots p_s$ con i p_i primi distinti. Provare che μ è moltiplicativa, cioè $\mu(n_1 n_2) = \mu(n_1)\mu(n_2)$ se $(n_1, n_2) = 1$. Provare inoltre che $\sum_{d|n} \mu(d)$ vale 1 per $n = 1$ e 0 altrimenti.

18. Provare la formula d'inversione di Möbius. Se f è una funzione sugli interi positivi a valori in un anello e $g(n) = \sum_{d|n} f(d)$, allora $f(n) = \sum_{d|n} \mu(n/d)g(d)$.

19. Per la funzione φ di Eulero, provare che $\varphi(n) = \sum_{d|n} \mu(n/d)d$.

20. Sia F un campo di $q < \infty$ elementi. Provare che i polinomi quadratici monici irriducibili a coefficienti in F sono $q(q-1)/2$ e i cubici monici irriducibili sono $q(q^2-1)/3$. (Si veda il corollario 2 del teorema 4.26, p. 289 dell'originale.)

2.16 Estensioni polinomiali dei domini fattoriali

Dimostriamo ora un importante teorema: se il dominio D è fattoriale, lo è anche il dominio $D[x]$ dei polinomi in un'indeterminata su D.

In un dominio fattoriale D ogni insieme finito di elementi non nulli ha un massimo comune divisore. Per $a_i \in D$ conviene porre $(a_1, \dots, a_k) = 0$ se tutti gli a_i sono nulli e, altrimenti, definirlo come massimo comune divisore degli a_i non nulli. Per $0 \neq f(x) = a_0 + a_1 x + \cdots + a_n x^n$ chiamiamo *contenuto* $c(f)$ il massimo comune divisore $(a_0, a_1, \dots, a_n) \neq 0$. Posto $d = c(f)$, scriviamo $a_i = da_i'$, $0 \leq i \leq n$, e $f(x) = df_1(x)$ con $f_1(x) = a_0' + a_1' x + \cdots + a_n' x^n$. Nel § 2.14 abbiamo visto che $(da, db) = d(a, b)$ a meno di associati. Per induzione $d(b_1, \dots, b_r) = (db_1, \dots, db_r)$ nello stesso senso. Ne segue che come contenuto $c(f_1)$ si può prendere 1. Un polinomio con questa proprietà si dice *primitivo*. Abbiamo così la fattorizzazione $f = c(f)f_1$ nel prodotto del contenuto e di un polinomio primitivo. Se anche $f = ef_2$ con e costante e $f_2 = a_0'' + a_1'' x + \cdots + a_n'' x^n$ primitivo, si ha $a_i = a_i'' e$ e 1 è un massimo comune divisore degli a_i''. Pertanto e è un massimo comune divisore degli a_i e $e \sim c(f)$.

È utile estendere questa decomposizione ai polinomi a coefficienti nel campo delle frazioni.

Lemma 1. *Siano D un dominio fattoriale, F il suo campo delle frazioni e $0 \neq f(x) \in F[x]$. Allora $f(x) = \gamma f_1(x)$, dove $\gamma \in F$ e $f_1(x) \in D[x]$ è primitivo. La fattorizzazione è unica a meno di fattori invertibili di D.*

Dimostrazione. Scriviamo $f(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ con $\alpha_i \in F$, $\alpha_n \neq 0$. Posti $\alpha_i = a_i b_i^{-1}$ con $a_i, b_i \in D$ e $b = \prod b_i$, si ha $bf \in D[x]$, dunque $bf = cf_1$ con $f_1 \in D[x]$ primitivo. Allora $f = \gamma f_1$ con $\gamma = cb^{-1} \in F$. Se anche $f = \delta f_2$, con $\delta \in F$ e $f_2 \in D[x]$ primitivo, scriviamo $\delta = de^{-1}$ con $d, e \in D$. Da $cb^{-1} f_1 = de^{-1} f_2$ segue $ce f_1 = bd f_2$. Il risultato già provato per coefficienti in D dà $f_1 \sim f_2$ e $ce \sim bd$. Quindi $bd = uce$ per un invertibile $u \in D$ e $de^{-1} = ucb^{-1}$. Pertanto $\delta = u\gamma$, come richiesto. □

Come nel caso di $D[x]$, chiamiamo *contenuto* di $f \in F[x]$ l'elemento γ, determinato a meno di invertibili di D. Dal lemma 1 segue immediatamente un corollario.

Corollario. *Due polinomi primitivi $f, g \in D[x]$ associati in $F[x]$ sono associati in $D[x]$.*

Dimostrazione. Si ha $f = \alpha g$ con $0 \neq \alpha \in F$. L'unicità del lemma 1 mostra che α è invertibile in D. □

Il lemma decisivo per la fattorialità di $D[x]$ è il seguente.

Lemma 2 (lemma di Gauss). *Il prodotto di polinomi primitivi è primitivo.*

Dimostrazione. Supponiamo primitivi f, g ma non $h = fg$. Esiste un irriducibile, e quindi primo, $p \in D$ tale che $p \nmid f$, $p \nmid g$, ma $p \mid h$. Dire che p è primo equivale a dire che $\bar{D} = D/(p)$ è un dominio, come si vede subito dalle definizioni. Quindi anche $\bar{D}[x]$ è un dominio. Applichiamo l'omomorfismo di $D[x]$ su $\bar{D}[x]$ che manda $a \in D$ nella classe $\bar{a} = a + (p)$ e x in x. Otteniamo $\bar{f}(x)\bar{g}(x) = \bar{h}(x) = 0$, ma $\bar{f}(x) \neq 0$, $\bar{g}(x) \neq 0$. Ciò contraddice il fatto che $\bar{D}[x]$ è un dominio. □

Lemma 3. *Se $f \in D[x]$ ha grado positivo ed è irriducibile in $D[x]$, allora è irriducibile in $F[x]$.*

Dimostrazione. Un siffatto f è primitivo. Se fosse riducibile in $F[x]$, avremmo $f = \varphi_1 \varphi_2$ con $\varphi_i \in F[x]$ di grado positivo. Scriviamo $\varphi_i = \alpha_i f_i$ con $\alpha_i \in F$, f_i primitivo in $D[x]$. Allora

$f = \alpha_1\alpha_2 f_1 f_2$, e $f_1 f_2$ è primitivo per il lemma di Gauss. Dunque f e $f_1 f_2$ differiscono per un fattore invertibile in D. Poiché $\deg f_i > 0$, ciò contraddice l'irriducibilità in $D[x]$. □

Possiamo ora dimostrare il teorema.

Teorema 2.25. *Se D è fattoriale, lo è anche $D[x]$.*

Dimostrazione. Sia $f \in D[x]$ non nullo e non invertibile. Scriviamo $f = df_1$ con $d \in D$ e f_1 primitivo. Se $\deg f_1 > 0$, f_1 non è invertibile. Se inoltre non è irriducibile, scriviamo $f_1 = f_{11} f_{12}$ con $0 < \deg f_{1i} < \deg f_1$. Chiaramente ciascun f_{1i} è primitivo. Per induzione sul grado otteniamo $f_1 = q_1 \cdots q_t$ con i q_i irriducibili in $D[x]$. Se d non è invertibile, abbiamo anche $d = p_1 \cdots p_s$ con i p_i irriducibili in D, e quindi irriducibili in $D[x]$. Usando le fattorizzazioni di d e f_1 quando questi non sono invertibili, otteniamo una fattorizzazione di f in irriducibili di $D[x]$.

Resta l'unicità a meno di fattori invertibili. Se f è primitivo, i suoi fattori irriducibili hanno tutti grado positivo. Siano quindi $f = q_1 \cdots q_h = q'_1 \cdots q'_k$ con fattori irriducibili di grado positivo. Per il lemma 3 sono irriducibili in $F[x]$, che è fattoriale. Dunque $h = k$ e, riordinando i q'_i, possiamo supporre che q_i, q'_i siano associati in $F[x]$ per ogni i. Per il corollario del lemma 1 sono associati anche in $D[x]$. Se f non è primitivo, i fattori irriducibili di grado positivo sono primitivi e così il loro prodotto. Pertanto in ogni fattorizzazione di f in irriducibili di $D[x]$ compaiono fattori appartenenti a D, e il loro prodotto è il contenuto di f. Modificando un fattore invertibile possiamo prenderlo identico nelle due fattorizzazioni. Per la fattorialità di D, i fattori irriducibili costanti si possono accoppiare a due a due in elementi associati. Il prodotto degli altri fattori, se ve ne sono, è primitivo e abbiamo già trattato questo caso. □

Ne segue immediatamente che se D è fattoriale, lo è anche $D[x_1, \ldots, x_r]$. Ad esempio sono fattoriali $\mathbb{Z}[x_1, \ldots, x_r]$ e $F[x_1, \ldots, x_r]$ per ogni campo F. La classe dei domini fattoriali è dunque più ampia di quella dei domini a ideali principali (si veda il § 2.11 e l'esercizio 5 qui sotto). Dalla fattorialità di $D[x]$ e dal lemma 3 segue anche questo importante corollario.

Corollario. *Se D è fattoriale e $f(x) \in D[x]$ è monico, ogni fattore monico di $f(x)$ in $F[x]$ appartiene a $D[x]$.*

Dimostrazione. Scriviamo $f = p_1^{e_1} \cdots p_r^{e_r}$ con p_i monici, irriducibili in $D[x]$ e distinti, $e_i > 0$. I fattori monici in $D[x]$ sono i prodotti $p_1^{f_1} \cdots p_r^{f_r}$ con $0 \le f_i \le e_i$. Passando a $F[x]$ i p_i restano irriducibili per il lemma 3. Pertanto i fattori monici di f in $F[x]$ sono gli stessi che in $D[x]$. □

Esercizi

1. Dimostrare che ogni radice razionale di un polinomio monico a coefficienti interi è intera.

2. Provare il criterio di irriducibilità di Eisenstein. Sia $f(x) = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{Z}[x]$. Se esiste un primo p tale che $p \mid a_i$ per $0 \le i < n$, $p \nmid a_n$ e $p^2 \nmid a_0$, allora f è irriducibile in $\mathbb{Q}[x]$.

3. Per un primo p di $\mathbb{Z}$, dimostrare che il polinomio ottenuto sostituendo $x+1$ a x in $x^{p-1} + x^{p-2} + \cdots + 1 = (x^p - 1)/(x-1)$ è irriducibile in $\mathbb{Q}[x]$. Dedurre che il polinomio «ciclotomico» $x^{p-1} + x^{p-2} + \cdots + 1$ è irriducibile in $\mathbb{Q}[x]$.

4. Scomporre in fattori irriducibili in $\mathbb{Z}[x]$ i polinomi $x^3 - 1$, $x^4 - 1$, $x^5 - 1$, $x^6 - 1$, $x^7 - 1$, $x^8 - 1$, $x^9 - 1$, $x^{10} - 1$.

5. Provare che se D è un dominio che non è un campo, $D[x]$ non è un dominio a ideali principali.

6. Sia F un campo e $f(x)$ irriducibile in $F[x]$. Dimostrare che $f(x)$ è irriducibile in $F(t)[x]$ per t indeterminata.

2.17 «Rng»: anelli senza unità

Nella maggior parte dei libri di algebra si definisce anello un insieme non vuoto R dotato di due operazioni binarie $+, \cdot$ e di un elemento 0, tale che $(R, +, 0)$ sia un gruppo abeliano, $(R, \cdot)$ sia un semigruppo (p. 29 dell'originale) e valgano le leggi distributive. In altri termini non si richiede l'esistenza di un'unità moltiplicativa. Considereremo brevemente questi sistemi e, per evitare conflitti con la terminologia precedente, chiameremo *rng*[8] le strutture per cui non si assume l'esistenza di un'unità. Osserviamo anzitutto che le proprietà elementari degli anelli ricordate nel § 2.1 (associatività e distributività generalizzate, regole per i multipli e così via) valgono anche per i rng. Lasciamo al lettore la verifica. Mostriamo ora che ogni rng si può immergere in un anello. Ciò permette di ricondurre al caso degli anelli la maggior parte delle questioni sui rng.

Sia R un rng. Per costruire un anello che lo contenga, prendiamo il prodotto cartesiano $S = \mathbb{Z} \times R$. Per $m, n \in \mathbb{Z}$, $a, b \in R$ definiamo la somma in S mediante

$$(m, a) + (n, b) = (m + n, a + b). \tag{53}$$

Poniamo $0 = (0, 0)$. È chiaro che $(S, +, 0)$ è un gruppo abeliano, precisamente il prodotto diretto, detto anche somma diretta, di $(\mathbb{Z}, +, 0)$ e $(R, +, 0)$. Definiamo il prodotto mediante

$$(m, a)(n, b) = (mn, mb + na + ab), \tag{54}$$

dove mb e na indicano rispettivamente l'm-esimo multiplo di b e l'n-esimo multiplo di a nel gruppo additivo $(R, +, 0)$. Abbiamo

$$\begin{aligned}
((m, a)(n, b))(q, c) &= (mn, mb + na + ab)(q, c) \\
&= ((mn)q, (mn)c + q(mb + na + ab) + (mb + na + ab)c) \\
&= ((mn)q, (mn)c + q(mb) + q(na) + q(ab) + (mb)c + (na)c + (ab)c), \\
(m, a)((n, b)(q, c)) &= (m, a)(nq, nc + qb + bc) \\
&= (m(nq), m(nc) + m(qb) + m(bc) + (nq)a + a(nc + qb + bc)).
\end{aligned}$$

Dall'associatività in $\mathbb{Z}$ e in R, dalla distributività in R e dalle proprietà dei multipli segue dunque l'associatività del prodotto in S. Se $1 = (1, 0)$, si ha $1(m, a) = (1, 0)(m, a) = (m, a) = (m, a)(1, 0) = (m, a)1$. Perciò $(S, \cdot, 1)$ è un monoide. Inoltre

$$\begin{aligned}
(m, a)[(n, b) + (q, c)] &= (m, a)(n + q, b + c) \\
&= (m(n + q), m(b + c) + (n + q)a + a(b + c)), \\
(m, a)(n, b) + (m, a)(q, c) &= (mn, mb + na + ab) + (mq, mc + qa + ac) \\
&= (mn + mq, mb + na + ab + mc + qa + ac).
\end{aligned}$$

Segue $(m, a)[(n, b) + (q, c)] = (m, a)(n, b) + (m, a)(q, c)$. Analogamente vale l'altra legge distributiva. Quindi $(S, +, \cdot, 0, 1)$ è un anello.

Consideriamo il sottoinsieme delle coppie $(0, a)$. Si ha $(0, a) + (0, b) = (0, a + b)$, $(0, a)(0, b) = (0, ab)$ e $0 = (0, 0)$ appartiene al sottoinsieme. Esso è dunque un sotto-rng isomorfo a R, con le ovvie definizioni di questi termini. Abbiamo così dimostrato quanto segue.

[8] L'autore suggerisce la pronuncia inglese «rŭngs» per il plurale *rngs*. Il termine gli è stato suggerito da Louis Rowen.

Teorema 2.26. *Ogni rng si può immergere in un anello.*

Osserviamo anche che R, identificato con il corrispondente sottoinsieme di S, ne è un ideale, poiché $(m,b)(0,a) = (0, ma + ba)$ e $(0,a)(m,b) = (0, ma + ab)$.

Esercizi

1. Un elemento a di un rng R si dice *quasi invertibile a destra* (a sinistra), o *quasi regolare a destra* (a sinistra), se esiste b tale che $a + b - ab = 0$ (rispettivamente $a + b - ba = 0$). Dimostrare che ciò equivale a dire che $1 - a$ ha inverso destro (sinistro) $1 - b$ in $S = \mathbb{Z} \times R$, con la struttura di anello sopra definita.

2. (Kaplansky.) Sia R un rng nel quale tutti gli elementi tranne uno sono quasi invertibili a destra. Provare che R ha un'unità ed è un corpo.

3. Sia R un rng per cui esiste un intero positivo k tale che $ka = 0$ per ogni $a \in R$. Sia $S_k = \mathbb{Z}/(k) \times R$. Scrivendo $\bar{m} = m + (k)$, si ponga

$$(\bar{m}, a) + (\bar{n}, b) = (\bar{m} + \bar{n}, a + b),$$
$$(\bar{m}, a)(\bar{n}, b) = (\overline{mn}, mb + na + ab), \qquad 0 = (\bar{0}, 0), \quad 1 = (\bar{1}, 0).$$

   Verificare che $(S_k, +, \cdot, 0, 1)$ è un anello di caratteristica k e che R vi si immerge.

4. Sia R un rng privo di divisori dello zero non nulli, ossia in cui $ab = 0$ implica $a = 0$ o $b = 0$. Supponiamo che contenga $a, b \neq 0$ tali che $ab + kb = 0$ per qualche intero positivo k. Mostrare che $ca + kc = 0 = ac + kc$ per ogni $c \in R$.

5. Sia R un rng privo di divisori dello zero non nulli e sia $S = \mathbb{Z} \times R$ l'anello del testo. Sia $Z = \{z \in S \mid za = 0 \text{ per ogni } a \in R\}$. Dimostrare che Z è un ideale di S e S/Z un dominio. Provare che $a \mapsto a + Z$ è un monomorfismo di R in S/Z.

Capitolo 3
Moduli su un dominio a ideali principali

Il concetto centrale dello sviluppo assiomatico dell'algebra lineare è quello di spazio vettoriale su un campo. L'assiomatizzazione dell'algebra lineare, realizzata negli anni Venti, fu motivata in larga misura dal desiderio di introdurre nozioni geometriche nello studio di certe classi di funzioni in analisi. In un primo tempo si considerarono esclusivamente spazi vettoriali sui reali o sui complessi. Divenne presto evidente che questa restrizione era alquanto artificiosa, poiché molti risultati dipendevano soltanto dalla risoluzione di equazioni lineari e valevano quindi su campi arbitrari. Si giunse così allo studio degli spazi vettoriali su un campo qualsiasi: è questo che costituisce oggi l'algebra lineare.

Il concetto di modulo generalizza immediatamente quello di spazio vettoriale. Basta sostituire al campo di base un anello qualsiasi. Perché introdurre questa generalizzazione? In primo luogo l'esperienza insegna che la struttura logica interna della matematica spinge a perseguire simili generalizzazioni «naturali». Spesso esse permettono di comprendere meglio la teoria stessa da cui hanno preso origine. Ne è un buon esempio lo studio di una trasformazione lineare di uno spazio vettoriale di dimensione finita su un campo, problema centrale dell'algebra lineare. Come vedremo nei §§ 3.2 e 3.10, data una trasformazione lineare T di uno spazio vettoriale V su F, possiamo servircene per trasformare V in un modulo sull'anello dei polinomi $F[\lambda]$, dove λ è un'indeterminata.[1] Lo studio di tale modulo condurrà alla teoria delle forme canoniche delle matrici di una trasformazione lineare e alla soluzione del problema della similitudine fra matrici.

Il passo dai moduli su $F[\lambda]$ ai moduli su un dominio a ideali principali qualsiasi è breve e ci darà ulteriori applicazioni. In particolare, prendendo come dominio $\mathbb{Z}$, otterremo la teoria della struttura dei gruppi abeliani finitamente generati e quindi dei gruppi abeliani finiti.

Sarebbe erroneo concludere che lo sviluppo storico della teoria dei moduli abbia seguito il percorso logico ora indicato, per estensione dell'algebra lineare. Il concetto di modulo sembra avere fatto la sua prima comparsa nella teoria algebrica dei numeri, nello studio di sottoinsiemi di anelli di numeri algebrici chiusi rispetto all'addizione e alla moltiplicazione per elementi di un sottoanello fissato. I moduli divennero uno strumento importante dell'algebra verso la fine degli anni Venti, soprattutto grazie all'intuizione di Emmy Noether, che per prima ne comprese le potenzialità. In particolare, osservò che questo concetto poteva collegare due importanti sviluppi dell'algebra che fino ad allora avevano proceduto affiancati e indipendenti: la teoria delle rappresentazioni (ossia degli omomorfismi) dei gruppi finiti mediante matrici, dovuta a Frobenius, Burnside e Schur, e la teoria della struttura delle algebre, dovuta a Molien, Cartan e Wedderburn. Tratteremo questi argomenti nel volume II. Più di recente si è sviluppata l'algebra omologica, in cui i moduli hanno ancora un ruolo centrale. Anche di questo si occupa il secondo volume.

L'argomento principale di questo capitolo è lo studio dei moduli finitamente generati su un dominio a ideali principali D, con i due casi particolari in cui D è $\mathbb{Z}$ oppure un anello di polinomi $F[\lambda]$ su un campo F. Come abbiamo osservato, ne risultano rispettivamente la teoria della struttura dei gruppi abeliani finitamente generati e le forme canoniche delle trasformazioni lineari. Naturalmente dovremo cominciare con un po' di teoria generale, che svilupperemo solo di poco oltre ciò che occorre per raggiungere i nostri obiettivi immediati. Gran parte della teoria generale dei moduli e altre applicazioni sono discusse nel nostro secondo volume.

[1] In questo capitolo indichiamo un'indeterminata con λ per riservare x ai vettori o, più in generale, agli elementi di un modulo.

3.1 L'anello degli endomorfismi di un gruppo abeliano

Sia M un gruppo abeliano. Usiamo in M la notazione additiva: $+$ per la composizione binaria, 0 per l'elemento neutro, $-a$ per l'inverso di a e ma, con $m \in \mathbb{Z}$, per la potenza m-esima. Indichiamo con $\operatorname{End} M$ l'insieme degli endomorfismi di M. Per definizione sono le applicazioni $\eta : M \to M$ tali che

$$\eta(x+y) = \eta(x) + \eta(y), \qquad \eta(0) = 0.$$

Abbiamo visto che la seconda condizione segue dalla prima. Dunque un'applicazione $\eta : M \to M$ è un endomorfismo se e solo se

$$\eta(x+y) = \eta(x) + \eta(y) \qquad (x, y \in M). \tag{1}$$

Ricordiamo che ne segue $\eta(mx) = m\eta(x)$ per ogni $m \in \mathbb{Z}$. Inoltre, se X genera M, l'azione di η su X determina η: se due endomorfismi η, ζ soddisfano $\eta(x) = \zeta(x)$ per ogni x di un insieme di generatori, allora $\eta = \zeta$.

Esempi.

1. Sia M il gruppo ciclico infinito $(\mathbb{Z}, +, 0)$. Allora 1 è un generatore e, se $\eta(1) = m$, si ha $\eta(x) = \eta(x1) = x\eta(1) = xm$. Dunque η è l'applicazione $x \mapsto mx$, $x \in M$, con $m = \eta(1)$. Viceversa, per ogni intero m, $x \mapsto mx$ è un endomorfismo in virtù della regola $m(x+y) = mx+my$ e manda 1 in m. Poiché un endomorfismo è determinato dall'immagine del generatore 1, abbiamo una corrispondenza biunivoca fra $\operatorname{End} M$ e $\mathbb{Z}$, che associa a η l'intero $\eta(1) = m$.

2. Sia $M = (\mathbb{Z}^{(2)}, +, 0)$ il prodotto diretto (o somma diretta) di due copie di $(\mathbb{Z}, +, 0)$. I suoi elementi sono coppie di interi $(x, y) = x(1, 0) + y(0, 1)$, sicché $e = (1, 0)$ e $f = (0, 1)$ generano $\mathbb{Z}^{(2)}$. Ogni $\eta \in \operatorname{End} \mathbb{Z}^{(2)}$ è determinato dalla coppia $\eta(e), \eta(f)$. Ogni coppia $(u, v) \in \mathbb{Z}^{(2)} \times \mathbb{Z}^{(2)}$ si ottiene in questo modo: dato (u, v), esiste un endomorfismo η con $\eta(e) = u$ e $\eta(f) = v$. Basta definire $\eta(x, y) = xu + yv$. Allora $(x', y') \mapsto x'u + y'v$ e $(x + x', y + y') \mapsto (x + x')u + (y + y')v = (xu + yv) + (x'u + y'v)$. Dunque η è un omomorfismo e ha le immagini prescritte. Otteniamo così una corrispondenza biunivoca fra $\operatorname{End} \mathbb{Z}^{(2)}$ e $\mathbb{Z}^{(2)} \times \mathbb{Z}^{(2)}$, che associa a η la coppia $(\eta(e), \eta(f))$. Le stesse considerazioni valgono per $M = \mathbb{Z}^{(n)}$, per ogni intero positivo n, e danno una corrispondenza biunivoca fra $\operatorname{End} \mathbb{Z}^{(n)}$ e il prodotto di n copie di $\mathbb{Z}^{(n)}$.

3. Sia M un gruppo ciclico finito. Possiamo prendere $M = (\mathbb{Z}/(n), +, 0)$, con n intero positivo e $\bar{x} = x + (n)$. Allora $\bar{1}$ è un generatore e abbiamo una corrispondenza biunivoca fra $\operatorname{End} \mathbb{Z}/(n)$ e $\mathbb{Z}/(n)$, data da $\eta \mapsto \eta(\bar{1})$.

Ora daremo a $\operatorname{End} M$, per ogni gruppo abeliano M, una struttura di anello. Se $\eta, \zeta \in \operatorname{End} M$, anche la composta $\eta\zeta$ appartiene a $\operatorname{End} M$, e vale l'associatività $(\eta\zeta)\rho = \eta(\zeta\rho)$. Anche l'identità $1 : x \mapsto x$ è un endomorfismo, sicché $(\operatorname{End} M, \cdot, 1)$ è un monoide. Tutto questo vale anche se M non è abeliano. Nel caso abeliano, tuttavia, possiamo dire molto di più: $\operatorname{End} M$, con la composizione come prodotto e con l'addizione e lo zero che ora definiremo, è un anello. Se $\eta, \zeta \in \operatorname{End} M$, poniamo

$$(\eta + \zeta)(x) = \eta(x) + \zeta(x). \tag{2}$$

Questa applicazione di M in sé è un endomorfismo, perché

$$\begin{aligned}(\eta+\zeta)(x+y) &= \eta(x+y)+\zeta(x+y)\\ &= \eta(x)+\eta(y)+\zeta(x)+\zeta(y)\\ &= \eta(x)+\zeta(x)+\eta(y)+\zeta(y)\\ &= (\eta+\zeta)(x)+(\eta+\zeta)(y).\end{aligned}$$

Nel passaggio dalla seconda alla terza riga è intervenuta la commutatività dell'addizione. Definiamo 0 mediante $x \mapsto 0$. È un endomorfismo e $\eta + 0 = \eta = 0 + \eta$ per ogni η. Sia $-\eta$ l'applicazione $x \mapsto -\eta(x)$, composta di η e dell'automorfismo $x \mapsto -x$ del gruppo abeliano M. Dunque $-\eta \in \operatorname{End} M$ e $\eta + (-\eta) = 0 = -\eta + \eta$. Poiché $((\eta+\zeta)+\rho)(x) = (\eta+\zeta)(x)+\rho(x) = \eta(x)+\zeta(x)+\rho(x)$ e $(\eta+(\zeta+\rho))(x) = \eta(x)+(\zeta+\rho)(x) = \eta(x)+\zeta(x)+\rho(x)$, l'addizione è associativa. È anche commutativa, perché $(\eta+\zeta)(x) = \eta(x)+\zeta(x) = \zeta(x)+\eta(x) = (\zeta+\eta)(x)$. Così $(\operatorname{End} M, +, 0)$ è un gruppo abeliano.

Sappiamo già che $(\operatorname{End} M, \cdot, 1)$ è un monoide. Per $\eta, \zeta, \rho \in \operatorname{End} M$ si ha poi

$$(\rho(\eta+\zeta))(x) = \rho(\eta(x)+\zeta(x)) = (\rho\eta)(x)+(\rho\zeta)(x) = (\rho\eta+\rho\zeta)(x).$$

Analogamente, $((\eta+\zeta)\rho)(x) = \eta(\rho(x))+\zeta(\rho(x))$. Valgono quindi entrambe le leggi distributive e abbiamo verificato il seguente teorema fondamentale.

Teorema 3.1. *Sia M un gruppo abeliano in notazione additiva e sia $\operatorname{End} M$ l'insieme dei suoi endomorfismi. Per $\eta, \zeta \in \operatorname{End} M$ si definiscano $\eta\zeta$ e $\eta+\zeta$ mediante $(\eta\zeta)(x) = \eta(\zeta(x))$ e $(\eta+\zeta)(x) = \eta(x)+\zeta(x)$, e si definiscano 1 e 0 mediante $1x = x$, $0x = 0$. Allora $(\operatorname{End} M, +, \cdot, 0, 1)$ è un anello.*

Chiameremo $(\operatorname{End} M, +, \cdot, 0, 1)$, o più brevemente $\operatorname{End} M$, l'*anello degli endomorfismi* del gruppo abeliano M. Riprendiamo gli esempi precedenti e identifichiamo questi anelli.

1. $M = (\mathbb{Z}, +, 0)$. Abbiamo visto che $\eta \mapsto \eta(1)$ è una biiezione di $\operatorname{End} M$ su $\mathbb{Z}$. Essa manda $\eta+\zeta$ in $(\eta+\zeta)(1) = \eta(1)+\zeta(1)$, $\eta\zeta$ in $(\eta\zeta)(1) = \eta(\zeta(1)) = \zeta(1)\eta(1) = \eta(1)\zeta(1)$ e 1 in $1(1) = 1$. Dunque $\eta \mapsto \eta(1)$ è un isomorfismo di $\operatorname{End} M$ con l'anello degli interi $\mathbb{Z}$. In questo senso l'anello degli endomorfismi di un gruppo ciclico infinito è $\mathbb{Z}$.

2. $M = (\mathbb{Z}^{(2)}, +, 0)$. Abbiamo la biiezione $\eta \mapsto (\eta(e), \eta(f))$ di $\operatorname{End} M$ su $\mathbb{Z}^{(2)} \times \mathbb{Z}^{(2)}$, con $e = (1,0)$ e $f = (0,1)$. Se $\eta(e) = (a,b)$ e $\eta(f) = (c,d)$, è evidente che

$$\eta \longmapsto \begin{pmatrix} a & c \\ b & d \end{pmatrix} \tag{3}$$

è una biiezione di $\operatorname{End} M$ sull'anello $M_2(\mathbb{Z})$ delle matrici intere 2×2. Mostriamo che è un isomorfismo. Se $\zeta(e) = (a',b')$ e $\zeta(f) = (c',d')$, allora

$$\zeta \longmapsto \begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix}. \tag{4}$$

Ora $(\eta+\zeta)(e) = \eta(e)+\zeta(e) = (a+a', b+b')$ e analogamente $(\eta+\zeta)(f) = (c+c', d+d')$. Ne segue

$$\eta+\zeta \longmapsto \begin{pmatrix} a+a' & c+c' \\ b+b' & d+d' \end{pmatrix},$$

la somma delle matrici (3) e (4). Calcoliamo poi $(\eta\zeta)(e) = \eta(\zeta(e)) = \eta(a'e + b'f) = \eta(a'e) + \eta(b'f) = a'\eta(e) + b'\eta(f) = a'(a,b) + b'(c,d) = (aa' + cb', ba' + db')$. Analogamente $(\eta\zeta)(f) = (ac' + cd', bc' + dd')$. Pertanto

$$\eta\zeta \longmapsto \begin{pmatrix} aa' + cb' & ac' + cd' \\ ba' + db' & bc' + dd' \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a' & c' \\ b' & d' \end{pmatrix},$$

il prodotto della matrice (3) per la (4). Inoltre $1(e) = (1,0)$ e $1(f) = (0,1)$, quindi 1 corrisponde alla matrice identità. Questo prova che la (3) è un isomorfismo di $\operatorname{End} M$ con $M_2(\mathbb{Z})$.

3. Per M ciclico di ordine n si vede come nel primo esempio che $\operatorname{End} M$ è isomorfo all'anello $\mathbb{Z}/(n)$.

Il fatto che $\operatorname{End} M$ sia un anello rispetto alle operazioni e agli elementi $0, 1$ così definiti è analogo al fatto che le applicazioni biiettive di un insieme in sé formino un gruppo con la composizione usuale e l'identità. Chiamiamo ora *anello di endomorfismi* qualsiasi sottoanello di un anello $\operatorname{End} M$, con M gruppo abeliano. Dimostreremo l'analogo per gli anelli del teorema di Cayley per i gruppi (p. 38 dell'originale).

Teorema 3.2. *Ogni anello è isomorfo a un anello di endomorfismi di un gruppo abeliano.*[2]

Dimostrazione. L'idea è identica a quella del teorema di Cayley. Dato l'anello R, prendiamo come M il gruppo additivo $(R, +, 0)$ e per $a \in R$ chiamiamo $a_L : x \mapsto ax$ la *moltiplicazione sinistra* determinata da a.[3] Poiché $a_L(x + y) = a(x + y) = ax + ay = a_L x + a_L y$, si ha $a_L \in \operatorname{End} M$. Inoltre $(a + b)_L x = (a + b)x = ax + bx = a_L x + b_L x = (a_L + b_L)x$ per definizione della somma di endomorfismi; $(ab)_L x = (ab)x = a(bx) = a_L(b_L x) = (a_L b_L)(x)$ e $1x = x$. Dunque $a \mapsto a_L$ è un omomorfismo di anelli da R a $\operatorname{End} M$. Da $a_L = b_L$ segue $a = a_L 1 = b_L 1 = b$, perciò è un monomorfismo. L'immagine è un sottoanello R_L di $\operatorname{End} M$ e $R \cong R_L$. □

È interessante considerare anche le moltiplicazioni destre. Definiamo $a_R : x \mapsto xa$; è un endomorfismo di $(R, +, 0)$ perché $(x + y)a = xa + ya$. Si vede subito che $a \mapsto a_R$ è un antiomomorfismo di R in $\operatorname{End} M$. L'immagine $R_R = \{a_R\}$ è un sottoanello di $\operatorname{End} M$ antiisomorfo a R. I sottoanelli R_L e R_R sono ciascuno il centralizzante dell'altro, come afferma il seguente risultato.

Teorema 3.3. *In* $\operatorname{End} M$ *valgono* $R_L = C(R_R)$ *e* $R_R = C(R_L)$.

Dimostrazione. Da $(ax)b = a(xb)$ segue $a_L b_R = b_R a_L$ per ogni $a, b \in R$. Sia ora η un endomorfismo di M con $a_L \eta = \eta a_L$ per ogni $a \in R$. Allora $\eta(x) = \eta(x1) = \eta(x_L 1) = x_L(\eta(1)) = x\eta(1)$, sicché $\eta = \eta(1)_R \in R_R$. Dunque $C(R_L) = R_R$ e, per simmetria, $C(R_R) = R_L$. □

Esercizi

1. Sia G un gruppo in notazione moltiplicativa e sia $F = G^G$ l'insieme delle applicazioni di G in sé. Per $\eta, \zeta \in F$ si definisca come al solito $\eta\zeta$ come la composizione di η dopo ζ. Si definisca $\eta + \zeta$ mediante $(\eta + \zeta)(x) = \eta(x)\zeta(x)$ e si ponga $1 : x \mapsto x$, $0 : x \mapsto 1$. Studiare le proprietà della struttura $(F, +, \cdot, 0, 1)$.

[2] In una situazione un po' più particolare, quella delle algebre, questo risultato sembra essere stato osservato per primo da Poincaré.

[3] Ricordiamo che per i gruppi preferivamo il termine «traslazione» per un'applicazione di questo tipo.

2. Osservare che per un gruppo abeliano M, $\operatorname{Aut} M$ è il gruppo delle unità (elementi invertibili) di $\operatorname{End} M$. Dedurne che per il gruppo ciclico di ordine n, $\operatorname{Aut} M$ è isomorfo al gruppo delle classi $\bar{m} = m + (n)$ di $\mathbb{Z}/(n)$ tali che $(m, n) = 1$.

3. Determinare $\operatorname{Aut} M$ per $M = (\mathbb{Z}^{(2)}, +, 0)$.

4. Determinare $\operatorname{End}(\mathbb{Q}, +, 0)$.

5. In vari casi abbiamo trovato $\operatorname{End}(R, +, 0) \cong R$ per un anello R. Vale in generale? Vale se R è un campo?

3.2 Moduli sinistri e destri

Il concetto di modulo sinistro è l'analogo, per gli anelli, dell'azione di un gruppo su un insieme. Come nel caso dei gruppi, esso nasce considerando un omomorfismo η di un anello dato R nell'anello $\operatorname{End} M$ degli endomorfismi di un gruppo abeliano M. Poiché $\eta(a) \in \operatorname{End} M$, si ha

$$\eta(a)(x + y) = \eta(a)(x) + \eta(a)(y) \qquad (x, y \in M);$$

inoltre, essendo η un omomorfismo,

$$\begin{aligned} \eta(a + b)(x) &= (\eta(a) + \eta(b))(x) = \eta(a)(x) + \eta(b)(x), \\ \eta(ab)(x) &= (\eta(a)\eta(b))(x) = \eta(a)(\eta(b)(x)), \\ \eta(1)(x) &= x \end{aligned}$$

per $x \in M$, $a, b \in R$. Consideriamo l'applicazione $(a, x) \mapsto \eta(a)(x)$ da $R \times M$ a M e scriviamo ax per $\eta(a)(x)$. Le identità precedenti diventano, per $x, y \in M$ e $a, b, 1 \in R$,

1. $a(x + y) = ax + ay$,
2. $(a + b)x = ax + bx$,
3. $(ab)x = a(bx)$,
4. $1x = x$.

Definizione 3.1. *Se R è un anello, un R-modulo sinistro è un gruppo abeliano M munito di un'applicazione $(a, x) \mapsto ax$ da $R \times M$ a M che soddisfa le proprietà 1–4.*

Abbiamo visto che un omomorfismo $\eta : R \to \operatorname{End} M$ determina su M una struttura di R-modulo sinistro ponendo $ax = \eta(a)(x)$ per $a \in R$, $x \in M$. Viceversa, sia dato un R-modulo sinistro M. Per $a \in R$ indichiamo con a_L l'applicazione $x \mapsto ax$ di M in sé. In virtù della proprietà 1, $a_L \in \operatorname{End} M$. Le proprietà 2–4 dicono inoltre che $a \mapsto a_L$ è un omomorfismo di R in $\operatorname{End} M$. Il modulo che il procedimento precedente associa a questo omomorfismo è il modulo sinistro di partenza. Reciprocamente, se partiamo da un omomorfismo η di R in $\operatorname{End} M$ e costruiamo il corrispondente R-modulo sinistro, l'omomorfismo associato $a \mapsto a_L$ coincide con η, dato che $a_L x = ax = \eta(a)(x)$. Pertanto la nozione di R-modulo sinistro equivale a quella di omomorfismo di R nell'anello degli endomorfismi di un gruppo abeliano.

La nozione di R-modulo destro è duale a quella di R-modulo sinistro.

Definizione 3.1′. *Un modulo destro su un anello R è un gruppo abeliano M munito di un'applicazione $(x, a) \mapsto xa$ da $M \times R$ a M che, per $a, b, 1 \in R$ e $x, y \in M$, soddisfa*

1′. $(x + y)a = xa + ya$,
2′. $x(a + b) = xa + xb$,
3′. $x(ab) = (xa)b$,
4′. $x1 = x$.

Indichiamo con a_R l'applicazione $x \mapsto xa$ di M in sé. Allora $a_R \in \operatorname{End} M$, e $(a+b)_R = a_R + b_R$, $(ab)_R = b_R a_R$, $1_R = 1$: dunque $a \mapsto a_R$ è un antiomomorfismo di R in $\operatorname{End} M$ (§ 2.8, p. 114 dell'originale). Viceversa, se $\eta : R \to \operatorname{End} M$ è un antiomomorfismo, il gruppo abeliano M diviene un R-modulo destro ponendo $xa = \eta(a)(x)$.

Ogni antiomomorfismo η di R si può considerare un omomorfismo dell'anello opposto R^0 (p. 113 dell'originale): infatti l'identità è un antiisomorfismo da R^0 su R, e la sua composta con η è un omomorfismo. Ne segue che, se M è un R-modulo destro (sinistro), ponendo $ax = xa$ (rispettivamente $xa = ax$) se ne fa un R^0-modulo sinistro (destro). Se R è commutativo, $R^0 = R$ come anelli, sicché ogni R-modulo sinistro (destro) è anche un R-modulo destro (sinistro) con $ax = xa$. Per gli anelli commutativi i due casi coincidono.

Consideriamo alcuni importanti esempi. Ogni gruppo abeliano M, scritto additivamente, è uno $\mathbb{Z}$-modulo: per $a \in \mathbb{Z}$ e $x \in M$ si definisce ax nel modo consueto. Le condizioni 1–4 seguono dalle proprietà dei multipli in un gruppo abeliano. Il fatto che i gruppi abeliani siano $\mathbb{Z}$-moduli permette di ricondurne la teoria a quella dei moduli. L'utilità di questa riduzione apparirà nel seguito.

Un esempio probabilmente familiare al lettore è uno spazio vettoriale V su un campo F. Ricordiamo che, per definizione assiomatica, esso è un gruppo abeliano V munito di un prodotto $ax \in V$ per $a \in F$, $x \in V$, che soddisfa le condizioni 1–4. Esso è dunque un F-modulo sinistro. Sia ora T una trasformazione lineare di V. Scriviamo Tx per $T(x)$. Le condizioni che la definiscono sono $T : V \to V$ e

$$T(x+y) = Tx + Ty, \qquad T(ax) = a(Tx) \qquad (a \in F,\ x, y \in V). \tag{5}$$

La prima dice che $T \in \operatorname{End} V$; la seconda che $a_L T = T a_L$ per ogni endomorfismo $a_L : x \mapsto ax$, $a \in F$. Il sottoanello $F_L[T]$ generato da $F_L = \{a_L \mid a \in F\}$ e T è perciò un sottoanello commutativo di $\operatorname{End} V$. Poiché $a \mapsto a_L$ è un omomorfismo di F, la proprietà fondamentale degli omomorfismi di $F[\lambda]$, con λ indeterminata (teorema 2.10, p. 122 dell'originale), mostra che l'applicazione

$$a_0 + a_1\lambda + \cdots + a_m\lambda^m \longmapsto (a_0)_L + (a_1)_L T + \cdots + (a_m)_L T^m$$

è un omomorfismo di $F[\lambda]$ in $F_L[T]$, e quindi in $\operatorname{End} V$. Pertanto V diviene un $F[\lambda]$-modulo sinistro ponendo

$$(a_0 + a_1\lambda + \cdots + a_m\lambda^m)x = a_0 x + a_1(Tx) + \cdots + a_m(T^m x)$$

per ogni $f(\lambda) = a_0 + a_1\lambda + \cdots + a_m\lambda^m \in F[\lambda]$. Vedremo che la teoria di una singola trasformazione lineare di uno spazio vettoriale di dimensione finita si deduce considerando in questo modo lo spazio come un $F[\lambda]$-modulo.

Come ultimo esempio prendiamo un anello R e il suo gruppo additivo $M = (R, +, 0)$. Facciamo operare R su M mediante moltiplicazione a sinistra: per $a \in R$, $x \in M$, ax è il prodotto definito in R. Le condizioni 1–4 sono evidenti e M è un R-modulo sinistro. Analogamente si ottiene un R-modulo destro prendendo per xa, con $x \in M$, $a \in R$, il prodotto dell'anello.

Esercizi

1. Sia M un R-modulo sinistro e sia $\eta : S \to R$ un omomorfismo di anelli. Dimostrare che M diventa un S-modulo sinistro ponendo $ax = \eta(a)(x)$ per $a \in S$, $x \in M$.

2. Per un R-modulo sinistro M si ponga $B = \{b \in R \mid bx = 0 \text{ per ogni } x \in M\}$. Verificare che B è un ideale di R. Dimostrare inoltre che, per ogni ideale C contenuto in B, M diventa un R/C-modulo sinistro ponendo $(a + C)x = ax$.

3. Sia M un R-modulo sinistro e sia S un sottoanello di R. Dimostrare che M è un S-modulo sinistro prendendo per bx, con $b \in S$, $x \in M$, il prodotto dato nell'R-modulo sinistro M. (È un caso particolare dell'esercizio 1.) In particolare R stesso si può considerare in questo modo un S-modulo sinistro.

4. Sia $V = \mathbb{R}^{(n)}$ lo spazio vettoriale delle n-uple di numeri reali con la somma e il prodotto per elementi di $\mathbb{R}$ consueti. Sia T la trasformazione lineare definita da $x = (x_1, x_2, \ldots, x_n) \mapsto Tx = (x_n, x_1, x_2, \ldots, x_{n-1})$. Considerare V come $\mathbb{R}[\lambda]$-modulo sinistro nel modo descritto nel testo. Determinare (a) λx, (b) $(\lambda^2 + 2)x$, (c) $(\lambda^{n-1} + \lambda^{n-2} + \cdots + 1)x$. Quali elementi soddisfano $(\lambda^2 - 1)x = 0$?

5. Nell'esempio dell'esercizio 4, sia B l'ideale di $\mathbb{R}[\lambda]$ definito come nell'esercizio 2. Descrivere esplicitamente B.

6. Sia M un gruppo abeliano scritto additivamente. Mostrare che esiste un solo modo di dare a M una struttura di $\mathbb{Z}$-modulo sinistro.

7. Sia M un $\mathbb{Q}$-modulo sinistro. Mostrare che l'azione data di $\mathbb{Q}$ è l'unica con cui si possa fare di M un $\mathbb{Q}$-modulo sinistro.

8. Sia M un gruppo abeliano finito diverso da zero. È possibile farne un $\mathbb{Q}$-modulo sinistro?

3.3 Nozioni e risultati fondamentali

D'ora in poi ci occuperemo quasi esclusivamente di moduli sinistri, che chiameremo semplicemente «moduli», «R-moduli» o «moduli su R», dove R è l'anello dato. Ciò che diremo vale naturalmente anche per i moduli destri. Useremo la specificazione «destro» quando vorremo enunciare esplicitamente un risultato per questi ultimi.

Sia M un R-modulo. Poiché $x \mapsto ax$ è un endomorfismo di $(M, +, 0)$, valgono $a0 = 0$ e $a(-x) = -ax$, per $x \in M$, $a \in R$. Poiché $a \mapsto a_L$ è un omomorfismo di R in $\operatorname{End} M$, valgono $0x = 0$ e $(-a)x = -ax$. Per induzione si ottiene inoltre $a(\sum x_i) = \sum ax_i$ e $(\sum a_i)x = \sum a_i x$.

Un *sottomodulo* N di M è un sottogruppo del gruppo additivo $(M, +, 0)$ chiuso rispetto all'azione degli elementi di R: se $a \in R$ e $y \in N$, allora $ay \in N$. Esplicitamente, affinché un sottoinsieme non vuoto N di M sia un sottomodulo occorre e basta che: (a) se $y_1, y_2 \in N$, allora $y_1 + y_2 \in N$; (b) se $y \in N$ e $a \in R$, allora $ay \in N$. I sottomoduli soddisfano certamente queste condizioni. Viceversa, se N le soddisfa, scelto $y \in N$, contiene $0 = 0y$ e contiene $-y = (-1)y$. È quindi un sottogruppo del gruppo additivo e perciò un sottomodulo di M.

Quali sono i sottomoduli negli esempi del § 3.2? Sia dapprima M uno $\mathbb{Z}$-modulo. Se N è un sottogruppo di $(M, +, 0)$, n è un intero positivo e $y \in N$, allora $ny = y + \cdots + y$ (n termini) appartiene a N. Anche $0y$ e $(-n)y$ appartengono a N, dunque N è uno $\mathbb{Z}$-sottomodulo. Il viceversa è evidente: gli $\mathbb{Z}$-sottomoduli di M sono precisamente i sottogruppi di $(M, +, 0)$. Se V è uno spazio vettoriale su un campo F, risulta dalle definizioni che i sottomoduli sono i sottospazi di V. Sia ora T una trasformazione lineare di V e consideriamo V come $F[\lambda]$-modulo secondo il § 3.2. I sottomoduli sono precisamente i sottospazi W stabili per T, ossia tali che $TW = T(W) \subset W$: questa condizione equivale a $\lambda w \in W$ per $w \in W$, da cui $(a_0 + a_1\lambda + \cdots + a_n\lambda^n)w \in W$. Consideriamo infine R come R-modulo sinistro ($M = (R, +, 0)$ e l'azione è la moltiplicazione a sinistra). I sottomoduli sono i sottoinsiemi non vuoti di R chiusi rispetto all'addizione e alla moltiplicazione a sinistra per ogni elemento di R. Un tale sottoinsieme si chiama *ideale sinistro* di R (cfr. esercizio 4 a p. 103 dell'originale). Analogamente i sottomoduli di R come R-modulo destro sono gli *ideali destri*, chiusi rispetto alla somma e alla moltiplicazione a destra per ogni elemento di R.

Se $\{N_\alpha\}$ è una famiglia di sottomoduli di M, anche $\bigcap N_\alpha$ è un sottomodulo. Per un sottoinsieme non vuoto S, l'intersezione $\langle S\rangle$ di tutti i sottomoduli di M che contengono S è quindi un sottomodulo, detto *sottomodulo generato da* S: contiene S ed è contenuto in ogni sottomodulo contenente S. È immediato che $\langle S\rangle$ consiste degli elementi $a_1y_1+\cdots+a_ry_r$ con $a_i\in R$ e $y_i\in S$. Se $\{N_\alpha\}$ è una famiglia di sottomoduli, il sottomodulo generato da $\bigcup N_\alpha$ consiste delle somme $y_{\alpha_1}+\cdots+y_{\alpha_r}$ con $y_{\alpha_k}\in N_{\alpha_k}$. Lo chiamiamo *sottomodulo generato dagli* N_α e lo indichiamo con $\sum N_\alpha$. Per una famiglia finita $\{N_1,\dots,N_m\}$ scriviamo $\sum N_i$ o $N_1+\cdots+N_m$.

Consideriamo ora il gruppo quoziente $\overline{M}=M/N$ di M rispetto al sottomodulo N. I suoi elementi sono le classi $\bar x=x+N$, con addizione $(x_1+N)+(x_2+N)=x_1+x_2+N$, elemento zero N e $-(x+N)=-x+N$. Se $a\in R$ e $x_1\equiv x_2 \pmod N$, ossia $x_2-x_1\in N$, allora $ax_2-ax_1=a(x_2-x_1)\in N$, sicché $ax_1\equiv ax_2 \pmod N$. Ponendo dunque

$$a\bar x=a(x+N)=ax+N=\overline{ax}, \tag{6}$$

la classe così ottenuta è indipendente dal rappresentante x. L'applicazione $(a,\bar x)\mapsto\overline{ax}$ è perciò definita su $R\times\overline{M}$ a valori in $\overline{M}$. Inoltre

$$a(\bar x_1+\bar x_2)=a\overline{x_1+x_2}=\overline{ax_1+ax_2}=\overline{ax_1}+\overline{ax_2}=a\bar x_1+a\bar x_2.$$

Analogamente $(a+b)\bar x=a\bar x+b\bar x$, $(ab)\bar x=a(b\bar x)$ e $1\bar x=\bar x$. Con l'azione (6), $\overline{M}=M/N$ è dunque un R-modulo, detto *modulo quoziente* di M rispetto a N.

Definiamo gli omomorfismi fra moduli soltanto quando gli anelli su cui sono definiti coincidono. Un *omomorfismo* (di moduli, o *R-omomorfismo*, o *omomorfismo su R*) da M a M' è un'applicazione $\eta: M\to M'$ che è un omomorfismo dei gruppi additivi e soddisfa $\eta(ax)=a\eta(x)$ per $a\in R$, $x\in M$. Dalla (6) segue che, dato un sottomodulo N, l'applicazione naturale $\nu: x\mapsto\bar x=x+N$ è un omomorfismo di moduli da M a $\overline{M}$.

Il nucleo di un omomorfismo $\eta: M\to M'$ è il nucleo $\eta^{-1}(0)$ dell'omomorfismo di gruppi. È un sottogruppo di M e, poiché da $\eta(y)=0$ segue $\eta(ay)=a\eta(y)=0$, $\ker\eta$ è un sottomodulo di M. L'immagine $\eta(M)$, ossia $\operatorname{im}\eta=\{\eta(x)\mid x\in M\}$, è un sottomodulo di M': è un sottogruppo e, se $y=\eta(x)$, allora $ay=a\eta(x)=\eta(ax)\in\eta(M)$. Come per i gruppi, se N è un sottomodulo contenuto in $\ker\eta$, l'applicazione

$$\bar\eta:\bar x=x+N\longmapsto\eta(x) \tag{7}$$

è un omomorfismo di moduli da M/N in M' con $\eta=\bar\eta\nu$, dove $\nu: x\mapsto x+N$. Inoltre $\bar\eta$ è un monomorfismo se e solo se $N=\ker\eta$. Otteniamo così il *teorema fondamentale di omomorfismo per i moduli*: ogni omomorfismo $\eta: M\to M'$ si fattorizza come $\bar\eta\nu$, dove ν è l'omomorfismo naturale di M su $\overline{M}=M/\ker\eta$ e $\bar\eta$ è il monomorfismo indotto di $\overline{M}$ in M'. Se η è suriettivo, lo è anche $\bar\eta$, che è allora un isomorfismo. Ogni immagine omomorfa di M è dunque isomorfa a un modulo quoziente.

I risultati dei §§ 1.9 e 1.10 sugli omomorfismi di gruppi si estendono ai moduli. Lasciamo la verifica al lettore e ci serviremo liberamente dei corrispondenti risultati per i moduli quando ci occorreranno.

Gli analoghi dei gruppi ciclici sono i *moduli ciclici*. Un tale modulo è generato da un solo elemento $x\in M$, quindi ha la forma $M=Rx=\{ax\mid a\in R\}$. Il ruolo del gruppo ciclico infinito $(\mathbb{Z},+,0)$ è ora assunto da R come R-modulo, generato da 1 poiché $R=R1$. Se $M=Rx$, consideriamo $\mu_x: R\to Rx$, $a\mapsto ax$. È un omomorfismo di gruppi e $\mu_x(ba)=(ba)x$, mentre $b\mu_x(a)=b(ax)$. Dunque $\mu_x(ba)=b\mu_x(a)$, per cui μ_x è un omomorfismo di moduli. È suriettivo e quindi $M=Rx\cong R/\ker\mu_x$. Il nucleo $\ker\mu_x=\{d\in R\mid dx=0\}$, essendo un sottomodulo

di R, è un ideale sinistro di R. Lo chiameremo *annullatore di* x (in R), indicandolo con $\operatorname{ann} x$. Abbiamo dunque la formula

$$Rx \cong R/\operatorname{ann} x. \tag{8}$$

Se $\operatorname{ann} x = 0$, si ha $Rx \cong R$. Nel caso $R = \mathbb{Z}$, o $\mathbb{Z}x \cong \mathbb{Z}$ oppure $\operatorname{ann} x = (n)$, ove $n > 0$ è il più piccolo intero positivo per cui $nx = 0$. È precisamente l'ordine dell'elemento x e del gruppo ciclico $\langle x \rangle$. L'annullatore di un elemento x di un modulo generalizza così l'ordine di un elemento di un gruppo. Per questo $\operatorname{ann} x$ si dice talvolta *ideale d'ordine* di x.

Siano M, N moduli e sia $\operatorname{Hom}(M, N)$, o $\operatorname{Hom}_R(M, N)$, l'insieme degli omomorfismi di M in N. Lo si rende un gruppo abeliano definendo $(\eta + \zeta)(x) = \eta(x) + \zeta(x)$ e $0(x) = 0$, quest'ultimo essendo lo zero di N. Per verificare che $\eta + \zeta, 0 \in \operatorname{Hom}(M, N)$ e che si ha un gruppo abeliano, occorre un solo passo in più rispetto alla verifica additiva per gli endomorfismi di un gruppo abeliano (p. 160 dell'originale): si deve mostrare che $(\eta + \zeta)(ax) = a((\eta + \zeta)(x))$. Ma il primo membro è $\eta(ax) + \zeta(ax) = a\eta(x) + a\zeta(x)$ e il secondo è $a(\eta(x) + \zeta(x)) = a\eta(x) + a\zeta(x)$. Dato un terzo modulo P e dati $\eta \in \operatorname{Hom}(M, N)$ e $\zeta \in \operatorname{Hom}(N, P)$, la composta $\zeta\eta$ è un omomorfismo di gruppi additivi da M a P; poiché

$$(\zeta\eta)(ax) = \zeta(\eta(ax)) = \zeta(a\eta(x)) = a\zeta(\eta(x)) = a((\zeta\eta)(x)),$$

essa appartiene a $\operatorname{Hom}(M, P)$. Come nel caso particolare di $\operatorname{End} M$ valgono le leggi distributive $(\zeta_1 + \zeta_2)\eta = \zeta_1\eta + \zeta_2\eta$ e $\zeta(\eta_1 + \eta_2) = \zeta\eta_1 + \zeta\eta_2$ per $\eta, \eta_1, \eta_2 \in \operatorname{Hom}(M, N)$ e $\zeta, \zeta_1, \zeta_2 \in \operatorname{Hom}(N, P)$. È chiaro inoltre che $1_N\eta = \eta = \eta 1_M$, e che, dato un quarto modulo Q, si ha $(\omega\zeta)\eta = \omega(\zeta\eta)$ per $\eta \in \operatorname{Hom}(M, N)$, $\zeta \in \operatorname{Hom}(N, P)$ e $\omega \in \operatorname{Hom}(P, Q)$. Come caso particolare segue che $(\operatorname{Hom}(M, M), +, \cdot, 0, 1)$ è un anello. Lo indicheremo con $\operatorname{End}_R M$ e lo chiameremo *anello degli endomorfismi del modulo* M.

Esercizi

1. Determinare $\operatorname{Hom}(\mathbb{Z}, \mathbb{Z}/(n))$ e $\operatorname{Hom}(\mathbb{Z}/(n), \mathbb{Z})$ per $n > 0$, come $\mathbb{Z}$-moduli.
2. Determinare $\operatorname{Hom}(\mathbb{Z}/(m), \mathbb{Z}/(n))$ per $m, n > 0$, come $\mathbb{Z}$-moduli.
3. Dimostrare che $\operatorname{Hom}(\mathbb{Z}^{(2)}, \mathbb{Z}) \cong (\mathbb{Z}^{(2)}, +, 0)$.
4. Per ogni anello R e ogni R-modulo M, dimostrare che $\operatorname{Hom}(R, M) \cong (M, +, 0)$.
5. Dimostrare che $\operatorname{End}_R M$ è il centralizzante in $\operatorname{End} M$ dell'insieme degli endomorfismi di gruppi a_L, $a \in R$.
6. Vale $a_L \in \operatorname{End}_R M$?
7. Un modulo M si dice *irriducibile* se $M \neq 0$ e i suoi soli sottomoduli sono 0 e M. Dimostrare che M è irriducibile se e solo se $M \neq 0$ ed è ciclico con ogni elemento non nullo come generatore.
8. Un ideale sinistro (destro) I di R si dice *massimale* se $R \neq I$ e non esistono ideali sinistri (destri) I' tali che $R \supsetneq I' \supsetneq I$. Dimostrare che M è irriducibile se e solo se $M \cong R/I$ per un ideale sinistro massimale I di R.
9. (Lemma di Schur.) Dimostrare che, per moduli irriducibili M_1, M_2, ogni omomorfismo non nullo di M_1 in M_2 è un isomorfismo. Dedurne che, se M è irriducibile, $\operatorname{End}_R M$ è un anello con divisione.

3.4 Moduli liberi e matrici

Sia R un anello e sia $R^{(n)}$ l'insieme delle n-uple $(x_1, \dots, x_n)$ con $x_i \in R$. Generalizziamo la consueta costruzione dello spazio vettoriale $\mathbb{R}^{(n)}$ introducendo su $R^{(n)}$ una somma, uno zero e una moltiplicazione per elementi di R:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n), \tag{9}$$

$$0 = (0, 0, \dots, 0), \tag{10}$$

$$a(x_1, \dots, x_n) = (ax_1, \dots, ax_n). \tag{11}$$

È chiaro che $(R^{(n)}, +, 0)$ è un gruppo abeliano: è un caso particolare della costruzione del prodotto diretto di p. 35 dell'originale. Dalla (11) discendono immediatamente le condizioni 1–4 per i moduli. Dunque $R^{(n)}$ è un modulo su R. Se $n = 1$, $R^{(1)}$ coincide con R considerato nel modo consueto come R-modulo sinistro. Poniamo

$$e_i = (0, \dots, 0, \overset{i}{1}, 0, \dots, 0). \tag{12}$$

Allora $x_i e_i = (0, \dots, 0, x_i, 0, \dots, 0)$ e

$$x = (x_1, \dots, x_n) = \sum_{i=1}^{n} x_i e_i. \tag{13}$$

Quindi gli n elementi e_i generano $R^{(n)}$ come R-modulo. Inoltre, in virtù della (13), $\sum x_i e_i = 0$ implica $(x_1, \dots, x_n) = 0$, cioè $x_i = 0$ per ogni i. Equivalentemente, $\sum x_i e_i = \sum y_i e_i$ implica $x_i = y_i$ per $1 \le i \le n$. Un sistema di generatori con queste proprietà si chiama *base*. L'esistenza di una base di n elementi caratterizza $R^{(n)}$ a meno di isomorfismo. Per provarlo stabiliamo prima un'altra proprietà fondamentale di $R^{(n)}$: dati un R-modulo qualsiasi M e una n-upla ordinata $(u_1, \dots, u_n)$ di suoi elementi, esiste un unico omomorfismo $\mu : R^{(n)} \to M$ tale che $\mu(e_i) = u_i$ per $1 \le i \le n$. Basta porre

$$\mu : (x_1, \dots, x_n) = \sum x_i e_i \longmapsto \sum x_i u_i. \tag{14}$$

L'applicazione è univocamente definita e una verifica diretta mostra che è un omomorfismo di moduli. Inoltre $\mu(e_i) = u_i$ per ogni i; poiché un omomorfismo è determinato dall'azione sui generatori (analogo per i moduli del teorema 1.7, p. 60 dell'originale), μ è l'unico omomorfismo con le immagini prescritte.

Supponiamo che gli u_i costituiscano una base di M nel senso ora definito. Allora $\operatorname{im}\mu$, sottomodulo di M, contiene i generatori $u_1, \dots, u_n$, perciò coincide con M. Se $x = (x_1, \dots, x_n) \in \ker\mu$, si ha $\sum x_i u_i = 0$, donde per definizione di base $x_i = 0$ per ogni i e $x = 0$. Così $\ker\mu = 0$ e μ è un isomorfismo. Abbiamo dunque provato che l'esistenza di una base di n elementi per M implica $M \cong R^{(n)}$. In tal caso diremo che M è un *R-modulo libero di rango n*.

Può accadere che $R^{(m)} \cong R^{(n)}$ per interi distinti m, n. Costruire esempi di tali anelli è piuttosto difficile. Per molte classi importanti di anelli vale infatti il noto risultato dell'algebra lineare sull'invarianza del numero degli elementi di una base. Mostriamolo per gli anelli commutativi.

Teorema 3.4. *Se R è commutativo, da $R^{(m)} \cong R^{(n)}$ segue $m = n$.*

Dimostrazione. Per il risultato sui moduli liberi, l'enunciato equivale a questo: se un modulo M su un anello commutativo R ha basi di m e di n elementi, allora $m = n$. Siano dunque $\{e_i \mid 1 \le i \le n\}$ e $\{f_j \mid 1 \le j \le m\}$ due basi di M. Si ha

$$f_j = \sum_{i=1}^{n} a_{ji} e_i, \qquad e_i = \sum_{j=1}^{m} b_{ij} f_j \qquad (a_{ji}, b_{ij} \in R).$$

Sostituendo si ottiene

$$f_j = \sum_{i=1}^{n}\sum_{j'=1}^{m} a_{ji}b_{ij'}f_{j'}, \qquad e_i = \sum_{j=1}^{m}\sum_{i'=1}^{n} b_{ij}a_{ji'}e_{i'}.$$

Poiché gli f_j e gli e_i formano basi,

$$\sum_{i=1}^{n} a_{ji}b_{ij'} = \begin{cases} 1 & j = j', \\ 0 & j \neq j', \end{cases} \tag{15}$$

$$\sum_{j=1}^{m} b_{ij}a_{ji'} = \begin{cases} 1 & i = i', \\ 0 & i \neq i', \end{cases} \tag{16}$$

con $j, j' = 1, \ldots, m$ e $i, i' = 1, \ldots, n$. Supponiamo $m < n$ e consideriamo le due matrici quadrate di ordine n

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & \cdots & b_{1m} & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nm} & 0 & \cdots & 0 \end{pmatrix}.$$

La (16) equivale a $BA = 1$. Poiché R è commutativo, ne segue $AB = 1$ (teorema 2.1, p. 96, ed esercizio 2, p. 97 dell'originale). D'altra parte la forma di A e B mostra che le ultime $n-m$ righe di AB sono nulle: dunque $AB \neq 1$. La contraddizione prova $m \geq n$. Per simmetria $n \geq m$, sicché $m = n$. □

Il ragionamento precedente mostra che, per due basi $(e_1, \ldots, e_n)$ e $(f_1, \ldots, f_n)$ e per $f_j = \sum_{i=1}^{n} a_{ji}e_i$, $e_i = \sum_{j=1}^{n} b_{ij}f_j$, le matrici $A = (a_{ij})$ e $B = (b_{ij})$ soddisfano $AB = 1 = BA$. Sono dunque invertibili: $A, B \in \mathrm{GL}_n(R)$, il gruppo delle matrici invertibili $n \times n$ a elementi in R. Viceversa, data una base $(e_1, \ldots, e_n)$ e una matrice $A \in \mathrm{GL}_n(R)$, definiamo $f_j = \sum_{i=1}^{n} a_{ji}e_i$ per $1 \leq j \leq n$. Anche $(f_1, \ldots, f_n)$ è una base. Infatti, se $B = A^{-1}$, si ha $\sum_j b_{kj}f_j = \sum_{i,j} b_{kj}a_{ji}e_i = e_k$ poiché $BA = 1$. Siccome gli e_i generano M, lo generano anche gli f_j. Se poi $\sum_j d_j f_j = 0$, si ha $\sum_{i,j} d_j a_{ji} e_i = 0$, e quindi $\sum_{j=1}^{n} d_j a_{ji} = 0$ per ogni i. Ne segue $\sum_{i,j=1}^{n} d_j a_{ji} b_{ih} = 0$ per ogni h; poiché $AB = 1$, si ha $d_h = 0$ per ogni h. Dunque gli f_j sono una base. Fissata una base ordinata $(e_1, \ldots, e_n)$ di un modulo libero su un anello commutativo, se ne ottengono pertanto tutte le basi ordinate applicando le matrici $A = (a_{ij}) \in \mathrm{GL}_n(R)$ nel senso appena precisato, cioè $f_j = \sum_i a_{ji}e_i$.

Abbandoniamo ora l'ipotesi di commutatività di R e consideriamo il gruppo additivo $\mathrm{Hom}(R^{(m)}, R^{(n)})$ degli omomorfismi di moduli per m, n arbitrari. Scegliamo basi $(e_1, \ldots, e_m)$ e $(f_1, \ldots, f_n)$ di $R^{(m)}$ e $R^{(n)}$. Per $\eta \in \mathrm{Hom}(R^{(m)}, R^{(n)})$ scriviamo

$$\begin{aligned} \eta(e_1) &= a_{11}f_1 + a_{12}f_2 + \cdots + a_{1n}f_n, \\ \eta(e_2) &= a_{21}f_1 + a_{22}f_2 + \cdots + a_{2n}f_n, \\ &\vdots \\ \eta(e_m) &= a_{m1}f_1 + a_{m2}f_2 + \cdots + a_{mn}f_n. \end{aligned} \tag{17}$$

Chiamiamo la matrice $m \times n$, $A = (a_{ij})$, di m righe e n colonne, *matrice di η rispetto alle basi ordinate* $(e_1, \ldots, e_m)$, $(f_1, \ldots, f_n)$. La matrice determina η: infatti, se vale (17) e $x = (x_1, \ldots, x_m) = \sum x_i e_i$, allora

$$\eta(x) = \eta\left(\sum x_i e_i\right) = \sum_i x_i \eta(e_i) = \sum_{i,j} x_i a_{ij} f_j.$$

Pertanto η è l'applicazione

$$(x_1, \ldots, x_m) \longmapsto (y_1, \ldots, y_n), \tag{18}$$

dove

$$y_j = \sum_{i=1}^{m} x_i a_{ij}, \qquad j = 1, \ldots, n. \tag{19}$$

Possiamo esprimerla anche in forma matriciale. In generale, per matrici $m \times n$ $A = (a_{ij}), B = (b_{ij})$ definiamo la somma $A + B = (a_{ij} + b_{ij})$, con elemento di posto (i, j) pari ad $a_{ij} + b_{ij}$. Per $A = (a_{ij})$ di formato $m \times n$ e $B = (b_{jk})$ di formato $n \times q$, il prodotto $P = AB$ è la matrice $m \times q$ il cui elemento di posto (i, k), $1 \leq i \leq m$, $1 \leq k \leq q$, è

$$p_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}. \tag{20}$$

Per esempio,

$$\begin{pmatrix} 0 & -1 & 2 \\ 3 & 5 & -4 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 6 & 3 \\ -1 & 5 \end{pmatrix} = \begin{pmatrix} -8 & 7 \\ 40 & -5 \end{pmatrix}.$$

Con la definizione (20), le (18) e (19) si riscrivono

$$\eta : (x_1, \ldots, x_m) \longmapsto (y_1, \ldots, y_n) = (x_1, \ldots, x_m)A. \tag{21}$$

L'insieme $M_{m,n}(R)$ delle matrici $m \times n$ a elementi in R è un gruppo rispetto all'addizione $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$ e ha per zero la matrice nulla $m \times n$. Dimostreremo che questo gruppo è isomorfo a $\mathrm{Hom}(R^{(m)}, R^{(n)})$ mediante $\eta \mapsto A$, dove A è la matrice di η rispetto alle basi scelte. Per la (21), A determina η, quindi l'applicazione è iniettiva. È anche suriettiva, perché, data $A \in M_{m,n}(R)$, possiamo porre $v_i = \sum_{j=1}^{n} a_{ij} f_j$. Esiste allora un η tale che $\eta(e_i) = v_i$ per $1 \leq i \leq m$, e la sua matrice è proprio A. Dunque $\eta \mapsto A$ è biiettiva. Se $\zeta \mapsto B = (b_{ij})$, si ha $\zeta(e_i) = \sum_j b_{ij} f_j$, onde

$$(\eta + \zeta)(e_i) = \sum_j a_{ij} f_j + \sum_j b_{ij} f_j = \sum_j (a_{ij} + b_{ij}) f_j.$$

Pertanto $\eta + \zeta \mapsto A + B$: la corrispondenza è un isomorfismo di gruppi.

Sia ora $\rho \in \mathrm{Hom}(R^{(n)}, R^{(q)})$ e sia $(g_1, \ldots, g_q)$ una base di $R^{(q)}$. La matrice $C = (c_{jk})$ di ρ rispetto alle basi (f_j) e (g_k) è data da $\rho(f_j) = \sum_{k=1}^{q} c_{jk} g_k$. Se η ha matrice $A = (a_{ij})$ rispetto alle basi $(e_i), (f_j)$, allora $\rho\eta \in \mathrm{Hom}(R^{(m)}, R^{(q)})$ e

$$(\rho\eta)(e_i) = \rho\left(\sum_j a_{ij} f_j\right) = \sum_j a_{ij} \rho(f_j) = \sum_{j,k} a_{ij} c_{jk} g_k.$$

Quindi la matrice di $\rho\eta$ rispetto a $(e_i), (g_k)$ è AC. Possiamo dedurne l'associatività del prodotto di matrici rettangolari, che naturalmente si può provare anche direttamente come per le matrici quadrate (p. 94 dell'originale). Introduciamo un quarto modulo libero $R^{(s)}$ con base (h_l) e sia $\tau \in \mathrm{Hom}(R^{(q)}, R^{(s)})$. Si ha $\tau(\rho\eta) = (\tau\rho)\eta \in \mathrm{Hom}(R^{(m)}, R^{(s)})$. Indichiamo con un asterisco la matrice di ciascun omomorfismo rispetto alle basi scelte, per esempio $\eta^* = A$, $\rho^* = C$. Allora $(\rho\eta)^* = \eta^*\rho^*$, onde

$$\eta^*(\rho^*\tau^*) = \eta^*(\tau\rho)^* = ((\tau\rho)\eta)^* = (\tau(\rho\eta))^* = (\rho\eta)^*\tau^* = (\eta^*\rho^*)\tau^*.$$

Le matrici η^*, ρ^*, τ^* possono essere arbitrarie dei formati $m \times n$, $n \times q$, $q \times s$, quindi l'associatività è provata in generale. Nello stesso modo si dimostrano le leggi distributive: per $A, A_1, A_2 \in M_{m,n}(R)$ e $C, C_1, C_2 \in M_{n,q}(R)$ valgono $(A_1 + A_2)C = A_1C + A_2C$ e $A(C_1 + C_2) = AC_1 + AC_2$.

Nel caso $\operatorname{End}_R R^{(n)} = \operatorname{Hom}(R^{(n)}, R^{(n)})$ il risultato dà un antiisomorfismo $\eta \mapsto \eta^* = A$ di $\operatorname{End}_R R^{(n)}$ con l'anello $M_n(R) = M_{n,n}(R)$. Qui A è la matrice di η nella base (e_i): se $\eta(e_i) = \sum_j a_{ij} e_j$, allora $A = (a_{ij})$. Se R è commutativo, la trasposizione $A \mapsto {}^tA$ è un antiautomorfismo di $M_n(R)$ (p. 111 dell'originale). Componendolo con $\eta \mapsto A$ si ottiene un isomorfismo $\eta \mapsto {}^tA$ di $\operatorname{End}_R R^{(n)}$ con $M_n(R)$. È questa la convenzione usata nell'esempio di $\mathbb{Z}^{(2)}$ di p. 161 dell'originale.

Nel caso particolare delle matrici associate alle applicazioni lineari fra spazi vettoriali queste considerazioni dovrebbero essere familiari. La discussione illustra il principio generale secondo cui in molte situazioni il passaggio dagli spazi vettoriali ai moduli liberi è del tutto naturale.

Esercizi

1. Sia R arbitrario e sia $(e_1, \dots, e_n)$ una base di $R^{(n)}$. Dimostrare che $(f_1, \dots, f_m)$, con $f_j = \sum_{j'=1}^{n} a_{jj'} e_{j'}$, è una base di $R^{(n)}$ se e solo se esiste una matrice B di formato $n \times m$ tale che $AB = 1_m$ e $BA = 1_n$, dove $A = (a_{ij})$ e $1_m, 1_n$ sono le matrici unità dei rispettivi ordini. Dedurne che $R^{(m)} \cong R^{(n)}$ se e solo se esistono $A \in M_{m,n}(R)$, $B \in M_{n,m}(R)$ con $AB = 1_m$, $BA = 1_n$.

2. Sia $\eta \in \operatorname{End}_R(R^{(n)})$ e sia A la sua matrice nella base $(e_1, \dots, e_n)$. Se $f_i = \sum_j p_{ij} e_j$, con $P = (p_{ij}) \in \mathrm{GL}_n(R)$, verificare che la matrice di η nella base $(f_1, \dots, f_n)$ è PAP^{-1}.

3. Sia R_n un R-modulo destro libero con base $(e_1, \dots, e_n)$. Per $\eta \in \operatorname{End}_R R_n$ si scriva $\eta(e_i) = \sum_{j=1}^{n} e_j a_{ji}$. Dimostrare che $\eta \mapsto A = (a_{ij})$ è un isomorfismo di $\operatorname{End}_R R_n$ con $M_n(R)$.

4. Sia R commutativo. Dimostrare che un endomorfismo suriettivo η di $R^{(n)}$ è biiettivo. La stessa conclusione vale se η è iniettivo?

5. Sia R commutativo e siano M, N due R-moduli. Per $a \in R$ e $\eta \in \operatorname{Hom}(M, N)$ si definisca $a\eta$ mediante $(a\eta)(x) = a(\eta(x)) = \eta(ax)$. Mostrare che $a\eta \in \operatorname{Hom}(M, N)$ e che questa azione rende $\operatorname{Hom}(M, N)$ un R-modulo. Dimostrare che $\operatorname{Hom}(R^{(m)}, R^{(n)})$ è libero di rango nm.

6. Per R commutativo e $(e_1, \dots, e_n)$ base di $R^{(n)}$, si ponga $f_i = \sum_j a_{ij} e_j$ con $A = (a_{ij}) \in M_n(R)$. Mostrare che gli f_i sono una base di un sottomodulo libero K di $R^{(n)}$ se e solo se $\det A$ non è un divisore dello zero. Mostrare che per ogni $\bar{x} = x + K \in R^{(n)}/K$ si ha $(\det A)\bar{x} = 0$. (*Suggerimento.* Basta provare che $(\det A)\bar{e}_i = 0$ per $1 \le i \le n$.)

3.5 Somme dirette di moduli

Definiamo ora l'analogo, per i moduli, del prodotto diretto di monoidi o di gruppi (p. 35 dell'originale). Siano $M_1, \dots, M_n$ moduli sul medesimo anello R e sia M il prodotto cartesiano $M_1 \times \cdots \times M_n$ delle n-uple $(x_1, \dots, x_n)$, $x_i \in M_i$. Come nel caso particolare del modulo libero $R^{(n)}$, vi introduciamo una somma, uno zero e un prodotto per gli elementi di R ponendo

$$\begin{aligned} (x_1, \dots, x_n) + (y_1, \dots, y_n) &= (x_1 + y_1, \dots, x_n + y_n), \\ 0 &= (0, \dots, 0), \\ a(x_1, \dots, x_n) &= (ax_1, \dots, ax_n), \qquad a \in R. \end{aligned}$$

Queste operazioni definiscono su M una struttura di modulo, detta *somma diretta* degli M_i e indicata con $M_1 \oplus M_2 \oplus \cdots \oplus M_n$ oppure con $\bigoplus_{i=1}^{n} M_i$.

Una proprietà fondamentale della somma diretta, relativa agli omomorfismi, è la seguente. Dati omomorfismi $\eta_i : M_i \to N$, $1 \le i \le n$, definiamo l'applicazione η da $\bigoplus M_i$ a N mediante

$$\eta : (x_1, \dots, x_n) \longmapsto \sum_{i=1}^{n} \eta_i(x_i).$$

Poiché

$$\begin{aligned}\eta(x_1 + y_1, \dots, x_n + y_n) &= \sum_i \eta_i(x_i + y_i) = \sum_i \eta_i(x_i) + \sum_i \eta_i(y_i) \\ &= \eta(x_1, \dots, x_n) + \eta(y_1, \dots, y_n)\end{aligned}$$

e $\eta(ax_1, \dots, ax_n) = \sum_i \eta_i(ax_i) = \sum_i a\eta_i(x_i) = a\sum_i \eta_i(x_i)$, η è un omomorfismo. Lo impiegheremo nella dimostrazione della prima parte del seguente teorema, che caratterizza la somma diretta attraverso proprietà interne.

Teorema 3.5. *Siano $M_1, \dots, M_n$ sottomoduli del modulo M con le proprietà seguenti:*

(i) *$M = M_1 + \cdots + M_n$ (cioè gli M_i generano M);*

(ii) *per ogni i, $M_i \cap (M_1 + \cdots + M_{i-1} + M_{i+1} + \cdots + M_n) = 0$.*

Allora $\iota : (x_1, \dots, x_n) \mapsto \sum_{i=1}^n x_i$ è un isomorfismo di $\bigoplus M_i$ su M. Viceversa, posti in $\bigoplus M_i$

$$M_i' = \{(0, \dots, 0, x_i, 0, \dots, 0) \mid x_i \in M_i\},$$

gli M_i' sono sottomoduli di $\bigoplus M_i$, isomorfi rispettivamente agli M_i, e soddisfano (i) e (ii).

Dimostrazione. Supponiamo che gli M_i soddisfino (i) e (ii). L'applicazione $\iota : (x_1, \dots, x_n) \mapsto \sum_i x_i$ è precisamente l'applicazione η appena costruita prendendo le identità $x_i \mapsto x_i$ da M_i a M_i e considerando le immagini in M. È dunque un omomorfismo. È suriettivo, perché la (i) dice che ogni $x \in M$ si può scrivere come $x = \sum_i x_i$, con $x_i \in M_i$, e allora $\iota(x_1, \dots, x_n) = x$. Per provare che è iniettivo basta mostrare che il suo nucleo è zero: da $\sum_i x_i = 0$ deve seguire $x_i = 0$ per ogni i. Infatti $-x_i = \sum_{j \ne i} x_j$ dà $x_i \in M_i \cap \sum_{j \ne i} M_j = 0$ in virtù della (ii). Dunque ι è un isomorfismo. Viceversa, nella somma diretta $\bigoplus M_i$ l'applicazione $\iota_i : x_i \mapsto (0, \dots, 0, x_i, 0, \dots, 0)$ è un monomorfismo di M_i. La sua immagine M_i' è perciò un sottomodulo isomorfo a M_i. Poiché

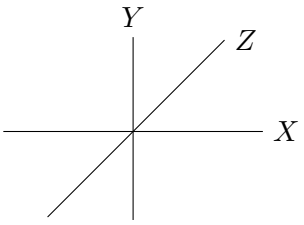
$$(x_1, 0, \dots, 0) + (0, x_2, 0, \dots, 0) + \cdots + (0, \dots, 0, x_n) = (x_1, \dots, x_n),$$

vale la (i). La somma $\sum_{j \ne i} M_j'$ consiste delle n-uple $(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n)$, quindi vale anche la (ii). La dimostrazione è completa. □

Il teorema permette di identificare M con $\bigoplus M_i$ quando gli M_i sono sottomoduli di M che soddisfano (i) e (ii). Diremo allora che M è la *somma diretta interna* dei suoi sottomoduli M_i e scriveremo ugualmente $M = \bigoplus M_i$ o $M = M_1 \oplus \cdots \oplus M_n$.

I sottomoduli M_i, $1 \le i \le n$, che soddisfano la (ii) si dicono *indipendenti*. Ciò equivale a dire che da ogni relazione $\sum_i x_i = 0$, con $x_i \in M_i$, segue $x_i = 0$ per ogni i; o ancora che da $\sum_i x_i = \sum_i y_i$, con $x_i, y_i \in M_i$, segue $x_i = y_i$ per ogni i. L'indipendenza è condizione più forte di $M_i \cap M_j = 0$ per $i \ne j$, ed è più forte persino delle condizioni $M_i \cap (\bigcup_{j \ne i} M_j) = 0$. Per esempio, nello spazio vettoriale reale bidimensionale $\mathbb{R}^{(2)}$ consideriamo

$$X = \{(x, 0) \mid x \in \mathbb{R}\}, \quad Y = \{(0, y) \mid y \in \mathbb{R}\}, \quad Z = \{(z, z) \mid z \in \mathbb{R}\}.$$



In figura, X è l'insieme dei vettori con estremità sull'asse x, Y quello dei vettori con estremità sull'asse y, Z quello dei vettori con estremità sulla retta inclinata di 45°. L'intersezione di ciascuna retta con l'unione delle altre due è l'origine. Tuttavia $X+Y = \mathbb{R}^{(2)}$, quindi $(X+Y)\cap Z = Z$. Dunque X, Y, Z non sono indipendenti.

I criteri di indipendenza formulati mediante gli elementi hanno le seguenti conseguenze.

I. Siano $M_1, \dots, M_n$ sottomoduli indipendenti. Posto $N_1 = M_1 + \cdots + M_{r_1}$, $N_2 = M_{r_1+1} + \cdots + M_{r_1+r_2}$, $N_3 = M_{r_1+r_2+1} + \cdots + M_{r_1+r_2+r_3}$, e così via, i sottomoduli $N_1, N_2, \dots$ sono indipendenti.

II. Siano $M_1, \dots, M_n$ indipendenti e sia $M_i = M_{i1} \oplus \cdots \oplus M_{ir_i}$ per $1 \le i \le n$, dove gli M_{ij} sono sottomoduli di M_i. Allora $M_{11}, \dots, M_{1r_1}, M_{21}, \dots, M_{2r_2}, \dots, M_{n1}, \dots, M_{nr_n}$ sono indipendenti.

Lasciamo la prova al lettore. Ne segue immediatamente il teorema seguente.

Teorema 3.6. *Sia $M = \bigoplus M_i$ la somma diretta dei suoi sottomoduli M_i. Posto $N_1 = M_1 + \cdots + M_{r_1}$, $N_2 = M_{r_1+1} + \cdots + M_{r_1+r_2}$ e così via, si ha $M = \bigoplus N_j$. Inoltre, se $M_i = \bigoplus_j M_{ij}$, $1 \le i \le n$, $1 \le j \le r_i$, allora $M = \bigoplus_{i,j} M_{ij}$.*

Omettiamo anche questa dimostrazione.

Esercizi

1. Sia V uno spazio vettoriale su un campo F. Dimostrare che i vettori non nulli x_i, $1 \le i \le n$, sono linearmente indipendenti se e solo se i sottospazi Fx_i sono indipendenti. Dimostrare inoltre che gli x_i formano una base se e solo se $V = \bigoplus Fx_i$.

2. Siano M_i, $1 \le i \le n$, sottomoduli di M tali che $M = \sum M_i$, e valgano le condizioni «triangolari» $M_1 \cap M_2 = 0$, $(M_1 + M_2) \cap M_3 = 0$, …, $(M_1 + \cdots + M_{n-1}) \cap M_n = 0$. Dimostrare che $M = \bigoplus M_i$.

3. Dimostrare che $\mathbb{Z}/(p^e)$, con p primo ed $e > 0$, considerato come $\mathbb{Z}$-modulo, non è somma diretta di due sottomoduli non nulli. Vale lo stesso per $\mathbb{Z}$? E per $\mathbb{Z}/(n)$ con altri interi positivi n?

4. Dimostrare che, se $M = M_1 \oplus M_2$, allora $M_1 \cong M/M_2$ e $M_2 \cong M/M_1$.

5. Siano M, N due R-moduli e siano $f : M \to N$, $g : N \to M$ omomorfismi di R-moduli con $fg(y) = y$ per ogni $y \in N$. Dimostrare che $M = \ker f \oplus \operatorname{im} g$.

3.6 Moduli finitamente generati su un dominio a ideali principali. Risultati preliminari

Possiamo ora affrontare l'obiettivo principale del capitolo: lo studio dei moduli finitamente generati su un dominio a ideali principali e le sue applicazioni ai gruppi abeliani finiti e alle

trasformazioni lineari. Sia M un modulo su un dominio a ideali principali D, generato da un insieme finito di elementi $x_1, \dots, x_n$, cosicché $M = \sum_{i=1}^{n} Dx_i$. Per studiarlo è naturale introdurre il modulo libero $D^{(n)}$ con base $(e_1, \dots, e_n)$ e l'epimorfismo $\eta : \sum_i a_i e_i \mapsto \sum_i a_i x_i$, $a_i \in D$, da $D^{(n)}$ su M. Allora $M \cong D^{(n)}/K$ con $K = \ker \eta$. Un primo fatto che ci occorrerà è che K è finitamente generato. Lo dedurremo dal risultato più forte che segue.

Teorema 3.7. *Sia D un dominio a ideali principali e sia $D^{(n)}$ il modulo libero di rango n su D. Ogni sottomodulo K di $D^{(n)}$ è libero e ha una base di $m \leq n$ elementi.*

Dimostrazione. Non escludendo $K = 0$, adottiamo la convenzione che il modulo costituito dal solo zero sia «libero di rango zero», con base vuota. Il risultato è allora banale per $n = 0$. Supponiamo $n > 0$ e ammettiamo il risultato per i sottomoduli di un modulo libero su D con una base di $n - 1$ elementi. Indichiamo con $D^{(n-1)}$ il sottomodulo generato da $e_2, \dots, e_n$. È libero con base $(e_2, \dots, e_n)$, quindi, se $K \subset D^{(n-1)}$, il risultato è acquisito. Possiamo supporre $K \not\subset D^{(n-1)}$. Consideriamo il sottoinsieme I di D formato dai b per i quali esiste un elemento $be_1 + y \in K$ con $y \in D^{(n-1)}$. Si tratta di un ideale di D, non nullo poiché $K \not\subset D^{(n-1)}$. Dunque $I = (d)$ con $d \neq 0$ ed esiste $f_1 = de_1 + y_1 \in K$, $y_1 \in D^{(n-1)}$. Sia ora $L = K \cap D^{(n-1)}$. È un sottomodulo di $D^{(n-1)}$ e per induzione ha una base $(f_2, \dots, f_m)$ di $m - 1 \leq n - 1$ elementi (può essere $m - 1 = 0$).

Mostriamo che $(f_1, f_2, \dots, f_m)$ è una base di K. Sia $x \in K$: si ha $x = be_1 + y$, con $b \in I = (d)$ e $y \in D^{(n-1)}$. Scrivendo $b = k_1 d$ si ottiene

$$x - k_1 f_1 = k_1 de_1 + y - k_1(de_1 + y_1) = y - k_1 y_1 \in L.$$

Pertanto $x - k_1 f_1 = \sum_{j=2}^{m} k_j f_j$ con $k_j \in D$, e $x = \sum_{j=1}^{m} k_j f_j$. Dunque gli f_j generano K. Se poi $\sum_{i=1}^{m} k_i f_i = 0$, allora $k_1 de_1 + k_1 y_1 + \sum_{j=2}^{m} k_j f_j = 0$. Poiché y_1 e gli f_j per $j \geq 2$ appartengono a $D^{(n-1)}$, ne risulta una relazione $k_1 de_1 + \sum_{k=2}^{n} l_k e_k = 0$, $l_k \in D$. Si ha così $k_1 d = 0$, quindi $k_1 = 0$ essendo $d \neq 0$. Ora $\sum_{j=2}^{m} k_j f_j = 0$ e, poiché gli $f_2, \dots, f_m$ sono una base di L, ogni $k_j = 0$. Dunque $(f_1, \dots, f_m)$ è una base di K. □

Ogni campo F è un dominio a ideali principali (i suoi soli ideali sono (0) e (1)). Per $D = F$ il teorema si riduce a un ben noto risultato di algebra lineare. Se V è uno spazio vettoriale di dimensione n su F, ossia un F-modulo libero di rango n, ogni suo sottospazio ha dimensione finita $m \leq n$.

Torniamo a $M \cong D^{(n)}/K$. Il teorema 3.7 ci dice che K ha una base di $m \leq n$ elementi. Il metodo che useremo funziona altrettanto bene partendo da un sistema finito di generatori e, nella pratica, è talvolta utile poter fare a meno di una base. Supponiamo dunque di avere generatori $f_1, \dots, f_m$ del sottomodulo K, dove m può superare n. Esprimiamoli mediante la base $(e_1, \dots, e_n)$:

$$\begin{aligned} f_1 &= a_{11}e_1 + a_{12}e_2 + \cdots + a_{1n}e_n, \\ f_2 &= a_{21}e_1 + a_{22}e_2 + \cdots + a_{2n}e_n, \\ &\vdots \\ f_m &= a_{m1}e_1 + a_{m2}e_2 + \cdots + a_{mn}e_n. \end{aligned} \tag{22}$$

La matrice $m \times n$ $A = (a_{ki})$ di queste relazioni è detta *matrice delle relazioni* del sistema ordinato di generatori $(f_1, \dots, f_m)$ rispetto alla base ordinata $(e_1, \dots, e_n)$. Non vi è naturalmente nulla di speciale nella scelta della base di $D^{(n)}$ e dei generatori di K. Vediamo che cosa accade cambiandoli. Ogni altra base di $D^{(n)}$ ha la forma $(e'_1, \dots, e'_n)$, con $e'_i = \sum_{j=1}^{n} p_{ij} e_j$, dove $P = (p_{ij})$ è invertibile in $M_n(D)$. Per i sistemi di generatori del sottomodulo K non si può fare

un'affermazione altrettanto generale. È tuttavia chiaro che, se $Q = (q_{kl})$ è invertibile in $M_m(D)$ e $Q^{-1} = (q^*_{kl})$, gli elementi $f'_k = \sum_{l=1}^m q_{kl} f_l$ generano ancora K. Infatti $f'_k \in K$ e $\sum_k q^*_{rk} f'_k = \sum_{k,l} q^*_{rk} q_{kl} f_l = f_r$, sicché ogni f_r appartiene al sottomodulo generato dagli f'_k.

Qual è la matrice delle relazioni degli f'_k rispetto agli e'_i? Si ha

$$f'_k = \sum_l q_{kl} f_l = \sum_{l,j} q_{kl} a_{lj} e_j = \sum_{l,j,i} q_{kl} a_{lj} p^*_{ji} e'_i, \qquad (p^*_{ij}) = P^{-1}.$$

La nuova matrice delle relazioni è dunque $A' = QAP^{-1}$.

Siamo così condotti a scegliere opportunamente Q e P in modo che le relazioni assumano una semplice forma «normale», dalla quale si possano ricavare informazioni importanti su $M \cong D^{(n)}/K$. Il problema matriciale ha interesse anche di per sé: nel prossimo paragrafo lo tratteremo a parte, prima di tornare allo studio del quoziente.

Esercizi

1. Trovare una base del sottomodulo di $\mathbb{Z}^{(3)}$ generato da $f_1 = (1, 0, -1)$, $f_2 = (2, -3, 1)$, $f_3 = (0, 3, 1)$, $f_4 = (3, 1, 5)$.
2. Trovare una base del sottomodulo di $\mathbb{Q}[\lambda]^{(3)}$ generato da $(2\lambda - 1, \lambda, \lambda^2 + 3)$, $(\lambda, \lambda, \lambda^2)$, $(\lambda + 1, 2\lambda, 2\lambda^2 - 3)$.
3. Trovare una base dello $\mathbb{Z}$-sottomodulo di $\mathbb{Z}^{(3)}$ costituito dalle terne (x_1, x_2, x_3) tali che $x_1 + 2x_2 + 3x_3 = 0$ e $x_1 + 4x_2 + 9x_3 = 0$.

3.7 Equivalenza delle matrici a coefficienti in un dominio a ideali principali

Due matrici A, B di tipo $m \times n$ su un dominio a ideali principali D si dicono *equivalenti* se esistono matrici invertibili $P \in M_m(D)$ e $Q \in M_n(D)$ tali che $B = PAQ$. Ciò definisce evidentemente una relazione di equivalenza in $M_{m,n}(D)$. Fra le matrici equivalenti a una matrice data A ne cerchiamo una la cui forma normale sia particolarmente semplice. Dimostreremo il seguente risultato.

Teorema 3.8. *Se $A \in M_{m,n}(D)$ e D è un dominio a ideali principali, A è equivalente a una matrice della forma «diagonale»*

$$\operatorname{diag}\{d_1, d_2, \dots, d_r, 0, \dots, 0\}, \qquad d_i \neq 0, \quad d_i \mid d_j \text{ se } i \leq j. \tag{23}$$

Qui la matrice, anche rettangolare, ha le entrate indicate sulla diagonale principale e zeri altrove.

Otterremo le matrici P, Q che trasformano A nella forma (23) come prodotti di matrici di certi tipi particolari che ora definiremo. Senza specificare l'ordine (m oppure n), introduciamo alcune matrici quadrate invertibili su D, che diremo *elementari*, e consideriamo l'effetto della moltiplicazione a sinistra o a destra per queste matrici.

Siano $b \in D$ e $i \neq j$. Poniamo $T_{ij}(b) = 1 + be_{ij}$, dove e_{ij} ha un unico 1 al posto (i, j) e zeri altrove. Questa matrice è invertibile, poiché

$$T_{ij}(b) T_{ij}(-b) = (1 + be_{ij})(1 - be_{ij}) = 1.$$

Sia poi u un elemento invertibile di D e poniamo $D_i(u) = 1 + (u - 1)e_{ii}$. Questa è una matrice diagonale con u al posto i e 1 in tutti gli altri posti diagonali. È invertibile, con $D_i(u)^{-1} =$

$D_i(u^{-1})$. Poniamo infine $P_{ij} = 1 - e_{ii} - e_{jj} + e_{ij} + e_{ji}$. Anche questa matrice è invertibile, perché $P_{ij}^2 = 1$. È facile verificare quanto segue.

I. Moltiplicando A a sinistra per la matrice $m \times m$ $T_{ij}(b)$, si aggiunge alla riga i di A la riga j moltiplicata per b, lasciando inalterate tutte le altre righe. Moltiplicando A a destra per la matrice $n \times n$ $T_{ij}(b)$, si aggiunge alla colonna j la colonna i moltiplicata per b, lasciando inalterate tutte le altre colonne.

II. Moltiplicare A a sinistra per la matrice $m \times m$ $D_i(u)$ equivale a moltiplicare la riga i per u, senza alterare le altre. Moltiplicarla a destra per la matrice $n \times n$ $D_i(u)$ equivale a moltiplicare la colonna i per u, senza alterare le altre.

III. Moltiplicare A a sinistra per la matrice $m \times m$ P_{ij} equivale a scambiare le righe i e j, lasciando inalterate le altre. Moltiplicarla a destra per la matrice $n \times n$ P_{ij} equivale a scambiare le colonne i e j, lasciando inalterate le altre.

Chiameremo $T_{ij}(b), D_i(u), P_{ij}$ *matrici elementari* rispettivamente del I, II e III tipo. La moltiplicazione di A a sinistra (a destra) per una di esse si dirà *trasformazione elementare sulle righe (sulle colonne)* del tipo corrispondente. Tali trasformazioni danno matrici equivalenti ad A.

Dimostrazione del Teorema 3.8. Trattiamo dapprima il caso particolare in cui D è euclideo, con applicazione δ da D a $\mathbb{N}$. Se $A = 0$ non c'è nulla da dimostrare. Altrimenti scegliamo un elemento $a_{ij} \neq 0$ di A con $\delta(a_{ij})$ minimo. Trasformazioni elementari sulle righe e sulle colonne lo portano al posto $(1,1)$. Supponiamo che vi si trovi già. Per $k > 1$ scriviamo $a_{1k} = a_{11}b_k + b_{1k}$, con resto nullo o con $\delta(b_{1k}) < \delta(a_{11})$. Sottraiamo dalla colonna k la prima colonna moltiplicata per b_k. Questa trasformazione elementare sostituisce a_{1k} con b_{1k}. Se $b_{1k} \neq 0$, otteniamo una matrice equivalente ad A per la quale il minimo di δ sugli elementi non nulli è diminuito. Riprendiamo il procedimento con questa matrice. Analogamente, se $a_{k1} = a_{11}b_k + b_{k1}$, con $b_{k1} \neq 0$ e $\delta(b_{k1}) < \delta(a_{11})$, mediante una trasformazione elementare del I tipo sulle righe diminuiamo il minimo di δ sugli elementi non nulli. Poiché il «grado» δ è un intero non negativo, dopo un numero finito di operazioni otteniamo una matrice equivalente $B = (b_{ij})$ nella quale $b_{11} \mid b_{1k}$ e $b_{11} \mid b_{k1}$ per ogni k. Trasformazioni elementari del I tipo sulle righe e sulle colonne danno allora una matrice equivalente della forma

$$\begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & c_{m2} & \cdots & c_{mn} \end{pmatrix}. \tag{24}$$

Possiamo ottenere anche $b_{11} \mid c_{kl}$ per ogni k, l. Infatti, se $b_{11} \nmid c_{kl}$, sommiamo la riga k alla prima: la nuova prima riga è $(b_{11}, c_{k2}, \ldots, c_{kl}, \ldots, c_{kn})$. Ripetendo il procedimento iniziale, sostituiamo c_{kl} con un elemento non nullo il cui grado è minore di quello di b_{11}. Un numero finito di passi del tipo indicato dà una matrice (24) equivalente ad A, nella quale $b_{11} \neq 0$ e $b_{11} \mid c_{kl}$ per ogni k, l. Ripetiamo il procedimento sulla sottomatrice (c_{kl}). Si ottiene una matrice equivalente della forma

$$\begin{pmatrix} b_{11} & 0 & 0 & \cdots & 0 \\ 0 & c_{22} & 0 & \cdots & 0 \\ 0 & 0 & d_{33} & \cdots & d_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & d_{m3} & \cdots & d_{mn} \end{pmatrix}, \qquad c_{22} \mid d_{pq} \quad \text{per ogni } p, q. \tag{25}$$

Le trasformazioni elementari sulle righe e sulle colonne di (c_{kl}) che danno (25) preservano la divisibilità per b_{11}. Quindi $b_{11} \mid c_{22}$ e $b_{11} \mid d_{pq}$. Proseguendo otteniamo la matrice diagonale equivalente $\operatorname{diag}\{d_1, d_2, \ldots, d_r, 0, \ldots, 0\}$ con $d_i \mid d_j$ per $i \le j$ ($d_1 = b_{11}$, $d_2 = c_{22}$, ecc.).

Nel caso generale l'argomento è assai simile. In luogo di $\delta(a)$ ragioniamo per induzione sulla lunghezza di un elemento non nullo di D. La lunghezza $\ell(a)$ di $a \neq 0$ è il numero dei fattori primi in una fattorizzazione $a = p_1 p_2 \cdots p_r$, con p_i primi. Per convenzione $\ell(u) = 0$ se u è un'unità. Oltre alle trasformazioni elementari sufficienti nel caso euclideo, avremo bisogno di moltiplicazioni per matrici della forma

$$\begin{pmatrix} x & s & \\ y & t & \\ & & I \end{pmatrix}, \qquad \begin{pmatrix} x & s \\ y & t \end{pmatrix} \text{ invertibile,} \tag{26}$$

dove i posti vuoti fuori dai blocchi diagonali sono nulli. Come prima, possiamo supporre $a_{11} \neq 0$ e $\ell(a_{11}) \le \ell(a_{ij})$ per ogni $a_{ij} \neq 0$. Supponiamo $a_{11} \nmid a_{1k}$. Scambiando le colonne 2 e k otteniamo $a_{11} \nmid a_{12}$. Scriviamo $a = a_{11}$, $b = a_{12}$ e $d = (a, b)$, cosicché $\ell(d) < \ell(a)$. Esistono $x, y \in D$ tali che $ax + by = d$. Poniamo $s = bd^{-1}$, $t = -ad^{-1}$. L'identità

$$\begin{pmatrix} -t & s \\ y & -x \end{pmatrix} \begin{pmatrix} x & s \\ y & t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

implica che entrambe le matrici sono invertibili, essendo D commutativo. È dunque invertibile anche (26). Moltiplicando A a destra per (26) otteniamo una matrice con prima riga $(d, 0, a_{13}, \ldots, a_{1n})$, e $\ell(d) < \ell(a_{11})$. Analogamente, se $a_{11} \nmid a_{k1}$ per qualche k, operazioni elementari seguite dalla moltiplicazione a sinistra per un'opportuna matrice (26) danno una matrice equivalente nella quale qualche elemento non nullo ha lunghezza minore di $\ell(a_{11})$. Possiamo così ottenere $a_{11} \mid a_{1k}$ e $a_{11} \mid a_{k1}$ per ogni k. Trasformazioni elementari danno allora una matrice della forma (24). Il resto dell'argomento coincide nella sostanza con il caso euclideo: a ogni passo diminuisce la lunghezza anziché il grado δ. □

Una matrice equivalente ad A che abbia la forma diagonale del Teorema 3.8 si chiama *forma normale* di A. I suoi elementi diagonali si chiamano *fattori invarianti* di A. Evidentemente ciascuno di essi si può sostituire con un associato, ossia con il suo prodotto per un'unità. Mostreremo che questa è l'unica libertà di scelta, ricavando formule che esprimono i fattori invarianti mediante gli elementi di A. Ricordiamo che A ha *rango (determinantale)* r se possiede un minore non nullo di ordine r e tutti i minori di ordine $r + 1$ sono nulli. Poiché i minori di ordine i sono somme di prodotti di minori di ordine $i - 1$ per elementi di D, quando il rango è r esistono minori non nulli di ogni ordine i, $1 \le i \le r$. Abbiamo quindi il seguente risultato.

Teorema 3.9. *Sia A una matrice $m \times n$ a elementi in un dominio a ideali principali D, di rango r. Per ogni $i \le r$ sia Δ_i un massimo comun divisore dei minori di ordine i di A. Allora i fattori invarianti non nulli di A coincidono, a meno di fattori invertibili, con*

$$d_1 = \Delta_1, \quad d_2 = \Delta_2 \Delta_1^{-1}, \quad \ldots, \quad d_r = \Delta_r \Delta_{r-1}^{-1}. \tag{27}$$

Si osservi che $\Delta_i \neq 0$ e $\Delta_{i-1} \mid \Delta_i$ per gli indici considerati.

Dimostrazione. Sia $Q = (q_{kl})$ una matrice $m \times m$ su D. L'elemento (k, i) di QA è $\sum_j q_{kj} a_{ji}$. Le righe di QA sono dunque combinazioni lineari delle righe di A a coefficienti in D. Per la multilinearità del determinante, i minori di ordine i di QA sono combinazioni lineari dei minori di ordine i di A. Il massimo comun divisore dei secondi divide pertanto il massimo comun

divisore dei primi. Analogamente, poiché le colonne di AP, $P \in M_n(D)$, sono combinazioni lineari delle colonne di A, il massimo comun divisore dei minori di ordine i di A divide quello dei minori di ordine i di AP. Riunendo i due fatti e usando la simmetria dell'equivalenza, vediamo che due matrici equivalenti A, B hanno gli stessi massimi comuni divisori dei minori di ciascun ordine, a meno di associati. Sia $B = \mathrm{diag}\{d_1, \ldots, d_r, 0, \ldots, 0\}$ una forma normale di A. Poiché $d_i \mid d_j$ per $i < j$, un massimo comun divisore dei minori di ordine i di B è $\Delta_i = d_1 d_2 \cdots d_i$. Ne segue l'asserto. □

Conseguenza immediata del Teorema 3.9 è che i fattori invarianti sono determinati a meno di fattori invertibili e che due matrici $m \times n$ sono equivalenti se e solo se hanno gli stessi fattori invarianti.

Esercizi.

1. Determinare una forma normale della matrice intera $\begin{pmatrix} 6 & 2 & 3 & 0 \\ 2 & 3 & -4 & 1 \\ -3 & 3 & 1 & 2 \\ -1 & 2 & -3 & 5 \end{pmatrix}$.

2. Determinare una forma normale della matrice

$$A = \begin{pmatrix} \lambda - 17 & 8 & 12 & -14 \\ -46 & \lambda + 22 & 35 & -41 \\ 2 & -1 & \lambda - 4 & 4 \\ -4 & 2 & 2 & \lambda - 3 \end{pmatrix}$$

   in $M_4(\mathbb{Q}[\lambda])$, dove λ è un'indeterminata. Determinare anche matrici invertibili P, Q tali che PAQ sia in forma normale.

3. Servirsi delle formule (27) per calcolare i fattori invarianti di $\begin{pmatrix} \lambda + 1 & 2 & -6 \\ 1 & \lambda & -3 \\ 1 & 1 & \lambda - 4 \end{pmatrix}$.

4. Dimostrare che, se D è euclideo, ogni matrice invertibile in $M_n(D)$ è un prodotto di matrici elementari. Dimostrare anche che ogni matrice elementare del III tipo è un prodotto di matrici elementari dei tipi I e II (si consideri dapprima il caso 2×2). Dedurne che se D è euclideo ogni matrice invertibile di $M_n(D)$ è un prodotto di matrici elementari dei tipi I e II.[4]

5. Dimostrare che, se F è un campo, ogni matrice di $M_n(F)$ di determinante 1 è un prodotto di matrici elementari del I tipo.

6. Sia D un dominio a ideali principali, siano $a_i \in D$, $1 \leq i \leq n$, e sia d un loro massimo comun divisore. Dimostrare che esiste una matrice invertibile $Q \in M_n(D)$ tale che $(a_1, a_2, \ldots, a_n)Q = (d, 0, \ldots, 0)$.

7. Dimostrare che, se $a_{11}, \ldots, a_{1n}$ sono primi fra loro, si possono scegliere $a_{kj} \in D$, $2 \leq k \leq n$, $1 \leq j \leq n$, in modo che (a_{ij}) sia invertibile in $M_n(D)$, con D dominio a ideali principali.

[4] Esistono domini a ideali principali nei quali alcune matrici invertibili non sono prodotti di matrici elementari. Un esempio si trova nel lavoro di P. M. Cohn, *On the structure of GL_2 of a ring*, Institut des Hautes Études Scientifiques, Publications mathématiques 30 (1966), pp. 5–54.

8. Sia $A \in M_n(D)$, con D euclideo e $\det A \neq 0$. Dimostrare che esiste $P \in M_n(D)$ invertibile per la quale PA ha la forma triangolare

$$\begin{pmatrix} d_1 & b_{12} & \cdots & b_{1n} \\ 0 & d_2 & \cdots & b_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_n \end{pmatrix},$$

con $d_i \neq 0$ e $\delta(b_{ji}) < \delta(d_i)$ per ogni elemento b_{ji} sopra la diagonale.

9. Dimostrare che per $A \in M_{m,n}(D)$, con D dominio a ideali principali, A e la sua trasposta hanno gli stessi fattori invarianti.

10. In un anello R si definiscano come sopra le matrici elementari $T_{ij}(a)$, $i \neq j$, $a \in R$. Verificare le relazioni seguenti, dette *relazioni di Steinberg*:

$$\begin{aligned} &\text{(i)} && T_{ij}(a)^{-1} = T_{ij}(-a), \\ &\text{(ii)} && T_{ij}(a)T_{ij}(b) = T_{ij}(a+b), \\ &\text{(iii)} && (T_{ij}(a), T_{jk}(b)) = T_{ik}(ab) \quad (k \neq i), \\ &\text{(iv)} && (T_{ij}(a), T_{kl}(b)) = 1 \quad (j \neq k,\ i \neq l), \end{aligned}$$

dove in generale $(x, y) = x^{-1}y^{-1}xy$.

3.8 Teorema di struttura dei moduli finitamente generati su un dominio a ideali principali

Siamo ora in grado di dimostrare il

Teorema fondamentale di struttura dei moduli finitamente generati su un dominio a ideali principali. *Se $M \neq 0$ è un modulo finitamente generato su un dominio a ideali principali D, M è somma diretta di moduli ciclici, $M = Dz_1 \oplus Dz_2 \oplus \cdots \oplus Dz_s$, i cui ideali d'ordine soddisfano*

$$\operatorname{ann} z_1 \supseteq \operatorname{ann} z_2 \supseteq \cdots \supseteq \operatorname{ann} z_s, \qquad \operatorname{ann} z_k \neq D. \tag{28}$$

Osservazione. Se $b \in \operatorname{ann} z$, allora $b(az) = a(bz) = 0$ per ogni $a \in D$. Quindi $\operatorname{ann}(az) \supseteq \operatorname{ann} z$. Ne segue che due generatori di uno stesso D-modulo ciclico hanno lo stesso annullatore. Pertanto $\operatorname{ann} z$ è indipendente dalla scelta del generatore z di Dz.

Dimostrazione. Abbiamo visto che, se $x_1, \ldots, x_n$ generano M, esiste l'epimorfismo η del modulo libero $D^{(n)}$, di base (e_i), $1 \leq i \leq n$, su M che manda e_i in x_i. Allora $M \simeq D^{(n)}/K$, e K è generato da un insieme finito di elementi $f_1, \ldots, f_m$, con $f_j = \sum_i a_{ji}e_i$. Abbiamo così la matrice delle relazioni $A = (a_{ji}) \in M_{m,n}(D)$. Sostituiamo ora la base (e_i) con (e'_i), dove $e'_i = \sum_{j=1}^n p_{ij}e_j$ e $P = (p_{ij})$ è invertibile in $M_n(D)$. Sostituiamo inoltre i generatori f_k, $1 \leq k \leq m$, con $f'_1, \ldots, f'_m$, dove $f'_k = \sum_{l=1}^m q_{kl}f_l$ e $Q = (q_{kl})$ è invertibile in $M_m(D)$. Come abbiamo visto nel § 3.6, la nuova matrice delle relazioni è QAP^{-1}. Per il Teorema 3.8 possiamo scegliere P, Q in modo che $QAP^{-1} = \operatorname{diag}\{d_1, \ldots, d_r, 0, \ldots, 0\}$, con $d_i \neq 0$ e $d_i \mid d_j$ per $i \leq j$. Le relazioni che legano i generatori f'_k di K alla base (e'_j) sono dunque

$$f'_1 = d_1e'_1, \ldots, f'_r = d_re'_r, \qquad f'_{r+1} = \cdots = f'_m = 0. \tag{29}$$

Poniamo $y_i = \sum_j p_{ij}x_j$, $1 \le i \le n$. Gli elementi $y_1, \dots, y_n$, immagini mediante η degli elementi della base (e'_j), costituiscono un altro insieme di generatori di M. Poiché $d_i e'_i = f'_i \in K$ per $1 \le i \le r$, si ha $d_i y_i = 0$ per gli indici corrispondenti. Supponiamo ora $\sum_{i=1}^n b_i y_i = 0$, con $b_i \in D$. Allora $\sum b_i e'_i \in K$, e quindi $\sum b_i e'_i = \sum c_i f'_i = \sum_{i=1}^r c_i d_i e'_i$. Poiché $(e'_1, \dots, e'_n)$ è una base di $D^{(n)}$, si ha $b_i = c_i d_i$ per $i \le r$ e $b_i = 0$ per $i > r$. Allora $b_i y_i = 0$ per ogni i. Abbiamo dimostrato che se $\sum b_i y_i = 0$ tutti gli addendi $b_i y_i$ sono nulli. Pertanto

$$M = \sum Dy_i = Dy_1 \oplus Dy_2 \oplus \cdots \oplus Dy_n.$$

Abbiamo inoltre provato che $b_i y_i = 0$ implica $b_i \in (d_i)$, ponendo $d_i = 0$ per $i > r$. Poiché $d_i y_i = 0$, si ha $\operatorname{ann} y_i = (d_i)$. Le condizioni di divisibilità sui d_i danno evidentemente $(d_1) \supseteq (d_2) \supseteq \cdots \supseteq (d_n)$. Ora, se d_i è un'unità, da $d_i y_i = 0$ segue $y_i = 0$. Possiamo dunque eliminare tale elemento dall'insieme dei generatori. Supponiamo che $d_1, \dots, d_t$ siano unità e $d_{t+1}, d_{t+2}, \dots$ non lo siano. Poniamo $z_1 = y_{t+1}$, $z_2 = y_{t+2}, \dots, z_s = y_n$, con $s = n-t$. Otteniamo $M = Dz_1 \oplus Dz_2 \oplus \cdots \oplus Dz_s$, con ogni $Dz_j \neq 0$, e valgono le condizioni (28). □

Esercizi.

1. Determinare la struttura di $\mathbb{Z}^{(3)}/K$ se K è generato da $f_1 = (2, 1, -3)$, $f_2 = (1, -1, 2)$.

2. Sia $D = \mathbb{Z}[\sqrt{-1}]$ l'anello degli interi di Gauss. Determinare la struttura di $D^{(3)}/K$, dove K è generato da $f_1 = (1, 3, 6)$, $f_2 = (2 + 3i, -3i, 12 - 18i)$, $f_3 = (2 - 3i, 6 + 9i, -18i)$, $i = \sqrt{-1}$. Mostrare che $M = D^{(3)}/K$ è finito, di ordine 352512.

3. Sia M l'ideale di $\mathbb{Z}[x]$ generato da 2 e x. Mostrare che M non è somma diretta di $\mathbb{Z}[x]$-moduli ciclici.

Gli esercizi restanti sviluppano una dimostrazione del teorema fondamentale di struttura indipendente dalla forma normale delle matrici (Teorema 3.8). In essi M è un modulo finitamente generato su un dominio a ideali principali D. Usiamo la lunghezza di un elemento di D definita nel § 3.7, estendendola allo zero con $\ell(0) = \infty$, maggiore di ogni intero. Per $x \in M$ poniamo inoltre $\ell(x) = \ell(d)$, dove $\operatorname{ann} x = (d)$.

4. Sia N un sottomodulo di M e sia $x \in M$. Mostrare che (i) $\operatorname{ann}(x + N) \supseteq \operatorname{ann} x$, con inclusione stretta se e solo se $Dx \cap N \neq 0$, (ii) $\ell(x + N) \le \ell(x)$, con disuguaglianza stretta se e solo se $Dx \cap N \neq 0$.

5. Siano $x_1, \dots, x_n$ un insieme di $n \ge 1$ generatori di M e $y = \sum_{i=1}^n a_i x_i$, con massimo comun divisore $(a_1, \dots, a_n) = 1$. Mostrare che esiste un insieme di n generatori $y_1, \dots, y_n$ con $y_1 = y$ (cfr. esercizio 7 del § 3.7). *Traccia.* Il caso $n = 1$ è evidente. Per $n = 2$ si scelgano $b_1, b_2 \in D$ con $a_1 b_1 + a_2 b_2 = 1$. Allora $y_1 = y$, $y_2 = -b_2 x_1 + b_1 x_2$ generano M. Per $n > 2$ si ponga $d = (a_2, \dots, a_n)$. Il caso $d = 0$, ossia $a_2 = \cdots = a_n = 0$, è immediato. Si supponga $d \neq 0$ e si scriva $a_j = da'_j$. Allora $(a'_2, \dots, a'_n) = 1$, quindi per induzione si hanno generatori $y_2 = \sum_{j=2}^n a'_j x_j, y_3, \dots, y_n$ per $N = \sum_{j=2}^n Dx_j$. Inoltre $(a_1, d) = 1$ e $y = a_1 x_1 + dy_2$. Il caso $n = 2$ mostra che esiste $z_2 \in P = Dx_1 + Dy_2$ tale che $Dy + Dz_2 = P$. Allora $y, z_2, y_3, \dots, y_n$ generano M.

6. Siano $x_1, \dots, x_n$ generatori di M tali che (i) n sia minimo e (ii) $\ell(x_1)$ sia minima tra tutti gli insiemi di n generatori di M. Mostrare che $M = Dx_1 \oplus N$, dove $N = \sum_{j=2}^n Dx_j$, e che $\operatorname{ann} x_1 \supseteq \operatorname{ann} y$ per ogni $y \in N$. Ne risulterà il teorema di struttura per induzione su n. *Traccia.* Se $Dx_1 \cap N \neq 0$, l'esercizio 4 dà $\ell(x_1 + N) < \ell(x_1)$. Allora $\operatorname{ann}(x_1 + N) = (a_1) \neq 0$ e $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$ per opportuni $a_i \in D$. Posti $d = (a_1, \dots, a_n)$ e $a_i = da'_i$, si ha

$(a'_1, \dots, a'_n) = 1$. Per l'esercizio 5 si hanno generatori $y_1 = \sum a'_i x_i, y_2, \dots, y_n$. Ma $dy_1 = 0$ e $\ell(y_1) \leq \ell(d) \leq \ell(a_1) < \ell(x_1)$, contro la scelta degli x_i. Per provare $\operatorname{ann} x_1 \supseteq \operatorname{ann} y$ per $y \in N$ basta provare l'inclusione per $y = x_j$, $j > 1$, e per simmetria basta $j = 2$. Supponiamo che non valga, e poniamo $\operatorname{ann} x_i = (d_i)$ per $i = 1, 2$. Allora $d_2 \neq 0$, per cui $d = (d_1, d_2) \neq 0$ e $\ell(d) < \ell(d_1) = \ell(x_1)$. Inoltre $d_i = dd'_i$ e $(d'_1, d'_2) = 1$, perciò si hanno generatori $y_1, y_2, \dots, y_n$ con $y_1 = d'_1 x_1 + d'_2 x_2$. Allora $\ell(y_1) < \ell(x_1)$, assurdo.

3.9 Moduli di torsione e componenti primarie. Teorema di invarianza

La decomposizione di un modulo finitamente generato su un dominio a ideali principali data dal teorema fondamentale di struttura non è in generale unica. Per esempio, se M è libero, ogni base $(e_1, \dots, e_n)$ determina una decomposizione $M = De_1 \oplus \cdots \oplus De_n$, con $\operatorname{ann} e_i = 0$. Passando a una base $(f_1, \dots, f_n)$ i cui elementi non siano semplicemente multipli degli e_i in un altro ordine, si ottiene una seconda decomposizione diretta $M = Df_1 \oplus \cdots \oplus Df_n$, diversa dalla prima. Nelle diverse decomposizioni di M in sottomoduli ciclici i cui ideali d'ordine soddisfano le inclusioni del teorema di struttura, vi è tuttavia qualcosa di invariante: la successione degli ideali d'ordine $\operatorname{ann} z_1, \operatorname{ann} z_2, \dots$ è la stessa per due qualsiasi decomposizioni di questo tipo. Il nostro prossimo obiettivo è dimostrarlo. Prima di passare alla dimostrazione, conviene introdurre il sottomodulo di torsione di un modulo su un dominio a ideali principali e studiarne alcune proprietà. Ciò faciliterà la dimostrazione del teorema di invarianza e ci aiuterà a comprendere meglio la struttura di questi moduli.

Sia M un modulo finitamente generato sul dominio a ideali principali D. Indichiamo con $\operatorname{tor} M$ il sottoinsieme degli $y \in M$ tali che $ay = 0$ per qualche $a \neq 0$ di D. Dunque $y \in \operatorname{tor} M$ se e solo se $\operatorname{ann} y \neq 0$. Se $a_i y_i = 0$, $i = 1, 2$, e $a_i \neq 0$, allora $a = a_1 a_2 \neq 0$ e $a(y_1 + y_2) = a_2 a_1 y_1 + a_1 a_2 y_2 = 0$. Inoltre $ay = 0$ implica $a(by) = b(ay) = 0$. Quindi $\operatorname{tor} M$ è un sottomodulo, che chiameremo *sottomodulo di torsione* di M. Diremo che M è un *modulo di torsione* se $M = \operatorname{tor} M$. Consideriamo ora la decomposizione $M = Dz_1 \oplus \cdots \oplus Dz_s$ con $\operatorname{ann} z_1 \supseteq \cdots \supseteq \operatorname{ann} z_s$. Supponiamo $\operatorname{ann} z_i \neq 0$ per $i \leq r$ e $\operatorname{ann} z_i = 0$ per $r < i \leq s$. Gli z_i con $i \leq r$ appartengono a $\operatorname{tor} M$, dunque $Dz_1 + \cdots + Dz_r \subseteq \operatorname{tor} M$. Viceversa, sia $y = b_1 z_1 + \cdots + b_s z_s \in \operatorname{tor} M$. Esiste $a \neq 0$ tale che $0 = ay = ab_1 z_1 + \cdots + ab_s z_s$. Ne segue $ab_i z_i = 0$ per ogni i, quindi $ab_i = 0$ per $i > r$. Poiché $a \neq 0$, abbiamo $b_i = 0$ per $i > r$ e $y \in Dz_1 + \cdots + Dz_r$. Pertanto

$$\operatorname{tor} M = Dz_1 \oplus \cdots \oplus Dz_r. \tag{30}$$

Si vede anche che $Dz_{r+1} + \cdots + Dz_s = Dz_{r+1} \oplus \cdots \oplus Dz_s$ è un sottomodulo libero di M, e $M = \operatorname{tor} M \oplus (Dz_{r+1} + \cdots + Dz_s)$. Abbiamo così il seguente teorema.

Teorema 3.10. *Ogni modulo finitamente generato su un dominio a ideali principali è somma diretta del proprio sottomodulo di torsione e di un sottomodulo libero.*

Se p è primo, definiamo la *p-componente* M_p di M come l'insieme degli $y \in M$ per i quali $p^k y = 0$ per qualche $k \in \mathbb{N}$. È un sottomodulo contenuto in $\operatorname{tor} M$. Se $p_1, \dots, p_h$ sono primi distinti (a meno di associati), le componenti corrispondenti sono indipendenti. Basta mostrare che $M_{p_1} \cap (M_{p_2} + \cdots + M_{p_h}) = 0$. Se y appartiene all'intersezione, si scrive $y = y_2 + \cdots + y_h$ con $y_i \in M_{p_i}$ e $p_i^{k_i} y_i = 0$ per qualche $k_i \in \mathbb{N}$. Quindi $p_2^{k_2} \cdots p_h^{k_h} y = 0$. D'altra parte $p_1^{k_1} y = 0$, poiché $y \in M_{p_1}$. Pertanto $\operatorname{ann} y$ contiene $p_1^{k_1}$ e $p_2^{k_2} \cdots p_h^{k_h}$, e contiene il loro massimo comun divisore 1. Ne segue $y = 0$. Mostreremo che quasi tutte le p-componenti, ossia tutte tranne un numero finito, sono nulle, e che $\operatorname{tor} M$ è somma diretta di tali componenti. Ciò seguirà dalla prima parte del lemma seguente.

Lemma. *(1) Se $M = Dx$, con $\operatorname{ann} x = (d)$, e $d = gh$ con $(g,h) = 1$, allora $M = Dy \oplus Dz$ con $\operatorname{ann} y = (g)$ e $\operatorname{ann} z = (h)$. (2) Se $M = Dy + Dz$, con $\operatorname{ann} y = (g)$, $\operatorname{ann} z = (h)$ e $(g,h) = 1$, allora $M = Dx$ con $\operatorname{ann} x = (gh)$.*

Dimostrazione. (1) Poniamo $y = hx$, $z = gx$. Si ha $x \in Dy + Dz$, perché esistono $a, b \in D$ tali che $ah + bg = 1$ e quindi $x = (ah + bg)x = a(hx) + b(gx) = ay + bz$. Poiché $M = Dx$, segue $M = Dy + Dz$. Se $u \in Dy \cap Dz$, si ha $gu = hu = 0$, essendo $gy = ghx = 0$ e $hz = hgx = 0$. Dunque $u = 1u = ahu + bgu = 0$, per cui $Dy \cap Dz = 0$ e $M = Dy \oplus Dz$. Si verifica inoltre che $\operatorname{ann} y = (g)$ e $\operatorname{ann} z = (h)$.

(2) Come in (1), abbiamo $M = Dy \oplus Dz$. Posto $x = y + z$, da $cx = 0$ segue $cy = cz = 0$. Allora c è multiplo di g e di h, quindi del loro minimo comune multiplo gh. Poiché $(gh)(y + z) = 0$, abbiamo $\operatorname{ann} x = (gh)$. Inoltre, se $ah + bg = 1$, si ha $y = ahy = ah(y + z) = ahx$, dunque $y \in Dx$. Segue $z = x - y \in Dx$, e così $Dx = M$. □

Dalla prima parte del lemma segue per induzione che, se $d = p_1^{e_1} \cdots p_t^{e_t}$ con p_i primi distinti e $\operatorname{ann} x = (d)$, allora $M = Dx$ si decompone in $Dx_1 \oplus \cdots \oplus Dx_t$, dove $\operatorname{ann} x_i = (p_i^{e_i})$. Ogni modulo ciclico di torsione è dunque somma diretta di moduli ciclici *primari*, nel senso che i loro ideali d'ordine sono della forma (p^e), con p primo. Possiamo servircene per dimostrare il seguente risultato.

Teorema 3.11. *Sia M un modulo di torsione finitamente generato su un dominio a ideali principali. Allora la componente primaria M_p è nulla per tutti i primi salvo un numero finito $p_1, \ldots, p_h$, e $M = M_{p_1} \oplus M_{p_2} \oplus \cdots \oplus M_{p_h}$.*

Dimostrazione. Siano $x_1, \ldots, x_n$ generatori, cosicché $M = Dx_1 + \cdots + Dx_n$. Siano $p_1, \ldots, p_h$ i distinti fattori primi di tutti i d_i per cui $\operatorname{ann} x_i = (d_i)$. Allora $Dx_i \subseteq M_{p_1} + \cdots + M_{p_h}$, quindi $M = M_{p_1} + \cdots + M_{p_h}$. Poiché questi sottomoduli sono indipendenti, la somma è diretta. Per ogni primo p diverso da tutti i p_i si ha ora $M_p = M_p \cap (M_{p_1} + \cdots + M_{p_h}) = 0$. □

Combinando il teorema fondamentale di struttura con la decomposizione di un modulo ciclico di torsione in moduli primari, otteniamo il risultato seguente.

Teorema 3.12. *Ogni modulo di torsione finitamente generato su un dominio a ideali principali è somma diretta di moduli ciclici primari.*

Basta infatti scrivere M come somma diretta di moduli ciclici mediante il teorema fondamentale. Come abbiamo visto, ciascuno di essi è somma diretta di sottomoduli ciclici primari, e tale è dunque anche M. Più precisamente, se $M = Dz_1 \oplus \cdots \oplus Dz_r$ e gli ideali $\operatorname{ann} z_i = (d_i)$ formano una catena decrescente, abbiamo $d_1 \mid d_2 \mid \cdots \mid d_r$. Possiamo supporre $d_i = p_1^{e_{1i}} \cdots p_h^{e_{hi}}$, con primi distinti ed $e_{j1} \le e_{j2} \le \cdots \le e_{jr}$, $1 \le j \le h$. Allora M è somma diretta di moduli ciclici con annullatori $(p_j^{e_{ji}})$, tralasciando gli esponenti nulli. Date queste potenze di primi si possono ricostruire i d_i. L'ultimo, d_s, è il minimo comune multiplo di tutte le potenze che compaiono. Eliminate le potenze di primi che sono fattori di d_s, il minimo comune multiplo di quelle restanti è d_{s-1}, e così via. Ad esempio, se $D = \mathbb{Z}$ e le potenze sono $3^2, 3^3, 3^4, 5^2, 5^4, 7, 7^3$, si ha $d_s = 3^4 5^4 7^3$, $d_{s-1} = 3^3 5^2 7$ e $d_{s-2} = d_1 = 3^2$. Se è data una decomposizione di M come somma diretta di sottomoduli ciclici primari, sommando opportunamente questi ultimi mediante la seconda parte del lemma precedente si ottiene la decomposizione in sottomoduli ciclici del teorema fondamentale. Nell'esempio, siano $x_1, \ldots, x_7$ i generatori degli addendi primari nell'ordine dato. Allora $Dx_3 + Dx_5 + Dx_7 = Dz_3$, $Dx_2 + Dx_4 + Dx_6 = Dz_2$, $Dx_1 = Dz_1$, e gli $\operatorname{ann} z_i = (d_i)$ soddisfano $\operatorname{ann} z_1 \supseteq \operatorname{ann} z_2 \supseteq \operatorname{ann} z_3$.

Siamo ora in grado di dimostrare il

Teorema di invarianza. *Sia* $M = Dz_1 \oplus \cdots \oplus Dz_s = Dw_1 \oplus \cdots \oplus Dw_t$, *con* $\operatorname{ann} z_1 \supseteq \cdots \supseteq \operatorname{ann} z_s$ *e* $\operatorname{ann} w_1 \supseteq \cdots \supseteq \operatorname{ann} w_t$, *e nessun addendo nullo. Allora* $s = t$ *e* $\operatorname{ann} z_i = \operatorname{ann} w_i$ *per* $1 \leq i \leq s$.

Dimostrazione. *I. Riduzione ai moduli di torsione.* Supponiamo che gli annullatori di z_i siano non nulli per $i \leq r$ e nulli per $i > r$, e che quelli di w_j siano non nulli per $j \leq u$ e nulli per $j > u$. Per (30) si ha

$$\operatorname{tor} M = Dz_1 \oplus \cdots \oplus Dz_r = Dw_1 \oplus \cdots \oplus Dw_u.$$

Inoltre $M/\operatorname{tor} M \simeq Dz_{r+1} \oplus \cdots \oplus Dz_s \simeq Dw_{u+1} \oplus \cdots \oplus Dw_t$. Questi ultimi sono moduli liberi, di ranghi $s - r$ e $t - u$. Per l'invarianza del rango dei moduli liberi su un anello commutativo (Teorema 3.4), $s - r = t - u$. Il numero degli $\operatorname{ann} z_i$ nulli è dunque uguale a quello degli $\operatorname{ann} w_j$ nulli. Resta da dimostrare il teorema per le decomposizioni di $\operatorname{tor} M$ appena scritte.

II. Riduzione ai moduli di torsione primari. Supponiamo ora che M sia di torsione e scomponiamo gli addendi ciclici Dz_i e Dw_j in somme dirette di sottomoduli ciclici primari. Per quanto si è visto, il teorema seguirà se due decomposizioni di M in sottomoduli ciclici primari hanno gli stessi ideali d'ordine, contando le molteplicità. Occorre cioè mostrare che per ogni potenza di primo p^e il numero di addendi ciclici con ideale d'ordine (p^e) è lo stesso nelle due decomposizioni. Fissiamo p e sommiamo in ciascuna decomposizione gli addendi con ideale d'ordine della forma (p^e), $e = 1, 2, \ldots$. Entrambe le somme coincidono con M_p. Basta quindi dimostrare il risultato per ogni M_p, ossia possiamo supporre $M = M_p$ primario.

III. Dimostrazione nel caso primario. Sia $M = M_p$. Allora $\operatorname{ann} z_i = (p^{e_i})$, $\operatorname{ann} w_j = (p^{f_j})$, e le inclusioni fra gli annullatori significano che $e_1 \leq \cdots \leq e_s$ e $f_1 \leq \cdots \leq f_t$. Per ogni $k \in \mathbb{N}$, $p^k M = \{p^k x \mid x \in M\}$ è un sottomodulo e $M \supseteq pM \supseteq p^2 M \supseteq \cdots$. Poniamo $M^{(k)} = p^k M / p^{k+1} M$. Ogni sua classe ha la forma $p^k x + p^{k+1} M$ e soddisfa $p(p^k x + p^{k+1} M) = p^{k+1} M = 0$ in $M^{(k)}$. Dunque l'ideale (p) annulla $M^{(k)}$, che si può considerare in modo naturale come un modulo su $\bar{D} = D/(p)$ (esercizio 2 del § 3.2). Poiché p è primo, $\bar{D}$ è un campo e $M^{(k)}$ è uno spazio vettoriale su $\bar{D}$. Possiamo metterne in relazione la dimensione con gli e_i e gli f_j. Si ha $p^k M = 0$ se $k \geq e_s$, mentre, se e_{q+1} è il primo $e_i > k$, si ha $p^k M = Dp^k z_{q+1} + \cdots + Dp^k z_s$. Le classi $p^k z_{q+1} + p^{k+1} M, \ldots, p^k z_s + p^{k+1} M$ formano una base di $M^{(k)}$ su $\bar{D}$. La sua dimensione è dunque il numero degli $e_i > k$. Analogamente è il numero degli $f_j > k$. Per ogni $k \in \mathbb{N}$ i due numeri coincidono, il che implica $s = t$ ed $e_i = f_i$ per $1 \leq i \leq s$. La dimostrazione è completa. □

Chiameremo gli ideali della successione $\operatorname{ann} z_1, \operatorname{ann} z_2, \ldots$, di cui abbiamo appena provato l'unicità, *ideali dei fattori invarianti* del modulo M. La dimostrazione mostra anche che, se M è di torsione, gli ideali d'ordine degli addendi ciclici primari in due decomposizioni qualsiasi di M sono invarianti. Li chiameremo *ideali dei divisori elementari* di M. Due moduli finitamente generati su un dominio a ideali principali sono evidentemente isomorfi se e solo se hanno gli stessi ideali dei fattori invarianti. Per due moduli di torsione, analogamente, si ha isomorfismo se e solo se coincidono gli ideali dei divisori elementari con le rispettive molteplicità.

Nel caso $D = \mathbb{Z}$ ogni ideale ha un unico generatore non negativo. Se $D = F[\lambda]$, con F campo, ogni ideale è generato da 0 o da un unico polinomio monico. In questi casi è naturale sostituire agli ideali i loro generatori normalizzati. Questi si chiamano *fattori invarianti* e *divisori elementari* del modulo.

Esercizi.

1. Sia $D = \mathbb{R}[\lambda]$ e sia M somma diretta di moduli ciclici i cui ideali d'ordine sono generati rispettivamente da $(\lambda-1)^3$, $(\lambda^2+1)^2$, $(\lambda-1)(\lambda^2+1)^4$, $(\lambda+2)(\lambda^2+1)^2$. Determinare i divisori elementari e i fattori invarianti di M.

2. Dimostrare che un modulo di torsione M su un dominio a ideali principali D è irriducibile (esercizio 7 del § 3.3) se e solo se $M = Dz$ con $\operatorname{ann} z = (p)$, p primo. Dimostrare che un modulo finitamente generato $M \neq 0$ è *indecomponibile*, nel senso che non è somma diretta di due sottomoduli non nulli, se e solo se $M = Dz$ con $\operatorname{ann} z = 0$ oppure $\operatorname{ann} z = (p^e)$, p primo.

3. Si definisca il *rango* di un modulo finitamente generato M su un dominio a ideali principali D come il rango del modulo libero $M/\operatorname{tor} M$. (È libero perché, scrivendo $M = \operatorname{tor} M \oplus F$ come nel Teorema 3.10, è isomorfo a F libero.) Se $M \simeq D^{(n)}/K$, mostrare che $\operatorname{rg} M = n - \operatorname{rg} K$. Mostrare inoltre che, per ogni sottomodulo N di M, N e M/N sono finitamente generati e $\operatorname{rg} M = \operatorname{rg} N + \operatorname{rg}(M/N)$.

4. Sia M un modulo di torsione su D con ideali dei fattori invarianti $(d_1) \supseteq \cdots \supseteq (d_s)$. Mostrare che ogni immagine omomorfa $\bar{M}$ di M è di torsione e i suoi ideali $(\bar{d}_1) \supseteq \cdots \supseteq (\bar{d}_t)$ soddisfano $t \leq s$, $\bar{d}_t \mid d_s$, $\bar{d}_{t-1} \mid d_{s-1}, \ldots, \bar{d}_1 \mid d_{s-t+1}$. (Si consideri dapprima il caso in cui M è primario.)

5. Siano $A, B \in M_n(D)$ con $\det(AB) \neq 0$, D dominio a ideali principali. Siano $\operatorname{diag}\{a_1, \ldots, a_n\}$, $\operatorname{diag}\{b_1, \ldots, b_n\}$ e $\operatorname{diag}\{c_1, \ldots, c_n\}$ forme normali di A, B, AB rispettivamente ($a_i \mid a_{i+1}$, ecc.). Dimostrare che $a_i \mid c_i$ e $b_i \mid c_i$ per $1 \leq i \leq n$.

6. Mostrare che l'asserzione dell'esercizio 4 relativa a un'immagine omomorfa $\bar{M}$ di M vale per ogni sottomodulo N di M.

7. Un sottomodulo N di M si dice *puro* se, per ogni $y \in N$ e $a \in D$, l'equazione $ax = y$ ammette soluzione in M se e solo se l'ammette in N. Mostrare che ogni addendo diretto è puro. Mostrare che, se N è puro in M e $\operatorname{ann}(x+N) = (d)$, si può scegliere il rappresentante x della classe $x + N$ in modo che $\operatorname{ann} x = (d)$.

8. Mostrare che ogni sottomodulo puro N di un modulo di torsione finitamente generato M su un dominio a ideali principali è un addendo diretto di M.

9. Sia M un modulo di torsione finitamente generato su un dominio a ideali principali. Mostrare che ogni sottomodulo ciclico Dz tale che $\operatorname{ann} z \subseteq \operatorname{ann} x$ per ogni $x \in M$ è puro. Per l'esercizio 8 Dz è dunque un addendo diretto.

3.10 Applicazioni ai gruppi abeliani e alle trasformazioni lineari

Specializziamo dapprima la teoria di struttura dei moduli finitamente generati M su un dominio a ideali principali D al caso $D = \mathbb{Z}$. Allora M è un gruppo abeliano dotato di un insieme finito di generatori. In particolare può essere un qualunque gruppo abeliano finito. Il teorema fondamentale di struttura afferma ora che ogni gruppo abeliano finitamente generato M è somma diretta di gruppi ciclici, $M = \langle z_1 \rangle \oplus \cdots \oplus \langle z_s \rangle$, dove $\operatorname{ann} z_i = (d_i)$ e $d_1 \mid d_2 \mid \cdots \mid d_s$. Se si normalizzano i d_i scegliendoli non negativi, l'ordine di z_i è d_i per $d_i > 0$ ed è infinito per $d_i = 0$. Il sottogruppo (ossia il sottomodulo) di torsione di M è l'insieme degli elementi di ordine finito. Nella decomposizione precedente coincide con $\langle z_1 \rangle + \cdots + \langle z_r \rangle$, dove $d_1 > 0, \ldots, d_r > 0$ e i d_i successivi sono nulli. Poiché $|\langle z_i \rangle| = d_i$ per $i \leq r$, il gruppo $\operatorname{tor} M$ è finito, di ordine $\prod_{i=1}^{r} d_i$. Il secondo teorema di struttura (Teorema 3.10) implica che ogni gruppo abeliano finitamente

generato è somma diretta di un gruppo finito e di un gruppo abeliano libero. In ciascuna decomposizione la componente finita è determinata univocamente, come sottogruppo di torsione. La componente libera può non essere unica, ma il suo rango è un invariante.

Il risultato sulla decomposizione di un modulo di torsione in somma diretta di moduli ciclici primari assume qui la forma seguente: ogni gruppo abeliano finito è somma diretta di gruppi ciclici di ordine una potenza di primo. Le potenze di primi che compaiono in una tale decomposizione, contate con le rispettive molteplicità, sono determinate univocamente. Si chiamano *invarianti* del gruppo abeliano finito. Due gruppi abeliani finiti sono evidentemente isomorfi se e solo se hanno gli stessi invarianti. Per comodità riuniamo i risultati principali nel seguente teorema.

Teorema 3.13. *Ogni gruppo abeliano finitamente generato è somma diretta di un gruppo finito, il suo sottogruppo di torsione, e di un gruppo abeliano libero. Il rango della componente libera è un invariante. Ogni gruppo abeliano finito è somma diretta di gruppi ciclici di ordine una potenza di primo. Tali ordini, con le rispettive molteplicità, sono determinati univocamente e costituiscono un sistema completo di invarianti: due gruppi abeliani finiti sono isomorfi se e solo se hanno gli stessi invarianti.*

Applichiamo ora i risultati ottenuti allo studio di una singola trasformazione lineare T in uno spazio vettoriale di dimensione finita V su un campo F. Sia $(u_1, \dots, u_n)$ una base di V su F e scriviamo

$$Tu_i = \sum_{j=1}^{n} a_{ij} u_j, \qquad i = 1, 2, \dots, n. \tag{31}$$

Allora $A = (a_{ij})$ è la matrice di T rispetto alla base data. Ricordiamo che, se $(v_1, \dots, v_n)$ è una seconda base e $v_i = \sum_j s_{ij} u_j$ con $S = (s_{ij})$ invertibile, la matrice di T rispetto a questa base è SAS^{-1}. Matrici legate in questo modo si dicono *simili*. Come nel § 3.2, possiamo fare di V un $F[\lambda]$-modulo definendo l'azione di ogni polinomio $g(\lambda) = b_0 + b_1\lambda + \cdots + b_m\lambda^m$ su $x \in V$ mediante

$$g(\lambda)x = b_0 x + b_1(Tx) + \cdots + b_m(T^m x).$$

Evidentemente questa azione estende quella di F in modo che $\lambda x = Tx$.

Osserviamo anzitutto che V è un $F[\lambda]$-modulo di torsione. Sia infatti $x \in V$ e consideriamo $x, \lambda x, \lambda^2 x, \dots$. Poiché $\dim_F V = n$, esiste $m \leq n$ tale che $\lambda^m x$ è combinazione lineare di $x, \lambda x, \dots, \lambda^{m-1} x$, diciamo $\lambda^m x = b_0 x + b_1 \lambda x + \cdots + b_{m-1}\lambda^{m-1} x$, con $b_i \in F$. Allora $g(\lambda) = \lambda^m - b_{m-1}\lambda^{m-1} - \cdots - b_0$ è un polinomio non nullo tale che $g(\lambda)x = 0$. Dunque $\operatorname{ann} x$ contiene $g(\lambda) \neq 0$ e non è nullo.

La base $(u_1, \dots, u_n)$ di V su F è evidentemente un insieme di generatori di V come $F[\lambda]$-modulo, sebbene in generale non ne sia una base. Abbiamo l'omomorfismo η del modulo libero $F[\lambda]^{(n)}$ di base $(e_1, \dots, e_n)$ su V che manda e_i in u_i, $1 \leq i \leq n$. Per studiare V come $F[\lambda]$-modulo con il nostro metodo, occorre un insieme di generatori di $K = \ker \eta$. Ce lo dà il seguente lemma.

Lemma. *Gli elementi $f_i = \lambda e_i - \sum_{j=1}^{n} a_{ij} e_j$, $1 \leq i \leq n$, formano una base di K.*

Dimostrazione. Da $Tu_i = \sum_j a_{ij} u_j$ segue $f_i \in K$. Le relazioni $\lambda e_i = f_i + \sum_j a_{ij} e_j$ consentono di scrivere ogni elemento $\sum g_i(\lambda) e_i$ nella forma $\sum h_i(\lambda) f_i + \sum b_i e_i$, con $b_i \in F$. Se l'elemento appartiene a K, anche $\sum b_i e_i \in K$, quindi $\sum b_i u_i = 0$. Poiché gli u_i formano una base di V su

F, tutti i b_i sono nulli e l'elemento di K dato è della forma $\sum h_i(\lambda)f_i$. Dunque gli f_i generano K. Supponiamo ora $\sum h_i(\lambda)f_i = 0$. Allora

$$\sum_{i=1}^{n} h_i(\lambda)\lambda e_i = \sum_{i,j=1}^{n} h_i(\lambda)a_{ij}e_j, \qquad h_i(\lambda)\lambda = \sum_j h_j(\lambda)a_{ji},$$

poiché gli e_i sono una base di $F[\lambda]^{(n)}$. Se qualche $h_i(\lambda)$ fosse non nullo, ne potremmo scegliere uno $h_r(\lambda)$ di grado massimo. La relazione $h_r(\lambda)\lambda = \sum_j h_j(\lambda)a_{jr}$ sarebbe impossibile per confronto dei gradi. Dunque tutti gli h_i sono nulli e gli f_i formano una base di K. □

La matrice che esprime la base (f_i) di K mediante la base (e_i) di $F[\lambda]^{(n)}$ è

$$\lambda 1 - A = \begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{pmatrix}. \tag{32}$$

La sua forma normale dà quindi i fattori invarianti di V come $F[\lambda]$-modulo, e di conseguenza la decomposizione di questo modulo in somma diretta di moduli ciclici. Il determinante $\det(\lambda 1 - A)$ si chiama *polinomio caratteristico* di A. Ha la forma

$$f(\lambda) = \det(\lambda 1 - A) = \lambda^n - a_1\lambda^{n-1} + \cdots + (-1)^n a_n. \tag{33}$$

Qui $a_1 = \sum_i a_{ii}$ si chiama *traccia* della matrice A, denotata $\operatorname{tr} A$, e $a_n = \det A$. In generale a_i è la somma dei minori principali (diagonali) di ordine i di A. Poiché $f(\lambda) \neq 0$ è il prodotto dei fattori invarianti monici di $\lambda 1 - A$, nessuno di questi è nullo. Ciò segue anche dal fatto che V è un modulo di torsione su $F[\lambda]$. Una forma normale di $\lambda 1 - A$ è quindi

$$\operatorname{diag}\{1, \ldots, 1, d_1(\lambda), \ldots, d_s(\lambda)\}, \tag{34}$$

con i $d_i(\lambda)$ monici di grado positivo e $d_i(\lambda) \mid d_j(\lambda)$ per $i \leq j$. Per i risultati del § 3.8, se P, Q sono matrici invertibili in $M_n(F[\lambda])$ tali che

$$P(\lambda 1 - A)Q = \operatorname{diag}\{1, \ldots, 1, d_1(\lambda), \ldots, d_s(\lambda)\}, \tag{35}$$

scrivendo $Q^{-1} = (q_{ij}^*)$ e ponendo $v_i = \sum_j q_{ij}^* u_j$, $z_i = v_{n-s+i}$, abbiamo

$$V = F[\lambda]z_1 \oplus F[\lambda]z_2 \oplus \cdots \oplus F[\lambda]z_s, \qquad \operatorname{ann} z_i = (d_i(\lambda)). \tag{36}$$

Serviamoci di (36) per ottenere una matrice canonica della trasformazione lineare T. Supponiamo dapprima $s = 1$, ossia che $V = F[\lambda]z$ sia ciclico come $F[\lambda]$-modulo. Allora $\operatorname{ann} z = (f(\lambda))$, con $f(\lambda)$ polinomio caratteristico di A dato dalla (33). Poiché f è il polinomio monico di grado minimo che annulla z, i vettori $z, \lambda z, \ldots, \lambda^{n-1}z$ sono linearmente indipendenti. Essi formano dunque una base di V su F. Si ha

$$\begin{aligned} Tz &= \lambda z, \\ T(\lambda z) &= \lambda(\lambda z) = \lambda^2 z, \\ &\vdots \\ T(\lambda^{n-2}z) &= \lambda^{n-1}z, \\ T(\lambda^{n-1}z) &= \lambda^n z = a_1\lambda^{n-1}z - a_2\lambda^{n-2}z + \cdots + (-1)^{n-1}a_n z. \end{aligned} \tag{37}$$

La matrice di T rispetto alla base $(z, \lambda z, \dots, \lambda^{n-1}z)$ è pertanto

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ (-1)^{n-1}a_n & (-1)^{n-2}a_{n-1} & \cdots & -a_2 & a_1 \end{pmatrix}.$$

In generale, dato un polinomio monico $d(\lambda) = \lambda^m - b_{m-1}\lambda^{m-1} - \cdots - b_0$, si chiama *matrice compagna* di $d(\lambda)$ la matrice

$$B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ b_0 & b_1 & \cdots & b_{m-2} & b_{m-1} \end{pmatrix}. \tag{38}$$

Nel caso ciclico la matrice di T rispetto alla base $(z, \lambda z, \dots, \lambda^{n-1}z)$ è dunque la matrice compagna del polinomio caratteristico $f(\lambda)$.

Consideriamo ora il caso generale della decomposizione (36). Si ottiene una base di V su F concatenando basi su F dei sottomoduli ciclici $F[\lambda]z_i$. Se $\deg d_i(\lambda) = n_i$, allora $(z_i, \lambda z_i, \dots, \lambda^{n_i-1}z_i)$ è una base di $F[\lambda]z_i$. Se

$$d_i(\lambda) = \lambda^{n_i} - b_{i,n_i-1}\lambda^{n_i-1} - \cdots - b_{i,0}, \tag{39}$$

si ha $T(\lambda^{n_i-1}z_i) = b_{i,0}z_i + b_{i,1}\lambda z_i + \cdots + b_{i,n_i-1}\lambda^{n_i-1}z_i$. La matrice di T rispetto alla base

$$(z_1, \lambda z_1, \dots, \lambda^{n_1-1}z_1;\ z_2, \lambda z_2, \dots, \lambda^{n_2-1}z_2;\ \dots;\ z_s, \lambda z_s, \dots, \lambda^{n_s-1}z_s)$$

ha chiaramente la forma

$$B = \operatorname{diag}(B_1, B_2, \dots, B_s), \tag{40}$$

dove B_i è la matrice compagna di $d_i(\lambda)$. La matrice B si chiama *forma canonica razionale* della trasformazione lineare T. La si può scrivere appena si conoscano i fattori invarianti di $\lambda 1 - A$, calcolabili mediante una successione di trasformazioni elementari sulle righe e sulle colonne di $\lambda 1 - A$.

Esempio. Sia T la trasformazione lineare in $V = \mathbb{Q}^{(3)}$ data da

$$Tu_1 = -u_1 - 2u_2 + 6u_3, \qquad Tu_2 = -u_1 + 3u_3, \qquad Tu_3 = -u_1 - u_2 + 4u_3.$$

Le matrici sono

$$A = \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}, \qquad \lambda 1 - A = \begin{pmatrix} \lambda + 1 & 2 & -6 \\ 1 & \lambda & -3 \\ 1 & 1 & \lambda - 4 \end{pmatrix}.$$

Abbiamo

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 2 - \lambda & -3 \end{pmatrix} (\lambda 1 - A) \begin{pmatrix} 1 & 3 & \lambda - 3 \\ 0 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} = \operatorname{diag}(1, \lambda - 1, (\lambda - 1)^2).$$

Le due matrici ai lati di $\lambda 1 - A$ hanno determinante 1, quindi sono invertibili in $M_3(\mathbb{Q}[\lambda])$. I fattori invarianti di V come $\mathbb{Q}[\lambda]$-modulo sono $\lambda - 1$ e $(\lambda-1)^2 = \lambda^2 - 2\lambda + 1$. La forma canonica razionale è

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Il metodo fornisce anche una matrice in $M_3(\mathbb{Q})$ che trasforma A nella sua forma canonica razionale. Con le notazioni precedenti,

$$Q = \begin{pmatrix} 1 & 3 & \lambda - 3 \\ 0 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \qquad Q^{-1} = \begin{pmatrix} 1 & \lambda & -3 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \end{pmatrix}.$$

Dunque $v_1 = u_1 + \lambda u_2 - 3u_3 = 0$, $v_2 = -u_2 + u_3$, $v_3 = -u_2$. Una base (z_1, z_2, z_3) che dà la forma canonica razionale è

$$z_1 = v_2 = -u_2 + u_3, \qquad z_2 = v_3 = -u_2, \qquad z_3 = \lambda v_3 = u_1 - 3u_3.$$

La matrice che esprime la nuova base in quella iniziale è

$$S = \begin{pmatrix} 0 & -1 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & -3 \end{pmatrix}.$$

Si può verificare che

$$S \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix} S^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Esiste una seconda forma canonica (ossia una seconda matrice canonica) per una trasformazione lineare T, la cosiddetta *forma di Jordan*, che si definisce se i fattori invarianti si scompongono in fattori lineari $\lambda - r$ in $F[\lambda]$. Ciò accade sempre se F è il campo dei numeri complessi $\mathbb{C}$ (si veda il capitolo 5). Sotto questa ipotesi i divisori elementari di V come $F[\lambda]$-modulo sono della forma $(\lambda - r)^e$, $r \in F$. A ciascuno corrisponde un addendo diretto ciclico $F[\lambda]w$ con $\operatorname{ann} w = ((\lambda - r)^e)$. Lo spazio $F[\lambda]w$ su F ha la base

$$w, (\lambda - r)w, (\lambda - r)^2 w, \ldots, (\lambda - r)^{e-1} w, \tag{41}$$

e si ha

$$\begin{aligned} Tw &= \lambda w = rw + (\lambda - r)w, \\ T((\lambda - r)w) &= \lambda(\lambda - r)w = r(\lambda - r)w + (\lambda - r)^2 w, \\ &\vdots \\ T((\lambda - r)^{e-2} w) &= r(\lambda - r)^{e-2} w + (\lambda - r)^{e-1} w, \\ T((\lambda - r)^{e-1} w) &= r(\lambda - r)^{e-1} w. \end{aligned}$$

La matrice della restrizione di T a $F[\lambda]w$ rispetto alla base (41) è pertanto

$$\begin{pmatrix} r & 1 & & 0 \\ & r & \ddots & \\ & & \ddots & 1 \\ 0 & & & r \end{pmatrix}. \tag{42}$$

Se $V = F[\lambda]w_1 \oplus \cdots \oplus F[\lambda]w_t$ e $\operatorname{ann} w_i = ((\lambda - r_i)^{e_i})$, concatenando le basi appena indicate per gli spazi ciclici si ottiene una base di V su F rispetto alla quale la matrice di T è la *forma canonica di Jordan*

$$\operatorname{diag}(C_1, C_2, \ldots, C_t), \tag{43}$$

dove C_i ha e_i righe ed e_i colonne ed è della forma (42) con $r = r_i$.

Esempio. Nell'esempio precedente i fattori invarianti erano $\lambda - 1$ e $(\lambda - 1)^2$. Essi coincidono in questo caso con i divisori elementari, e la forma canonica di Jordan è

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

I risultati ottenuti possono anche enunciarsi in termini di matrici anziché di trasformazioni lineari. Data $A \in M_n(F)$, possiamo definire la trasformazione lineare T in $V = F^{(n)}$ tale che $Tu_i = \sum_j a_{ij}u_j$, dove $(u_1, \ldots, u_n)$ è una base di V su F. Le matrici simili ad A sono le matrici di T rispetto alle diverse basi ordinate di V su F. Chiamiamo *forma canonica razionale* di A (oppure *forma canonica di Jordan* di A, quando è definita) quella di T. Ne segue immediatamente che due matrici A, B sono simili se e solo se hanno la stessa forma canonica razionale, oppure la stessa forma canonica di Jordan, quando questa è definita, a meno dell'ordine dei blocchi.

Anche i risultati classici sui polinomi caratteristico e minimo delle matrici sono conseguenze dei nostri risultati. Li dedurremo ora. Con A, T, u_i come sopra, sia $P(\lambda 1 - A)Q = \operatorname{diag}\{1, \ldots, 1, d_1(\lambda), \ldots, d_s(\lambda)\}$ la forma normale di $\lambda 1 - A$, con P, Q invertibili in $M_n(F[\lambda])$ e $d_i(\lambda)$ monici di grado positivo. Allora $\det P$ e $\det Q$ sono elementi non nulli di F, e il polinomio caratteristico di A è

$$f(\lambda) = \det(\lambda 1 - A) = d_1(\lambda)d_2(\lambda)\cdots d_s(\lambda). \tag{44}$$

Inoltre $d_i(\lambda) \mid d_j(\lambda)$ per $i \leq j$ e $V = F[\lambda]z_1 \oplus \cdots \oplus F[\lambda]z_s$, con $\operatorname{ann} z_i = (d_i(\lambda))$. Poniamo $m(\lambda) = d_s(\lambda)$. Poiché $d_i(\lambda) \mid m(\lambda)$, abbiamo $m(\lambda)z_i = 0$. Ogni $x \in V$ si scrive $x = \sum g_i(\lambda)z_i$, perciò $m(\lambda)x = 0$. Dunque $m(T) = 0$ o, equivalentemente, $m(A) = 0$. Se $g(T) = 0$, ossia $g(A) = 0$, allora $g(\lambda)z_s = 0$, e quindi $m(\lambda) \mid g(\lambda)$. Pertanto $m(\lambda)$ è il polinomio monico di grado minimo per cui $m(A) = 0$. Dalla (44) segue $f(A) = 0$. Poiché ogni $d_i(\lambda)$ divide $m(\lambda)$, $f(\lambda)$ e $m(\lambda)$ hanno gli stessi fattori irriducibili e possono differire solo nelle rispettive molteplicità. Infine, dalle formule del Teorema 3.9 per i fattori invarianti risulta $m(\lambda) = d_s(\lambda) = f(\lambda)/\Delta_{n-1}(\lambda)$, dove $\Delta_{n-1}(\lambda)$ è il massimo comun divisore monico dei minori di ordine $n-1$ di $\lambda 1 - A$. Riuniamo questi risultati, dovuti a Hamilton, Cayley e Frobenius, nel seguente teorema.

Teorema 3.14. *Sia $A \in M_n(F)$, con F campo, e sia $f(\lambda) = \det(\lambda 1 - A)$ il suo polinomio caratteristico. Allora $f(A) = 0$. Sia inoltre $\Delta_{n-1}(\lambda)$ il massimo comun divisore monico dei minori di ordine $n-1$ di $\lambda 1 - A$, e si ponga $m(\lambda) = f(\lambda)/\Delta_{n-1}(\lambda)$. Allora $m(A) = 0$ e $m(\lambda)$ divide ogni polinomio $g(\lambda)$ tale che $g(A) = 0$. Inoltre $m(\lambda)$ e $f(\lambda)$ hanno gli stessi fattori primi in $F[\lambda]$.*

Esercizi.

1. Determinare il numero di gruppi abeliani di ordine 360 a due a due non isomorfi.

2. Sia $\mathbb{Z}^{(n)}$ il $\mathbb{Z}$-modulo libero di base $(e_1, \ldots, e_n)$, e sia K il sottomodulo generato dagli elementi $f_i = \sum_{j=1}^{n} a_{ij} e_j$, con $a_{ij} \in \mathbb{Z}$ e $d = \det(a_{ij}) \neq 0$. Mostrare che $|\mathbb{Z}^{(n)}/K| = |d|$.

3. Sia $a + b\sqrt{-1}$ un elemento non nullo dell'anello degli interi di Gauss $\mathbb{Z}[\sqrt{-1}]$. Mostrare che $\left|\mathbb{Z}[\sqrt{-1}]/(a + b\sqrt{-1})\right| = a^2 + b^2$.

4. Verificare che il polinomio caratteristico di

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -2 & -2 & 0 & 1 \\ -2 & 0 & -1 & -2 \end{pmatrix}$$

è un prodotto di fattori lineari in $\mathbb{Q}[\lambda]$. Determinare le forme canoniche razionale e di Jordan di A in $M_4(\mathbb{Q})$. Trovare anche matrici che realizzino le due similitudini.

5. Dimostrare che, se F è un campo, $A, B \in M_n(F)$ sono simili se e solo se $\lambda 1 - A$ e $\lambda 1 - B$ sono equivalenti in $M_n(F[\lambda])$.

6. Dimostrare che ogni matrice quadrata A è simile alla propria trasposta.

7. Mostrare che il modulo su $F[\lambda]$ determinato da T è ciclico se e solo se il polinomio caratteristico $f(\lambda)$ è il polinomio minimo di T.

8. Dimostrare che ogni matrice nilpotente di $M_n(F)$ è simile a $\mathrm{diag}(N_1, \ldots, N_s)$, dove ogni blocco N_i ha la forma

$$\begin{pmatrix} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}.$$

9. Mostrare che $A \in M_n(\mathbb{C})$ è simile a una matrice diagonale $\mathrm{diag}\{r_1, \ldots, r_n\}$, $r_i \in \mathbb{C}$, se e solo se il polinomio minimo $m(\lambda)$ ha radici distinte.

10. Mostrare che, se $A^2 = A$, allora A è simile a una matrice $\mathrm{diag}\{1, \ldots, 1, 0, \ldots, 0\}$.

11. (Weyr.) Mostrare che $A, B \in M_n(\mathbb{C})$ sono simili se e solo se, per ogni $a \in \mathbb{C}$ e ogni $k = 1, 2, 3, \ldots$, $\mathrm{rg}(a1 - A)^k = \mathrm{rg}(a1 - B)^k$.

12. Mostrare che le seguenti matrici di $M_p(\mathbb{Z}/(p))$, con p primo, sono simili:

$$\begin{pmatrix} 0 & 1 & & 0 \\ & 0 & \ddots & \\ & & \ddots & 1 \\ 1 & & & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 & & 0 \\ & 1 & \ddots & \\ & & \ddots & 1 \\ 0 & & & 1 \end{pmatrix}.$$

13. Sia P la matrice compagna di un polinomio monico irriducibile $p(\lambda)$ di grado m e sia $N = e_{1m}$. Mostrare che il polinomio minimo della matrice $em \times em$

$$B = \begin{pmatrix} P & N & & 0 \\ & P & \ddots & \\ & & \ddots & N \\ 0 & & & P \end{pmatrix}$$

è $p(\lambda)^e$. Dedurne che, se i divisori elementari di $\lambda 1 - A$ sono $p_1(\lambda)^{e_1}, \dots, p_t(\lambda)^{e_t}$, con $p_i(\lambda)$ irriducibili, allora A è simile a $\operatorname{diag}(B_1, \dots, B_t)$, dove B_i ha la forma di B con P matrice compagna di $p_i(\lambda)$ e con e_i blocchi diagonali.

14. Mostrare che ogni matrice di $M_n(\mathbb{R})$ è simile a una matrice diagonale a blocchi, ciascuno dei quali ha una delle forme

$$\begin{pmatrix} r & 1 & & 0 \\ & r & \ddots & \\ & & \ddots & 1 \\ 0 & & & r \end{pmatrix}, \qquad \begin{pmatrix} H & I_2 & & 0 \\ & H & \ddots & \\ & & \ddots & I_2 \\ 0 & & & H \end{pmatrix}, \qquad H = \begin{pmatrix} 0 & 1 \\ -b & a \end{pmatrix},$$

con $a^2 < 4b$.

15. Sia R un anello commutativo, $R^{(n)}$ il modulo libero di base $(e_1, \dots, e_n)$ e η l'R-endomorfismo dato da $\eta e_i = \sum_j a_{ij} e_j$, con $A = (a_{ij}) \in M_n(R)$. Si faccia di $R^{(n)}$ un $R[\lambda]$-modulo come nel caso di un campo: ax, per $a \in R$, è definito come in $R^{(n)}$, e $\lambda x = \eta x$. Si hanno allora le relazioni

$$\begin{aligned} (\lambda - a_{11})e_1 - a_{12}e_2 - \cdots - a_{1n}e_n &= 0, \\ -a_{21}e_1 + (\lambda - a_{22})e_2 - \cdots - a_{2n}e_n &= 0, \\ &\vdots \\ -a_{n1}e_1 - a_{n2}e_2 - \cdots + (\lambda - a_{nn})e_n &= 0. \end{aligned}$$

Sia A_{ij} il cofattore dell'elemento (i, j) in $\lambda 1 - A$. Moltiplicare le relazioni precedenti rispettivamente per $A_{1i}, A_{2i}, \dots, A_{ni}$ e sommare. Mostrare che si ottengono le relazioni $f(\lambda)e_i = 0$, dove $f(\lambda) = \det(\lambda 1 - A)$. Allora $f(\eta) = 0$ e, per la corrispondenza fra $\operatorname{End}_R R^{(n)}$ e $M_n(R)$ (§ 3.4), si ottiene il teorema di Hamilton–Cayley per le matrici su R: $f(A) = 0$.

3.11 L'anello degli endomorfismi di un modulo finitamente generato su un dominio a ideali principali

Un problema interessante consiste nel determinare le matrici B di ordine n, a coefficienti in un campo F, che commutano con una matrice data $A \in M_n(F)$. In termini geometrici, il problema è determinare le trasformazioni lineari U di uno spazio vettoriale V di dimensione n su F che commutano con una trasformazione lineare data T di V. Una tale U è un endomorfismo del gruppo additivo di V per cui $U(ax) = a(Ux)$, $a \in F$, e $U(Tx) = T(Ux)$. Considerando come prima V quale $F[\lambda]$-modulo, l'ultima condizione diventa $U(\lambda x) = \lambda(Ux)$, da cui $U(\lambda^k x) = \lambda^k(Ux)$. Allora $U(f(\lambda)x) = f(\lambda)(Ux)$ per ogni polinomio $f(\lambda) \in F[\lambda]$, cosicché U è un endomorfismo di V considerato come $F[\lambda]$-modulo. Viceversa, questa condizione è sufficiente affinché U sia una trasformazione lineare di V su F che commuta con T: essa comprende infatti le condizioni che U sia un endomorfismo di gruppi, che $U(ax) = a(Ux)$ per $a \in F$ e che $U(Tx) = U(\lambda x) = \lambda(Ux) = T(Ux)$.

Più in generale, consideriamo il problema di determinare esplicitamente l'anello D' degli endomorfismi, ossia $\operatorname{Hom}(M, M)$, di un modulo finitamente generato M su un dominio a ideali principali D. Partiamo da una decomposizione $M = Dz_1 \oplus \cdots \oplus Dz_s$ con $\operatorname{ann} z_1 \supseteq \cdots \supseteq \operatorname{ann} z_s$ e $\operatorname{ann} z_i = (d_i) \neq 0$ per $i \leq r$, mentre $\operatorname{ann} z_i = 0$ per $i > r$. Sia $\eta \in D'$ e sia $\eta z_i = \eta(z_i) = w_i \in M$, $1 \leq i \leq s$. Per $x = \sum_{i=1}^{s} a_i z_i \in M$, $a_i \in D$, abbiamo

$$\eta x = \eta\Big(\sum a_i z_i\Big) = \sum \eta(a_i z_i) = \sum a_i(\eta z_i) = \sum a_i w_i.$$

Come già sappiamo, η è determinato dall'azione sui generatori z_i. Inoltre $d_i w_i = d_i(\eta z_i) = \eta(d_i z_i) = 0$. Dunque $\operatorname{ann} w_i \supseteq \operatorname{ann} z_i$. Se $\operatorname{ann} w_i = (g_i)$, questa condizione non impone alcun vincolo su g_i per $i > r$, mentre richiede $g_i \mid d_i$ per $i \le r$.

Viceversa, si scelga per ogni i un elemento $w_i \in M$ con $\operatorname{ann} w_i \supseteq \operatorname{ann} z_i$. Se $x = \sum a_i z_i = \sum b_i z_i$ sono due espressioni dello stesso elemento di M, allora $a_i - b_i \in \operatorname{ann} z_i \subseteq \operatorname{ann} w_i$, quindi $\sum a_i w_i = \sum b_i w_i$. La regola $\eta\colon \sum a_i z_i \mapsto \sum a_i w_i$ definisce dunque un'applicazione da M a M. Si verifica direttamente che $\eta \in D'$.

Abbiamo stabilito una biiezione $\eta \mapsto (w_1, \dots, w_s)$ fra l'anello $D' = \operatorname{Hom}(M, M)$ e l'insieme delle s-uple di elementi di M che soddisfano $\operatorname{ann} w_i \supseteq \operatorname{ann} z_i$. Scriviamo ora $w_i = \sum_j b_{ij} z_j$, $b_{ij} \in D$, e associamo alla s-upla ordinata $(w_1, \dots, w_s)$ la matrice

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1s} \\ b_{21} & b_{22} & \cdots & b_{2s} \\ \vdots & \vdots & & \vdots \\ b_{s1} & b_{s2} & \cdots & b_{ss} \end{pmatrix} \in M_s(D). \tag{45}$$

Questa matrice può non essere univocamente determinata, perché, per $j \le r$, ogni b_{ij} può essere sostituito da un $b'_{ij} \equiv b_{ij} \pmod{d_j}$. Questa è l'unica modifica che lascia invariati i w_i. La condizione $\operatorname{ann} w_i \supseteq \operatorname{ann} z_i$ equivale a

$$d_i b_{ij} \equiv 0 \pmod{d_j}. \tag{46}$$

Ciò significa che esistono $c_{ij} \in D$ tali che $d_i b_{ij} = c_{ij} d_j$. Dunque (46) equivale alla condizione che per la matrice B della (45) esista una $C \in M_s(D)$ tale che

$$\operatorname{diag}\{d_1, \dots, d_s\} B = C \operatorname{diag}\{d_1, \dots, d_s\}. \tag{47}$$

L'insieme R delle matrici B che soddisfano (47) è un sottoanello di $M_s(D)$. Ogni $B \in R$ determina un $\eta \in D'$ tale che $\eta z_i = \sum_j b_{ij} z_j$. Come nel caso particolare del modulo libero trattato nel § 3.4, si verifica facilmente che $B \mapsto \eta$ è un antiomomorfismo suriettivo di R su D' (ossia si ottiene un epimorfismo usando le matrici trasposte). Si ha $\eta = 0$ se e solo se $b_{ij} \equiv 0 \pmod{d_j}$ per $B = (b_{ij})$. Il nucleo K è pertanto l'insieme delle matrici della forma

$$B = Q \operatorname{diag}\{d_1, \dots, d_s\}, \qquad Q \in M_s(D). \tag{48}$$

Osserviamo che tali matrici soddisfano automaticamente la (47). Ne risulta il seguente teorema.

Teorema 3.15. *Sia $M = Dz_1 \oplus \cdots \oplus Dz_s$, con ideali d'ordine $\operatorname{ann} z_i = (d_i)$ che soddisfano $\operatorname{ann} z_1 \supseteq \cdots \supseteq \operatorname{ann} z_s$. Allora l'anello D' degli endomorfismi del D-modulo M è anti-isomorfo a R/K, dove R è l'anello delle matrici $B \in M_s(D)$ per le quali esiste $C \in M_s(D)$ tale che $\operatorname{diag}\{d_1, \dots, d_s\} B = C \operatorname{diag}\{d_1, \dots, d_s\}$, e K è l'ideale delle matrici della forma $Q \operatorname{diag}\{d_1, \dots, d_s\}$, $Q \in M_s(D)$.*

Se M è libero tutti i d_i sono nulli, dunque $R = M_s(D)$ e $K = 0$. Si ritrova il risultato del § 3.4. Se $s = 1$, ossia se M è ciclico, la condizione per $B = (b)$ è automaticamente soddisfatta per la commutatività di D. Allora $D' \simeq D/(d)$, con $d = d_1$.

Possiamo determinare R in modo più esplicito usando le condizioni $d_i \mid d_j$ per $i \le j \le r$ e $d_i = 0$ per $i > r$. Le condizioni (46) danno:

1. b_{ij} è arbitrario per $i \ge j$, perché in questo caso $d_i \equiv 0 \pmod{d_j}$;
2. $b_{ij} = 0$ per $i \le r$ e $j > r$, perché in questo caso $d_i \ne 0$ e $d_j = 0$;

3. b_{ij} è arbitrario per $i, j > r$, perché $d_i = d_j = 0$;

4. $b_{ij} \equiv 0 \pmod{d_i^{-1} d_j}$ per $i < j \leq r$.

Cambiando leggermente notazione, vediamo che B ha la forma

$$\left(\begin{array}{cccc|ccc} b_{11} & b_{12}d_1^{-1}d_2 & \cdots & b_{1r}d_1^{-1}d_r & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & b_{2r}d_2^{-1}d_r & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rr} & 0 & \cdots & 0 \\ \hline b_{r+1,1} & b_{r+1,2} & \cdots & b_{r+1,r} & b_{r+1,r+1} & \cdots & b_{r+1,s} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ b_{s1} & b_{s2} & \cdots & b_{sr} & b_{s,r+1} & \cdots & b_{ss} \end{array}\right). \tag{49}$$

L'angolo superiore destro è nullo, tutti i b_{ij} indicati sono arbitrari e l'elemento (i, j) per $i < j \leq r$ è $b_{ij}d_i^{-1}d_j$. Affinché la matrice appartenga a K occorre e basta che $b_{ij} = 0$ per $j > r$, che b_{ij} sia divisibile per d_j se $i \geq j$ e $j \leq r$, e che b_{ij} sia divisibile per d_i se $i < j \leq r$. Se il modulo è di torsione si ha $r = s$ e (49) si riduce al blocco superiore sinistro.

Specializziamo ora a $M = V$, dove V è il modulo su $F[\lambda]$ determinato da una trasformazione lineare T di uno spazio vettoriale V di dimensione finita su F. Questo è un modulo di torsione. Per $i \geq j$ possiamo sostituire ogni b_{ij} con un b'_{ij} della stessa classe modulo d_j. Possiamo dunque supporre $\deg b_{ij} < n_j = \deg d_j$ per $i \geq j$. Analogamente possiamo supporre $\deg b_{ij} < n_i$ per $i < j$. Diremo *normalizzate* le matrici $B \in R$ che soddisfano queste condizioni. La corrispondenza $B \mapsto \eta$, ristretta alle matrici normalizzate, è una biiezione su D'. Vi è un modo naturale di considerare D' e R come spazi vettoriali su F. In R basta moltiplicare tutte le entrate di $B \in R$ per lo scalare $a \in F$. Per D' definiamo $a\eta$, $a \in F$, $\eta \in D'$, mediante $(a\eta)x = a(\eta x) = \eta(ax)$ (cfr. esercizio 5 del § 3.4). Rispetto a queste strutture, l'insieme S delle matrici normalizzate di R è chiaramente un sottospazio e $B \mapsto \eta$ è un isomorfismo F-lineare di S su D'. Vogliamo calcolare la dimensione di D' su F, ossia, in termini matriciali, la dimensione dello spazio delle matrici che commutano con una matrice data. L'isomorfismo appena stabilito ci permette di calcolare a tal fine $\dim S$. Sia S_{ij}, $1 \leq i, j \leq s$, il sottospazio delle matrici normalizzate le cui entrate sono nulle salvo eventualmente quella al posto (i, j). Allora $\dim S_{ij} = n_j$ per $i \geq j$, e $\dim S_{ij} = n_i$ per $i < j$. Poiché S è somma diretta dei sottospazi S_{ij}, si ottiene

$$\dim S = \sum_{j=1}^{s} (s - j + 1)n_j + \sum_{i=1}^{s-1} (s - i)n_i = \sum_{j=1}^{s} (2s - 2j + 1)n_j.$$

Il risultato si enuncia in termini di matrici nella forma seguente.

Teorema 3.16 (Frobenius). *Sia $A \in M_n(F)$, con F campo. Siano $d_1(\lambda), \ldots, d_s(\lambda)$ i fattori invarianti diversi da 1 di $\lambda 1 - A$ e si ponga $n_i = \deg d_i(\lambda)$. La dimensione dello spazio vettoriale su F delle matrici che commutano con A è*

$$N = \sum_{j=1}^{s} (2s - 2j + 1)n_j. \tag{50}$$

Naturalmente il teorema si può enunciare anche in termini di trasformazioni lineari. In questa forma dà il seguente corollario.

Corollario. *Una trasformazione lineare T è ciclica, cioè è ciclico il corrispondente $F[\lambda]$-modulo, se e solo se le sole trasformazioni lineari che commutano con T sono i polinomi in T.*

Dimostrazione. T è ciclica se e solo se $s = 1$. Sappiamo inoltre che $d_s(\lambda)$ è il polinomio minimo $m(\lambda)$ di T. Pertanto n_s è la dimensione su F dell'anello $F[T]$ dei polinomi in T a coefficienti in F (si veda l'esercizio 1 a p. 133 dell'originale). Se $s = 1$, la (50) dà $N = n_1 = \dim F[T]$. Lo spazio delle trasformazioni lineari che commutano con T, che certamente contiene $F[T]$, coincide dunque con $F[T]$. Se $s > 1$, la (50) dà $N > \sum_{j=1}^{s} n_j > n_s$. Esistono quindi trasformazioni lineari che commutano con T e non sono polinomi in T. □

Esempio. Sia $F = \mathbb{Q}$ e sia

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Se T è la corrispondente trasformazione lineare e si considera lo spazio vettoriale quale $\mathbb{Q}[\lambda]$-modulo tramite T, si ha $V = \mathbb{Q}[\lambda]f_1 \oplus \mathbb{Q}[\lambda]f_2$. I fattori invarianti sono $\lambda - 1$ e $(\lambda - 1)^2$. Le matrici normalizzate di R hanno la forma

$$\begin{pmatrix} b_{11} & b_{12}(\lambda - 1) \\ b_{21} & b_{22} + b'_{22}\lambda \end{pmatrix}, \qquad b_{ij}, b'_{22} \in \mathbb{Q}. \tag{51}$$

Poiché $\lambda f_1 = f_1$ e $\lambda^2 f_2 = (2\lambda - 1)f_2 = -f_2 + 2\lambda f_2$, la trasformazione U corrispondente alla (51) soddisfa

$$\begin{aligned} Uf_1 &= b_{11}f_1 - b_{12}f_2 + b_{12}(\lambda f_2), \\ Uf_2 &= b_{21}f_1 + b_{22}f_2 + b'_{22}(\lambda f_2), \\ U(\lambda f_2) &= b_{21}f_1 - b'_{22}f_2 + (b_{22} + 2b'_{22})(\lambda f_2). \end{aligned}$$

Pertanto la forma generale di una matrice che commuta con A è

$$\begin{pmatrix} b_{11} & -b_{12} & b_{12} \\ b_{21} & b_{22} & b'_{22} \\ b_{21} & -b'_{22} & b_{22} + 2b'_{22} \end{pmatrix}.$$

Ritorniamo al caso generale di un D-modulo finitamente generato M su un dominio a ideali principali D, con $M = Dz_1 \oplus \cdots \oplus Dz_s$ come sopra. Poiché D è commutativo, l'anello D' degli endomorfismi di M contiene tutte le applicazioni $x \mapsto ax$, $a \in D$. Esse appartengono evidentemente al centro di D'. Dimostreremo ora il seguente teorema.

Teorema 3.17. *Il centro di $D' = \mathrm{Hom}(M, M)$ è l'insieme delle applicazioni $x \mapsto ax$, $a \in D$.*

Dimostrazione. La descrizione ottenuta di D' mostra che per ogni i, $1 \le i \le s$, esiste un endomorfismo ε_{is} tale che $\varepsilon_{is}z_s = z_i$ e $\varepsilon_{is}z_j = 0$ per $j \neq s$. Sia γ nel centro di D' e scriviamo $\gamma z_s = \sum a_j z_j$, con $a_j \in D$. Allora

$$\gamma z_s = \gamma\varepsilon_{ss}z_s = \varepsilon_{ss}\gamma z_s = \varepsilon_{ss}\Big(\sum_j a_j z_j\Big) = \sum_j a_j \varepsilon_{ss} z_j = a_s z_s.$$

Inoltre $\gamma z_i = \gamma\varepsilon_{is}z_s = \varepsilon_{is}\gamma z_s = \varepsilon_{is}(\sum_j a_j z_j) = \sum_j a_j \varepsilon_{is} z_j = a_s z_i$. Ne segue che γ è l'applicazione $x \mapsto a_s x$. □

Specializzando il Teorema 3.17 al modulo determinato da una trasformazione lineare, otteniamo il seguente risultato.

Corollario 1. *Se U è una trasformazione lineare di uno spazio vettoriale di dimensione finita che commuta con ogni trasformazione lineare che commuta con una trasformazione lineare data T, allora U è un polinomio in T.*

Ponendo $T = 1$ si ottiene l'immediata conseguenza seguente.

Corollario 2. *Il centro dell'anello delle trasformazioni lineari di uno spazio vettoriale di dimensione finita su un campo F è l'insieme delle moltiplicazioni scalari $x \mapsto ax$, $a \in F$.*

Esercizi.

1. Sia G un gruppo abeliano finito, somma diretta di gruppi ciclici di ordini $n_1, \ldots, n_s$, con $n_i \mid n_j$ per $i \leq j$. Mostrare che il numero degli endomorfismi di G è $N = \prod_{j=1}^{s} n_j^{2s-2j+1}$.

2. Determinare le matrici di $M_5(\mathbb{Q})$ che commutano con $\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$.

3. Determinare le matrici di $M_4(\mathbb{Q})$ che commutano con $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -3 & 3 \end{pmatrix}$.

4. Dimostrare che una trasformazione lineare T di uno spazio vettoriale di dimensione finita su un campo è ciclica se e solo se l'anello delle trasformazioni lineari che commutano con T è commutativo.

5. Dimostrare la seguente estensione del Teorema 3.17. I soli endomorfismi di M che commutano con ogni elemento idempotente di D' sono le applicazioni $x \mapsto ax$, $a \in D$.

Capitolo 4
Teoria di Galois delle equazioni

Questo capitolo ha per obiettivo principale lo studio di due problemi classici: la risolubilità per radicali delle equazioni polinomiali e le costruzioni con riga e compasso. Diamo anzitutto qualche cenno alla loro storia.

In algebra elementare si ricava la formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

per risolvere l'equazione di secondo grado $ax^2 + bx + c = 0$. Nella sostanza essa era già nota ai Babilonesi. Durante il Rinascimento italiano si dedicarono notevoli sforzi alla sua generalizzazione alle equazioni di grado superiore. Se ne ebbe uno dei grandi successi della matematica rinascimentale: le formule per le radici delle equazioni di terzo e quarto grado. La prima si deve a Scipione del Ferro, professore all'Università di Bologna dal 1496 al 1526. Ignoriamo la data esatta della scoperta, ma sappiamo che, prima del 1541, Niccolò Tartaglia, forse al corrente dell'esistenza della soluzione di del Ferro, riuscì a scoprirla da sé. La soluzione di Tartaglia fu pubblicata da Girolamo Cardano nell'*Ars Magna* (1545) ed è generalmente conosciuta con il nome di «formule di Cardano» per la soluzione delle equazioni cubiche.[1] È facile, per noi, vedere che la soluzione dell'equazione cubica $x^3 + ax^2 + bx + c = 0$ si riconduce al caso «ridotto» $x^3 + px + q = 0$, sostituendo a x l'espressione $x - \frac{1}{3}a$. Indichiamo con x_1, x_2, x_3 le radici dell'equazione ridotta e poniamo

$$\delta = -4p^3 - 27q^2, \qquad \zeta = -\tfrac{1}{2} - \tfrac{1}{2}\sqrt{-3},$$
$$y_1 = x_1 + \zeta^2 x_2 + \zeta x_3, \qquad y_2 = x_1 + \zeta x_2 + \zeta^2 x_3.$$

Le formule di Cardano sono allora

$$y_1 = \sqrt[3]{-\tfrac{27}{2}q + \tfrac{3}{2}\sqrt{-3\delta}}, \qquad y_2 = \sqrt[3]{-\tfrac{27}{2}q - \tfrac{3}{2}\sqrt{-3\delta}},$$

per opportune determinazioni delle radici cubiche (si vedano le pp. 264–266 dell'originale). Dalla forma dell'equazione ridotta segue $x_1 + x_2 + x_3 = 0$. Per determinare le radici x_i basta pertanto risolvere le tre equazioni lineari

$$x_1 + x_2 + x_3 = 0, \qquad x_1 + \zeta^2 x_2 + \zeta x_3 = y_1, \qquad x_1 + \zeta x_2 + \zeta^2 x_3 = y_2.$$

Un metodo generale per risolvere le equazioni di quarto grado, anch'esso pubblicato da Cardano nell'*Ars Magna*, è attribuito al suo assistente Ludovico Ferrari. Lo indicheremo più avanti. Osserviamo qui soltanto che, come per le cubiche, le soluzioni si esprimono mediante estrazioni di radici e operazioni razionali eseguite sui coefficienti dell'equazione data.

Dalla metà del Cinquecento all'inizio dell'Ottocento alcuni dei massimi matematici del tempo, fra cui Eulero e Lagrange, tentarono di ottenere risultati analoghi per le equazioni di quinto grado. Lagrange andò ben oltre gli altri aspiranti risolutori della quintica: analizzò con grande acutezza le soluzioni già note delle cubiche e delle quartiche e mostrò che esse erano risolubili per radicali perché si poteva ricondurne la soluzione a quella di equazioni «risolventi» di grado inferiore. Trovò però che, applicato a una quintica, lo stesso metodo conduceva a una risolvente

[1] Per una discussione più ampia della storia della teoria delle equazioni algebriche rimandiamo a C. A. Boyer, *A History of Mathematics*, New York, Wiley, 1968, oppure a E. T. Bell, *The Development of Mathematics*, New York, McGraw-Hill, 1940.

di sesto grado. Era un forte indizio che le equazioni di grado superiore al quarto non fossero, in generale, risolubili per radicali. Eppure, la scoperta che così stavano effettivamente le cose destò grande sorpresa. Vi giunsero indipendentemente Ruffini (pubblicazione del 1813) e N. H. Abel (pubblicazione del 1827). Il loro risultato, solitamente attribuito ad Abel, afferma che l'equazione «generale» di grado n superiore a quattro, ossia l'equazione $x^n + t_1x^{n-1} + \cdots + t_n = 0$ con coefficienti indeterminati t_i, non è risolubile per radicali. Le dimostrazioni di Ruffini e di Abel sono piuttosto oscure e forse incomplete in qualche particolare. Per noi hanno un interesse soltanto storico: furono presto superate dal coronamento di questa linea di ricerca, le scoperte di Galois nella teoria delle equazioni. Galois ottenne i suoi risultati giovanissimo: morì in duello nel 1832, prima di compiere ventun anni. Il suo lavoro fornì sia una dimostrazione del teorema di Ruffini–Abel sia un criterio di risolubilità per radicali per ogni equazione $x^n + a_1x^{n-1} + \cdots = 0$, anche a coefficienti assegnati. Soprattutto, la sua scoperta principale, la corrispondenza biunivoca tra i sottocampi di un certo tipo di estensione e i sottogruppi di un gruppo finito, il *gruppo di Galois*, è divenuta un risultato centrale per tutta l'algebra, di importanza ben superiore al problema da cui ebbe origine.[2] Nel secolo successivo alla pubblicazione della teoria, avvenuta nel 1846, quasi quindici anni dopo la sua morte, essa fu notevolmente semplificata e affinata, soprattutto con l'introduzione di idee più astratte.

Il secondo problema sul quale concentreremo l'attenzione ha origine nella matematica greca. I Greci non riuscirono a decidere se certe costruzioni geometriche fossero possibili con il solo impiego di riga non graduata e compasso. Fra le più note ricordiamo: (1) la trisezione di un angolo arbitrario; (2) la duplicazione del cubo, cioè la costruzione dello spigolo di un cubo avente volume doppio di quello di un cubo dato; (3) la costruzione dell'ettagono regolare, il poligono regolare di sette lati; (4) la quadratura del cerchio, cioè la costruzione di un quadrato della stessa area di un cerchio dato. Ogni problema di costruzione con riga e compasso si può formulare come un problema algebrico sui campi. Fatto ciò, è facile vedere che i primi tre problemi ammettono risposta negativa: basta applicare la formula fondamentale delle dimensioni per i campi (teorema 4.2). L'impossibilità della quadratura del cerchio segue dal fatto, dimostrato per la prima volta da F. Lindemann nel 1882, che π è trascendente, cioè non è algebrico su $\mathbb{Q}$. Gauss risolse nelle *Disquisitiones Arithmeticae* (1801) il problema generale di determinare gli interi n per i quali è costruibile con riga e compasso il poligono regolare di n lati. Segue dai suoi risultati che la costruzione è possibile per $n = 17$, 257 o 65537. La prima scoperta matematica di Gauss di cui si abbia notizia fu un metodo per costruire il poligono regolare di 17 lati, problema rimasto aperto dal tempo dei Greci per circa duemila anni. Gauss giunse a questi risultati attraverso calcoli elementari ma alquanto lunghi sulle radici dell'unità. Vedremo che la teoria di Galois permette di ottenerli rapidamente, senza calcoli.

Oltre a questi risultati, ci occuperemo in questo capitolo di alcune conseguenze della teoria di Galois, fra cui lo studio dei campi finiti, anch'esso iniziato da Galois. Introdurremo altre nozioni fondamentali della teoria dei campi: norme, tracce, elementi primitivi e basi normali.

4.1 Risultati preliminari, alcuni già noti, altri nuovi

Abbiamo definito l'anello fondamentale di un anello R come il suo più piccolo sottoanello e abbiamo visto che esso coincide con l'insieme $\mathbb{Z}1$ dei multipli interi di 1 (§ 2.7). Inoltre $\mathbb{Z}1 \cong \mathbb{Z}$ oppure $\mathbb{Z}1 \cong \mathbb{Z}/(k)$, con $k \neq 0$. Nel primo caso R ha caratteristica zero, nel secondo ha

[2] Uno dei più grandi matematici di questo secolo, Hermann Weyl, ha così valutato il contributo di Galois nel suo libro *Symmetry*, Princeton University Press, 1952, p. 138: «Le idee di Galois, rimaste per parecchi decenni un libro chiuso da sette sigilli, ma destinate a esercitare poi un'influenza sempre più profonda sull'intero sviluppo della matematica, sono contenute in una lettera d'addio scritta a un amico alla vigilia della morte, cui andò incontro in un insensato duello all'età di ventun anni. A giudicarla dalla novità e dalla profondità delle idee che contiene, questa lettera è forse lo scritto più denso di sostanza di tutta la letteratura dell'umanità».

caratteristica k. Se R è un dominio, k è un numero primo p. Supponiamo ora che $R = F$ sia un campo. Definiamo il *sottocampo fondamentale* di F come il più piccolo sottocampo di F. Se F ha caratteristica $p \neq 0$, il sottoanello fondamentale è un campo, poiché è isomorfo a $\mathbb{Z}/(p)$; in questo caso, dunque, il sottoanello e il sottocampo fondamentali di F coincidono. Se F ha caratteristica zero, il monomorfismo $m \mapsto m1$ di $\mathbb{Z}$ nel sottocampo fondamentale si estende a un monomorfismo di $\mathbb{Q}$ nel medesimo sottocampo. Quest'ultimo è pertanto isomorfo a $\mathbb{Q}$. In questo senso possiamo dire che ogni campo contiene l'anello $\mathbb{Z}/(p)$ per un certo primo p (la caratteristica del campo), oppure il campo $\mathbb{Q}$ dei numeri razionali.

Sia E un'estensione del campo F, ossia un campo che contiene F come sottocampo. Per ogni sottoinsieme S di E, ricordiamo che $F[S]$ denota il sottoanello di E generato da F e da S, o, come diremo d'ora innanzi, il *sottoanello di E/F generato da S*. Useremo la notazione $F(S)$ per il *sottocampo di E/F generato da S*, intendendo naturalmente il sottocampo di E generato da F e da S. Come per gli anelli, se T è un secondo sottoinsieme di E, si ha subito $F(S)(T) = F(S \cup T)$ (§ 2.10). Se $u \in E$, scriviamo $F(u)$ per $F(\{u\})$; più in generale, per un insieme finito poniamo $F(u_1, \ldots, u_n) = F(\{u_1, \ldots, u_n\})$. Come si presenta $F(u)$? Ricordiamo anzitutto che, se x è un'indeterminata, abbiamo l'omomorfismo $g(x) \mapsto g(u)$ dell'anello dei polinomi $F[x]$ in E che si riduce all'identità su F e manda x in u (teorema 2.10, p. 122 dell'originale). Se il nucleo è zero, $F[u] \cong F[x]$. Altrimenti esiste un polinomio monico $f(x)$ di grado positivo tale che il nucleo è l'ideale principale $(f(x))$, e allora $F[u] \cong F[x]/(f(x))$. Il polinomio $f(x)$ è primo, altrimenti $F[x]/(f(x))$ non sarebbe un dominio. Perciò $F[x]/(f(x))$ è un campo (teorema 2.16, p. 131) ed è chiaro che $F(u) = F[u]$. Nell'altro caso, $F[x] \cong F[u]$, l'omomorfismo $g(x) \mapsto g(u)$ è un monomorfismo e si estende univocamente a un monomorfismo del campo delle frazioni $F(x)$ di $F[x]$. Allora $F(u) \supsetneq F[u]$, $F(u) \cong F(x)$ e gli elementi di $F(u)$ sono i quozienti $g(u)h(u)^{-1}$, con $g(x), h(x) \in F[x]$ e $h(x) \neq 0$. In questo caso u è trascendente. Se invece $F(u) \cong F[x]/(f(x))$, con $f(x)$ di grado positivo, u è algebrico e, quando $f(x)$ è monico, esso è il polinomio minimo dell'elemento u. In ogni caso, se $E = F(u)$, diciamo che E è un'*estensione semplice* di F e che u è un *elemento primitivo*, ossia un generatore di E/F come campo.

Nello studio di un'estensione E rispetto a un sottocampo F è utile considerare E come spazio vettoriale (o modulo) su F. Il gruppo abeliano sottostante è il gruppo additivo di E, e l'operazione di modulo ay, con $a \in F$ e $y \in E$, è il prodotto in E. Le estensioni che incontreremo più spesso in questo capitolo hanno dimensione finita sul campo base F. Indichiamo la dimensione con $[E : F]$. Mostreremo anzitutto che $u \in E$ è algebrico su F se e solo se $[F(u) : F] < \infty$ e che in tal caso $[F(u) : F]$ è il grado del polinomio minimo di u su F. Chiameremo questo numero *grado di u su F*. Sia u algebrico e sia $f(x) = x^n + a_1 x^{n-1} + \cdots + a_n \in F[x]$ il suo polinomio minimo. Abbiamo $F(u) = F[u]$. Per ogni $g(x) \in F[x]$, scriviamo $g(x) = f(x)q(x) + r(x)$ con $\deg r(x) < \deg f(x) = n$. Allora $g(u) = 0q(u) + r(u)$, sicché ogni elemento di $F[u]$ è della forma

$$r(u) = b_0 + b_1 u + \cdots + b_{n-1} u^{n-1}, \qquad b_i \in F.$$

Poiché $f(x)$ ha il minimo grado possibile fra i polinomi non nulli che si annullano in u, una relazione $b_0 + b_1 u + \cdots + b_{n-1} u^{n-1} = 0$, con $b_i \in F$, vale solamente quando tutti i b_i sono nulli. Dunque $(1, u, \ldots, u^{n-1})$ è una base di $F(u)/F$. La dimensione dell'estensione è perciò $n = \deg f(x)$. Se invece u è trascendente, gli elementi $1, u, u^2, \ldots$ di $F(u)$ sono linearmente indipendenti su F, e quindi $F(u)$ non ha dimensione finita su F. Riuniamo parte dei risultati nell'enunciato seguente.

Teorema 4.1. *Sia u un elemento di un'estensione E del campo F. Allora u è algebrico su F se e solo se $F(u)$ ha dimensione finita su F. In tal caso $F(u)$ coincide con il sottoanello $F[u]$ generato da F e da u.*

Supponiamo ora di avere un'estensione di F a due livelli, cioè $F \subset E \subset K$, dove K è un campo ed E, F sono sottocampi. Possiamo considerare K come spazio vettoriale su E e su F,

ed E come spazio vettoriale su F. Indichiamo questi spazi rispettivamente con K/E, K/F ed E/F. Fra le loro dimensioni vale la seguente importante relazione.

Teorema 4.2. *Se $K \supset E \supset F$ sono campi, $[K : F]$ è finito se e solo se sono finiti $[K : E]$ ed $[E : F]$. In tal caso vale la relazione*

$$[K : F] = [K : E][E : F]. \tag{1}$$

Dimostrazione. Se $[K : F] < \infty$, anche $[E : F]$ è finito, poiché E è un sottospazio di K/F. Se $(u_1, \dots, u_n)$ è una base di K/F, ogni elemento di K è combinazione lineare degli u_i a coefficienti in F, e a maggior ragione a coefficienti in E. Per un noto risultato di algebra lineare si può dunque estrarre da $\{u_1, \dots, u_n\}$ una base di K/E. Pertanto $[K : E] < \infty$. Viceversa, supponiamo finiti $[K : E]$ ed $[E : F]$. Siano $(v_1, \dots, v_m)$ una base di K/E e $(w_1, \dots, w_r)$ una base di E/F. Dato $z \in K$, si ha $z = \sum_{i=1}^{m} a_i v_i$ per opportuni $a_i \in E$, e $a_i = \sum_{j=1}^{r} b_{ij} w_j$ per opportuni $b_{ij} \in F$. Ne segue $z = \sum_{i,j} b_{ij} w_j v_i$: ogni elemento di K è combinazione F-lineare degli mr elementi $w_j v_i$. Supponiamo ora $\sum_{i,j} b_{ij} w_j v_i = 0$ con $b_{ij} \in F$. Allora $\sum_i a_i v_i = 0$ con $a_i = \sum_j b_{ij} w_j$. Poiché i v_i formano una base di K/E, tutti gli a_i sono nulli. Poiché i w_j formano una base di E/F, da $a_i = \sum_j b_{ij} w_j = 0$ segue che ogni b_{ij} è nullo. I prodotti $w_j v_i$ sono dunque indipendenti su F e costituiscono una base di K/F. Si ha così $[K : F] = mr = [K : E][E : F] < \infty$. □

Segue immediatamente che, se $[K : F] < \infty$, la dimensione di ogni sottocampo E/F divide la dimensione di K/F. In particolare, se $[K : F]$ è primo, gli unici sottocampi di K/F sono K e F.

Esercizi

1. Sia $E = \mathbb{Q}(u)$, con $u^3 - u^2 + u + 2 = 0$. Esprimere $(u^2 + u + 1)(u^2 - u)$ e $(u - 1)^{-1}$ nella forma $au^2 + bu + c$, con $a, b, c \in \mathbb{Q}$.

2. Determinare $[\mathbb{Q}(\sqrt{2}, \sqrt{3}) : \mathbb{Q}]$.

3. Sia p primo e sia $v \in \mathbb{C}$ tale che $v \neq 1$, $v^p = 1$ (per esempio $v = \cos(2\pi/p) + i \sin(2\pi/p)$). Dimostrare che $[\mathbb{Q}(v) : \mathbb{Q}] = p-1$. (*Suggerimento.* Usare l'esercizio 3, p. 154 dell'originale.)

4. Sia $w = \cos(\pi/6) + i \sin(\pi/6) \in \mathbb{C}$. Si osservi che $w^{12} = 1$, ma $w^r \neq 1$ per $1 \leq r < 12$: w genera quindi il gruppo ciclico delle radici dodicesime di 1. Dimostrare che $[\mathbb{Q}(w) : \mathbb{Q}] = 4$ e determinare il polinomio minimo di w su $\mathbb{Q}$.

5. Sia $E = F(u)$, dove u è algebrico di grado dispari (cioè di polinomio minimo di grado dispari). Dimostrare che $E = F(u^2)$.

6. Siano E_i, $i = 1, 2$, sottocampi di K/F tali che $[E_i : F]$ sia finito. Dimostrare che, se E è il sottocampo di K generato da E_1 ed E_2, allora $[E : F] \leq [E_1 : F][E_2 : F]$.

7. Sia E un'estensione algebrica di F, nel senso che ogni elemento di E è algebrico su F. Dimostrare che ogni sottoanello di E/F è un sottocampo. Dedurne che ogni sottoanello di un'estensione E/F di dimensione finita è un sottocampo.

8. Sia $E = F(u)$ con u trascendente e sia $K \neq F$ un sottocampo di E/F. Dimostrare che u è algebrico su K.

9. Sia E un'estensione di F con le proprietà seguenti: (i) $[E : F] < \infty$; (ii) per ogni coppia di sottocampi E_1, E_2 contenenti F, o $E_1 \supset E_2$ oppure $E_2 \supset E_1$. Dimostrare che E ha un elemento primitivo su F.

4.2 Costruzioni con riga e compasso

Il problema delle costruzioni euclidee, ossia con riga e compasso, si può formulare come segue. Dato un insieme finito di punti $S = \{P_1, P_2, \ldots, P_n\}$ in un piano ω, definiamo per induzione sottoinsiemi S_m di ω, con $m = 1, 2, \ldots$. Poniamo $S_1 = S$ e otteniamo S_{r+1} come unione di S_r con i seguenti tre insiemi: (1) i punti d'intersezione delle coppie di rette che congiungono punti distinti di S_r; (2) i punti d'intersezione delle rette di (1) con tutte le circonferenze che hanno centro in S_r e raggio uguale a un segmento di estremi in S_r; (3) i punti d'intersezione delle coppie di circonferenze descritte in (2). Sia

$$C(P_1, \ldots, P_n) = \bigcup_{i=1}^{\infty} S_i.$$

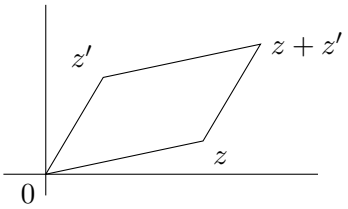
Diremo che un punto P di ω *si può costruire da* $P_1, \ldots, P_n$ *con riga e compasso* se $P \in C(P_1, \ldots, P_n)$. In caso contrario, P non è costruibile a partire dai P_i.

Qual è il rapporto con la nozione di costruibilità della geometria euclidea? I dati di una costruzione euclidea sono punti, rette, circonferenze e angoli, in numero finito per ciascun tipo. Una retta è determinata da due suoi punti; una circonferenza dal centro e da un suo punto; un angolo dal vertice e da due punti sui suoi lati, equidistanti dal vertice. Con queste sostituzioni possiamo supporre dato un insieme finito $S_1 = \{P_1, \ldots, P_n\}$ nel piano ω. I punti degli insiemi successivi $S_2, S_3, \ldots$ si ottengono certamente da S_1 con costruzioni euclidee. Osserviamo inoltre che, talvolta, nelle costruzioni euclidee si prescrive di servirsi di un punto o di una lunghezza «arbitrari», richiedendo soltanto che il punto appartenga a una certa regione o che la lunghezza soddisfi una disuguaglianza. Si tratta così di scegliere punti in certi aperti non vuoti del piano. Vedremo fra poco che, se S_1 contiene almeno due punti distinti, l'insieme $C(P_1, \ldots, P_n)$ è denso nel piano. Ogni istruzione che richieda di scegliere un punto in un aperto non vuoto si può dunque eseguire scegliendo un punto di $C(P_1, \ldots, P_n)$. La nostra definizione di punti costruibili, che ha il vantaggio di essere precisa, equivale pertanto a quanto si intendeva fare nella geometria euclidea.

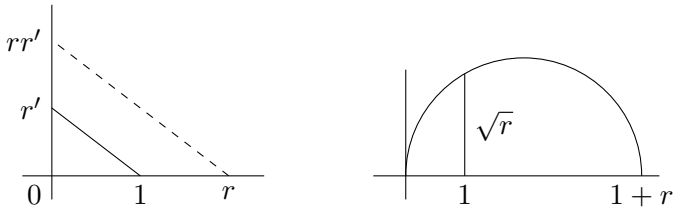
Consideriamo, per esempio, la trisezione di un angolo di 60°. Sono dati $P_1 = (0, 0)$ (il vertice), $P_2 = (1, 0)$ e $P_3 = (\cos 60°, \sin 60°) = (\frac{1}{2}, \frac{1}{2}\sqrt{3})$. Il punto $P = (\cos 20°, \sin 20°)$ appartiene a $C(P_1, P_2, P_3)$? L'angolo di 60° si può trisecare con la sola riga e il compasso se e solo se la risposta è affermativa.

Traduciamo ora la definizione in termini algebrici. Supponiamo $n \geq 2$, poiché altrimenti $C(P_1, \ldots, P_n) = \{P_1\}$. Scegliamo coordinate cartesiane in modo che $P_1 = (0, 0)$ sia l'origine e $P_2 = (1, 0)$. Al punto $P = (x, y)$ associamo il numero complesso $x + iy$. Identifichiamo così il piano con il campo $\mathbb{C}$ e l'insieme dato con un insieme di complessi $\{z_1, \ldots, z_n\}$ dove $z_1 = 0$, $z_2 = 1$. Qual è l'insieme $C(z_1, \ldots, z_n)$ dei numeri complessi corrispondenti ai punti di $C(P_1, \ldots, P_n)$? È naturale chiamarli *numeri complessi costruibili da* $z_1, \ldots, z_n$ *con riga e compasso*. Dimostreremo la seguente caratterizzazione: $C(z_1, \ldots, z_n)$ è il più piccolo sottocampo di $\mathbb{C}$ che contiene gli z_i ed è chiuso rispetto all'estrazione di radici quadrate e al coniugio. Esso contiene cioè ogni z tale che z^2 appartenga all'insieme e, con $z = x + iy$, x, y reali, contiene anche $\bar{z} = x - iy$. Come di consueto, «più piccolo» significa che $C(z_1, \ldots, z_n)$ ha le proprietà indicate ed è contenuto in ogni sottoinsieme di $\mathbb{C}$ che le possieda.

Siano $z, z' \in C(z_1, \ldots, z_n)$. Si costruisce $z + z'$ con l'usuale regola del parallelogramma per la somma di vettori:



Delle due intersezioni della circonferenza di centro z e raggio $|z'|$ (la lunghezza di $0z'$) con quella di centro z' e raggio $|z|$, quella indicata in figura è $z+z'$. È anche chiaro che $-z \in C(z_1, \ldots, z_n)$. Questo insieme è dunque un sottogruppo del gruppo additivo di $\mathbb{C}$. Per provarne la chiusura rispetto a moltiplicazione, inversione ed estrazione di radice quadrata, scriviamo z in forma polare, $z = re^{i\theta}$, dove il modulo è $r = (x^2 + y^2)^{1/2}$ se $z = x + iy$, e l'argomento θ è l'angolo dall'asse x a $0z$. Se $z' = r'e^{i\theta'}$, allora $zz' = rr'e^{i(\theta+\theta')}$ ha per modulo il prodotto rr' dei moduli e per argomento la somma degli argomenti. È facile costruire la semiretta di argomento $\theta + \theta'$. La figura a sinistra indica come costruire rr'.



La retta tratteggiata, parallela alla retta per 1 e r', si costruisce come le parallele della prima figura. Invertendo la costruzione, con r e r' sull'asse y, si ottiene sull'asse x il punto r/r'. È quindi costruibile $z(z')^{-1}$ se $z' \neq 0$. È noto, e si vede facilmente, che con riga e compasso si può bisecare ogni angolo. La figura a destra mostra come costruire $\sqrt{r}$. Ne segue che, se z è costruibile, anche $z^{1/2}$ lo è. Lo stesso vale per $\bar{z}$: basta condurre da z la perpendicolare all'asse x (la retta P_1P_2) e prendere il punto simmetrico di z rispetto a quest'asse. È così dimostrato che $C(z_1, \ldots, z_n)$ è un sottocampo di $\mathbb{C}$ chiuso rispetto alle radici quadrate e al coniugio.

Sia ora C' un sottocampo qualsiasi di $\mathbb{C}$ che contiene gli z_i, $1 \leq i \leq n$, ed è chiuso rispetto alle radici quadrate e al coniugio. Per la definizione induttiva $C(z_1, \ldots, z_n) = \bigcup_i S_i$, per provare $C' \supset C(z_1, \ldots, z_n)$ basta mostrare che appartengono ancora a C' i punti comuni a due rette determinate da punti di C', a una tale retta e a una circonferenza con centro in C' e raggio uguale a un segmento di estremi in C', oppure a due circonferenze di questo tipo. Per prima cosa, C' è chiuso rispetto al coniugio e contiene $i = \sqrt{-1}$: se $z = x + iy \in C'$ con x, y reali, anche $x, y \in C'$. Perciò ogni retta per due punti distinti di C' ha un'equazione $ax + by + c = 0$ con a, b, c reali in C'. Ogni circonferenza con centro e raggio come sopra ha un'equazione $x^2 + y^2 + dx + ey + f = 0$ con d, e, f reali in C'. Per la regola di Cramer le coordinate dell'intersezione di due rette non parallele $ax + by + c = 0$, $a'x + b'y + c' = 0$ sono quozienti di determinanti formati dai coefficienti a, b, c, a', b', c': sono pertanto numeri reali di C'. Le ascisse delle intersezioni di $y = mx + b$ con $x^2 + y^2 + dx + ey + f = 0$ si ottengono risolvendo

$$x^2 + (mx + b)^2 + dx + e(mx + b) + f = 0.$$

La formula dell'equazione di secondo grado mostra che le soluzioni sono reali e appartengono a C' se i coefficienti m, b, d, e, f sono reali in C' e la retta incontra la circonferenza. Si tratta analogamente una retta $x = a$. Infine, i punti comuni alle circonferenze $x^2 + y^2 + dx + ey + f = 0$ e $x^2 + y^2 + d'x + e'y + f' = 0$ sono i punti comuni alla prima circonferenza e alla retta $(d - d')x + (e - e')y + f - f' = 0$. Dunque le intersezioni di rette e circonferenze con coefficienti reali in C' hanno coordinate (p, q) esprimibili mediante operazioni razionali ed estrazioni di radici

quadrate sui coefficienti; così $p + qi \in C'$. Ciò dimostra la nostra affermazione sulla minimalità di $C(z_1, \ldots, z_n)$.

Si osservi che $C(z_1, \ldots, z_n)$ contiene tutti i complessi $p + iq$ con p e q razionali. Questo sottoinsieme è denso in $\mathbb{C}$: ogni disco aperto contiene almeno un suo punto. Dalla caratterizzazione appena ottenuta si ricava il seguente criterio.

Criterio A. *Siano $z_1, \ldots, z_n \in \mathbb{C}$ e sia $F = \mathbb{Q}(z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n)$. Un numero complesso z è costruibile da $z_1, \ldots, z_n$ se e solo se è contenuto in un sottocampo di $\mathbb{C}$ della forma $F(u_1, \ldots, u_r)$, dove $u_1^2 \in F$ e, per ogni $i > 1$, $u_i^2 \in F(u_1, \ldots, u_{i-1})$.*

Un campo $F(u_1, \ldots, u_r)$ con queste proprietà si chiamerà *torre di radici quadrate su F*.

Dimostrazione del criterio. Poiché $C(z_1, \ldots, z_n)$ è chiuso rispetto alle radici quadrate e al coniugio, contiene F e ogni torre di radici quadrate su F. Pertanto contiene l'insieme C' dei numeri complessi che soddisfano la condizione indicata. Siano $z, z' \in C'$, con z nella torre $F(u_1, \ldots, u_r)$ e z' nella torre $F(u'_1, \ldots, u'_s)$. Allora $z + z'$, zz' e, se $z \neq 0$, z^{-1} appartengono alla torre $F(u_1, \ldots, u_r, u'_1, \ldots, u'_s)$. Quindi C' è un sottocampo di $\mathbb{C}$. Evidentemente è chiuso rispetto alle radici quadrate e, poiché $\bar{F} = F$, da

$$\overline{F(u_1, \ldots, u_r)} = F(\bar{u}_1, \ldots, \bar{u}_r)$$

segue che è chiuso anche rispetto al coniugio. Si ha così $C' \supset C(z_1, \ldots, z_n)$; vale dunque l'uguaglianza che dimostra il criterio. □

Per le applicazioni che seguono sarà sufficiente una semplice conseguenza del criterio.

Corollario. *Sia $F = \mathbb{Q}(z_1, \ldots, z_n, \bar{z}_1, \ldots, \bar{z}_n)$. Ogni numero complesso z costruibile da $z_1, \ldots, z_n$ è algebrico su F e il suo grado è una potenza di due.*

Dimostrazione. Se $L = K(u)$ con $u^2 = a \in K$, si ha immediatamente $L = K$ oppure $[L : K] = 2$. Applicando ripetutamente il teorema 4.2, vediamo che ogni torre di radici quadrate su F ha dimensione su F uguale a una potenza di due. Se z appartiene alla torre, ancora per il teorema 4.2 si ha $[F(z) : F] = 2^s$ per un certo $s \geq 0$. □

In molti problemi si assegnano soltanto due punti o, equivalentemente, un segmento. Con una scelta opportuna delle coordinate essi si possono prendere uguali a 0 e a 1. Allora $F = \mathbb{Q}$ e chiameremo $C = C(z_1, z_2)$ il *campo dei numeri complessi costruibili in senso euclideo*. Il corollario mostra che tali numeri sono algebrici su $\mathbb{Q}$ e hanno grado una potenza di due.

Usiamo questo corollario per risolvere tre dei quattro problemi classici elencati sopra. Il quarto, la quadratura del cerchio, sarà trattato nel § 4.12, dove dimostreremo che π è trascendente.

Trisezione degli angoli. Non ogni angolo si può trisecare con riga e compasso. In particolare non si può trisecare l'angolo di 60°. Abbiamo visto che per costruire un angolo di 20° da uno di 60° occorre costruire $P = (\cos 20°, \sin 20°)$ a partire da $P_1 = (0, 0)$, $P_2 = (1, 0)$ e $P_3 = (\frac{1}{2}, \frac{1}{2}\sqrt{3})$. Sarebbe allora costruibile $Q = (\cos 20°, 0)$, piede della perpendicolare da P a P_1P_2. È più facile applicare a Q il criterio e il corollario. I numeri complessi in esame sono $z_1 = 0$, $z_2 = 1$, $z_3 = \frac{1}{2} + \frac{1}{2}\sqrt{3}\, i$, e il campo $F = \mathbb{Q}(z_1, z_2, z_3, \bar{z}_1, \bar{z}_2, \bar{z}_3)$ è $\mathbb{Q}(\sqrt{-3})$. Per il corollario, la trisezione richiederebbe che $\cos 20°$ avesse grado una potenza di due su F, e quindi anche su $\mathbb{Q}$, poiché $[\mathbb{Q}(\sqrt{-3}) : \mathbb{Q}] = 2$. Ma l'identità $\cos 3\theta = 4\cos^3\theta - 3\cos\theta$ dà $4a^3 - 3a = \frac{1}{2}$ per $a = \cos 20°$. Il polinomio minimo di a su $\mathbb{Q}$ divide dunque $4x^3 - 3x - \frac{1}{2}$. Se quest'ultimo è irriducibile, determina il polinomio minimo a meno di un fattore costante; il grado di a sarà tre, che non è una potenza di due, e la trisezione risulterà impossibile. Il polinomio dato è irriducibile se e solo se lo è

$$4(\tfrac{1}{2}x)^3 - 3(\tfrac{1}{2}x) - \tfrac{1}{2} = \tfrac{1}{2}x^3 - \tfrac{3}{2}x - \tfrac{1}{2}.$$

Moltiplicando per due si ottiene $x^3 - 3x - 1$. Ogni sua radice razionale sarebbe intera e dovrebbe dividere 1. Poiché né 1 né -1 sono radici, $x^3 - 3x - 1$ è irriducibile.

Duplicazione del cubo. Dobbiamo mostrare che $\sqrt[3]{2}$ non è un numero complesso costruibile. Ciò segue da $[\mathbb{Q}(\sqrt[3]{2}) : \mathbb{Q}] = 3$, essendo $x^3 - 2$ irriducibile in $\mathbb{Q}[x]$.

Costruzione dei poligoni regolari di p lati, con p primo. Occorre costruire il numero complesso $z = \cos(2\pi/p) + i\sin(2\pi/p)$. Si ha $z^p = 1$ e, poiché $x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \cdots + 1)$, risulta $z^{p-1} + z^{p-2} + \cdots + 1 = 0$. Il polinomio $x^{p-1} + x^{p-2} + \cdots + 1$ è irriducibile su $\mathbb{Q}$ (esercizio 3, p. 154), onde $[\mathbb{Q}(z) : \mathbb{Q}] = p - 1$. Condizione necessaria per la costruibilità del poligono è dunque che $p - 1 = 2^s$ per un intero $s \geq 0$. Poiché 6 non è una potenza di due, l'ettagono regolare non è costruibile. Osserviamo ora che per la primalità di $2^s + 1$, con $s > 0$, occorre che $s = 2^t$ per un intero $t \geq 0$. Infatti, se $s = uv$ ha un divisore dispari $u > 1$, dall'identità $x^u + 1 = (x+1)(x^{u-1} - x^{u-2} + \cdots + 1)$ segue

$$2^s + 1 = (2^v + 1)(2^{(u-1)v} - 2^{(u-2)v} + \cdots + 1),$$

e quindi $2^s + 1$ è composto. Si ottiene così una condizione necessaria più precisa per il poligono regolare di p lati: il primo p deve essere della forma $2^{2^t} + 1$. I primi di questa forma si chiamano *primi di Fermat*, da Pierre Fermat, che congetturò erroneamente che ogni intero di tale forma fosse primo.[3] I primi di Fermat noti sono $3, 5, 17, 257, 65537$, per $t = 0, 1, 2, 3, 4$. Sulla base dell'evidenza empirica si è congetturato che i primi di Fermat siano in numero finito, e non è escluso che l'elenco precedente li esaurisca.

Nel § 4.11 daremo una condizione necessaria e sufficiente per la costruibilità del poligono regolare di n lati, per ogni intero n. Ne seguirà il viceversa del risultato precedente: se p è primo di Fermat, il poligono regolare di p lati è costruibile. Concludiamo questo paragrafo calcolando $z = \cos(2\pi/17) + i\sin(2\pi/17)$ mediante operazioni razionali ed estrazioni di radici quadrate. Seguirà che z è costruibile, e con esso il poligono regolare di 17 lati.

Poniamo $\theta = 2\pi/17$ e $z = \cos\theta + i\sin\theta$. Allora $z^k = \cos k\theta + i\sin k\theta$, e queste potenze sono distinte per $1 \leq k \leq 17$. Poiché $(z^k)^{17} = (z^{17})^k = 1$, esse forniscono diciassette radici diciassettesime distinte dell'unità. L'equazione $x^{17} - 1 = 0$ ha al più diciassette radici (teorema 2.17, p. 132), quindi le z^k sono tutte le sue radici e formano un sottogruppo ciclico di $\mathbb{C}^*$. Il polinomio minimo di z su $\mathbb{Q}$ è il polinomio irriducibile $f(x) = \sum_{i=0}^{16} x^i$, che ha come radici $z, \ldots, z^{16}$ in $\mathbb{C}$. Dunque $f(x) = \prod_{k=1}^{16}(x - z^k)$ e

$$\sum_{k=1}^{16} z^k = -1. \tag{2}$$

Poiché $z\bar{z} = 1$ dà $\bar{z} = z^{-1}$, si ha $(z^k)^{-1} = \bar{z}^k = \cos k\theta - i\sin k\theta$ e

$$z^k + z^{-k} = 2\cos k\theta. \tag{3}$$

Il gruppo moltiplicativo di $\mathbb{Z}/(17)$ è ciclico (teorema 2.18, p. 132). La classe $\bar{3} = 3 + (17)$ è un generatore, poiché

$$3^2 \equiv 9, \quad 3^4 \equiv 13, \quad 3^8 \equiv 16 \pmod{17}. \tag{4}$$

L'ordine di $\bar{3}$ non può quindi essere $1, 2, 2^2$ o 2^3 e deve dividere 2^4, dunque è 2^4. Poniamo

$$\begin{aligned} x_1 &= z + z^{3^2} + z^{3^4} + z^{3^6} + z^{3^8} + z^{3^{10}} + z^{3^{12}} + z^{3^{14}}, \\ x_2 &= z^3 + z^{3^3} + z^{3^5} + z^{3^7} + z^{3^9} + z^{3^{11}} + z^{3^{13}} + z^{3^{15}}. \end{aligned} \tag{5}$$

[3] Eulero dimostrò che $2^{32} + 1 = 641 \cdot 6700417$. Una deduzione semplice di questa fattorizzazione si trova in G. H. Hardy ed E. M. Wright, *The Theory of Numbers*, New York, Oxford, 1938, p. 14.

Da $3^0+3^8\equiv 0 \pmod{17}$ si deducono $3^2+3^{10}\equiv 0$, $3^4+3^{12}\equiv 0$ e $3^6+3^{14}\equiv 0 \pmod{17}$. Perciò

$$\begin{aligned} x_1 &= (z+z^{3^8})+(z^{3^2}+z^{3^{10}})+(z^{3^4}+z^{3^{12}})+(z^{3^6}+z^{3^{14}}) \\ &= (z+z^{-1})+(z^8+z^{-8})+(z^4+z^{-4})+(z^2+z^{-2}). \end{aligned}$$

Ne segue

$$x_1 = 2(\cos\theta+\cos 8\theta+\cos 4\theta+\cos 2\theta) \tag{6}$$

e analogamente

$$x_2 = 2(\cos 3\theta+\cos 7\theta+\cos 5\theta+\cos 6\theta). \tag{7}$$

Si ha $x_1+x_2=\sum_{k=1}^{16} z^k=-1$. Moltiplicando direttamente e usando $2\cos\alpha\cos\beta=\cos(\alpha+\beta)+\cos(\alpha-\beta)$, si trova $x_1x_2=4(x_1+x_2)=-4$. Pertanto x_1,x_2 sono le radici di $(x-x_1)(x-x_2)=x^2+x-4=0$, ossia $\frac{1}{2}(-1\pm\sqrt{17})$. Poiché $\theta=2\pi/17$, si ha $\cos 3\theta>0$ e $\cos 7\theta,\cos 5\theta,\cos 6\theta<0$. Inoltre $|\cos 6\theta|=\cos(\pi-6\theta)=\cos(5\pi/17)>\cos(6\pi/17)=\cos 3\theta$. Quindi $x_2<0$ e

$$x_1=\tfrac{1}{2}(-1+\sqrt{17}), \qquad x_2=\tfrac{1}{2}(-1-\sqrt{17}). \tag{8}$$

Poniamo ora

$$\begin{aligned} y_1 &= z+z^{-1}+z^4+z^{-4} = 2(\cos\theta+\cos 4\theta), \\ y_2 &= z^8+z^{-8}+z^2+z^{-2} = 2(\cos 8\theta+\cos 2\theta), \\ y_3 &= z^3+z^{-3}+z^5+z^{-5} = 2(\cos 3\theta+\cos 5\theta), \\ y_4 &= z^7+z^{-7}+z^6+z^{-6} = 2(\cos 7\theta+\cos 6\theta). \end{aligned}$$

Usando direttamente l'identità sui coseni già ricordata, otteniamo

$$y_1y_2 = 4(\cos\theta+\cos 4\theta)(\cos 8\theta+\cos 2\theta) = 2\sum_{k=1}^{8}\cos k\theta = -1.$$

Analogamente $y_3y_4=-1$. Poiché $y_1+y_2=x_1$ e $y_3+y_4=x_2$, i numeri y_1,y_2 sono le radici di $x^2-x_1x-1=0$, e y_3,y_4 sono le radici di $x^2-x_2x-1=0$. Si ha $\cos\theta>\cos 2\theta$, $\cos 4\theta>0$ e $\cos 8\theta<0$, quindi $y_1>y_2$. Analogamente $y_3>y_4$. Dunque

$$\begin{aligned} y_1 &= \tfrac{1}{2}(x_1+\sqrt{x_1^2+4}), \quad y_2=\tfrac{1}{2}(x_1-\sqrt{x_1^2+4}), \\ y_3 &= \tfrac{1}{2}(x_2+\sqrt{x_2^2+4}), \quad y_4=\tfrac{1}{2}(x_2-\sqrt{x_2^2+4}). \end{aligned} \tag{9}$$

Infine poniamo $z_1=z+z^{-1}=2\cos\theta$ e $z_2=z^4+z^{-4}=2\cos 4\theta$. Allora $z_1>z_2$, $z_1+z_2=y_1$ e $z_1z_2=4\cos\theta\cos 4\theta=2(\cos 5\theta+\cos 3\theta)=y_3$. Pertanto

$$2\cos\theta=\tfrac{1}{2}(y_1+\sqrt{y_1^2-4y_3}). \tag{10}$$

Con (8) e (9) possiamo ora ottenere una formula esplicita, alquanto spaventosa, per $\cos\theta$. Otteniamo poi $\sin\theta=\sqrt{1-\cos^2\theta}$ e $z=\cos\theta+i\sin\theta$. Il numero complesso z è dunque costruibile e perciò si può costruire con riga e compasso il poligono regolare di 17 lati.[4] Abbiamo evitato di motivare i singoli passi del calcolo. Le ragioni di queste scelte si chiariranno quando avremo sviluppato la teoria di Galois.

[4]Per la costruzione geometrica effettiva si veda H. S. M. Coxeter, *Introduction to Geometry*, New York, Wiley, 1961, p. 27.

Esercizi

1. Dimostrare che il pentagono regolare è costruibile con riga e compasso.
2. Dimostrare che l'angolo $\arccos(11/16)$ si può trisecare con riga e compasso.
3. Dimostrare che l'ennagono regolare non è costruibile.

4.3 Campo di spezzamento di un polinomio

I matematici dell'Ottocento si occuparono quasi esclusivamente del campo $\mathbb{C}$ dei numeri complessi e dei suoi sottocampi. Dal punto di vista algebrico, la proprietà importante di $\mathbb{C}$ è la sua adeguatezza alla soluzione delle equazioni algebriche in un'incognita: è *algebricamente chiuso*, nel senso che ogni equazione polinomiale $x^n + a_1 x^{n-1} + \cdots + a_n = 0$ con $a_i \in \mathbb{C}$ ha una radice in $\mathbb{C}$. Del ruolo centrale di $\mathbb{C}$ nell'algebra ottocentesca è indizio il fatto che proprio allora si chiamò «il teorema fondamentale dell'algebra» il risultato che $\mathbb{C}$ è algebricamente chiuso. Conserviamo questa terminologia per rispetto del passato, benché il teorema non occupi più una posizione centrale nell'algebra. Per esempio, ci interessano anche i campi di caratteristica $p \neq 0$, fra l'altro per la loro utilità nella teoria dei numeri, e questi non s'immergono in $\mathbb{C}$. Partiremo da un campo base F arbitrario. Dato $f(x) \in F[x]$, vorremmo disporre di un'estensione E di F che contenga, in qualche senso, tutte le radici di $f(x) = 0$. Ricordiamo che $f(r) = 0$ se e solo se $x - r$ divide $f(x)$ (p. 130 dell'originale). Diremo che $f(x)$, supposto monico, *si spezza* in E se $f(x) = \prod_{i=1}^{n}(x - r_i)$, cioè se è prodotto di fattori lineari (di primo grado) in $E[x]$. Se r è una radice in E, da $0 = f(r) = \prod_i (r - r_i)$ segue che r è uno degli r_i. La medesima fattorizzazione vale anche in $R[x]$, dove $R = F(r_1, \ldots, r_n)$. Quando ci occupiamo del solo polinomio $f(x)$, conviene spostare l'attenzione da E a $F(r_1, \ldots, r_n)$, costruito appositamente per lo studio di f. Giungiamo così alla seguente importante definizione.

Definizione 4.1. *Siano F un campo e $f(x) \in F[x]$ un polinomio monico. Un'estensione E/F si dice campo di spezzamento di $f(x)$ su F se*

$$\begin{aligned} &\text{(i)} \quad f(x) = (x - r_1)(x - r_2)\cdots(x - r_n) \quad \text{in } E[x], \\ &\text{(ii)} \quad E = F(r_1, r_2, \ldots, r_n), \end{aligned}$$

cioè E è generato su F dalle radici di $f(x)$.

Dimostriamo anzitutto l'esistenza di un campo di spezzamento per ogni polinomio di grado positivo. Estenderemo un procedimento impiegato per la prima volta da A. Cauchy per costruire $\mathbb{C}$ a partire da $\mathbb{R}$ (aggiunzione di $\sqrt{-1}$) e adoperato in seguito da L. Kronecker per costruire una radice di un polinomio irriducibile.

Teorema 4.3. *Ogni polinomio monico $f(x) \in F[x]$ di grado positivo ha un campo di spezzamento E/F.*

Dimostrazione. Sia $f(x) = f_1(x) f_2(x) \cdots f_k(x)$ la fattorizzazione in polinomi monici irriducibili. Evidentemente $k \leq n = \deg f(x)$. Procediamo per induzione su $n - k$. Se $n - k = 0$, tutti i fattori sono lineari e F stesso è un campo di spezzamento. Sia dunque $n - k > 0$, e sia, per esempio, $f_1(x)$ un fattore di grado maggiore di uno. Poniamo $K = F[x]/(f_1(x))$. Poiché $f_1(x)$ è irriducibile, K è un campo. Identificando $a \in F$ con $a + (f_1(x))$, si ha un'estensione di F, e $K = F(r)$ dove $r = x + (f_1(x))$ è una radice di $f_1(x) = 0$. Conviene ora dimenticare i particolari della costruzione e tenere a mente soltanto che disponiamo di un'estensione K/F generata su F da una sola radice r del polinomio irriducibile $f_1(x)$. Poiché $K \supset F$ e $f(x), f_i(x) \in F[x] \subset K[x]$, si ottiene la fattorizzazione di f in fattori monici irriducibili in $K[x]$ scomponendovi ciascun $f_i(x)$. Ora $f_1(r) = 0$, quindi $f_1(x) = (x - r)g(x)$ in $K[x]$. Se ℓ è il numero di fattori irriducibili

di $f(x)$ in $K[x]$, si ha pertanto $\ell > k$ e $n - \ell < n - k$. Applicando l'ipotesi induttiva a $f(x)$ e a K, otteniamo un'estensione $E = K(r_1, \ldots, r_n)$ tale che $f(x) = \prod_{i=1}^{n}(x - r_i)$ in $E[x]$. Siccome $f_1(r) = 0$ e $f_1(x)$ divide $f(x)$, si ha $f(r) = 0$ e dunque $r = r_i$ per un certo i. Allora

$$E = K(r_1, \ldots, r_n) = F(r)(r_1, \ldots, r_n) = F(r_1, \ldots, r_n)$$

è un campo di spezzamento di $f(x)$ su F. □

Esempi

1. Sia $f(x) = x^2 + ax + b$. Se $f(x)$ è riducibile in $F[x]$, con F arbitrario, F stesso è un campo di spezzamento. Altrimenti poniamo $E = F[x]/(f(x)) = F(r_1)$, dove $r_1 = x + (f(x))$. Allora E è un campo di spezzamento: poiché $f(r_1) = 0$, si ha $f(x) = (x - r_1)(x - r_2)$ in $E[x]$. Dunque $E = F(r_1) = F(r_1, r_2)$. Poiché f è il polinomio minimo di r_1 su F, $[E : F] = 2$.

2. Sia $F = \mathbb{Z}/(2)$, il campo di due elementi, e $f(x) = x^3 + x + 1$. Poiché $1 + 1 + 1 \not\equiv 0$ e $0 + 0 + 1 \not\equiv 0 \pmod 2$, f non ha radici in F ed è dunque irriducibile in $F[x]$. Poniamo $r_1 = x + (f(x))$ in $F[x]/(f(x))$. Allora $F(r_1)$ è un campo e in $F(r_1)[x]$ si ha

   $$x^3 + x + 1 = (x + r_1)(x^2 + ax + b).$$

   Possiamo scrivere + al posto di − perché la caratteristica è due. Confrontando i coefficienti, $a = r_1$ e $b = 1 + r_1^2$. Gli elementi di $F(r_1)$ sono i numeri $c + dr_1 + er_1^2$, $c, d, e \in F$. Sono otto:

   $$0,\ 1,\ r_1,\ 1 + r_1,\ r_1^2,\ 1 + r_1^2,\ r_1 + r_1^2,\ 1 + r_1 + r_1^2.$$

   Sostituendoli in $x^2 + r_1 x + 1 + r_1^2$, si trova $(r_1^2)^2 + r_1(r_1^2) + 1 + r_1^2 = r_1^4 + r_1^3 + 1 + r_1^2 = 0$, perché $r_1^3 = r_1 + 1$ e $r_1^4 = r_1^2 + r_1$. Il fattore $x^2 + ax + b$ si scompone quindi in fattori lineari in $F(r_1)[x]$ ed $E = F(r_1)$ è un campo di spezzamento di $x^3 + x + 1$ su F.

3. Siano $F = \mathbb{Q}$ e $f(x) = (x^2 - 2)(x^2 - 3)$. Le eventuali radici razionali di $x^2 - 2$ e di $x^2 - 3$ dovrebbero essere intere (esercizio 1, p. 154), perciò questi polinomi sono irriducibili in $\mathbb{Q}[x]$. Formiamo $K = \mathbb{Q}(r_1)$, dove $r_1 = x + (x^2 - 2)$ in $\mathbb{Q}[x]/(x^2 - 2)$. Gli elementi di K sono della forma $a + br_1$ con $a, b \in \mathbb{Q}$. Affermiamo che $x^2 - 3$ è irriducibile in $K[x]$. Altrimenti esisterebbero a, b razionali con $(a + br_1)^2 = 3$. Ne seguirebbe $(a^2 + 2b^2) + 2abr_1 = 3$, dunque $ab = 0$ e $a^2 + 2b^2 = 3$. Se $b = 0$, si avrebbe $a^2 = 3$, impossibile perché $\sqrt{3}$ è irrazionale. Se $a = 0$, si avrebbe $b^2 = 3/2$ e quindi $(2b)^2 = 6$, impossibile per l'irrazionalità di $\sqrt{6}$. Pertanto $x^2 - 3$ è irriducibile in $K[x]$. Formiamo ora $E = K[x]/(x^2 - 3)$. È un campo di spezzamento di $(x^2 - 2)(x^2 - 3)$ su $\mathbb{Q}$ e $[E : \mathbb{Q}] = [E : K][K : \mathbb{Q}] = 2 \cdot 2 = 4$.

4. Siano $F = \mathbb{Q}$, $f(x) = x^p - 1$ e p primo. Si ha $x^p - 1 = (x - 1)(x^{p-1} + x^{p-2} + \cdots + 1)$ e il secondo fattore è irriducibile in $\mathbb{Q}[x]$. Sia $E = \mathbb{Q}(z)$ con $z = x + (x^{p-1} + \cdots + 1)$ in $\mathbb{Q}[x]/(x^{p-1} + \cdots + 1)$. Si ha $z^p = 1$ e, poiché $x^{p-1} + \cdots + 1$ è il polinomio minimo di z su $\mathbb{Q}$, gli elementi $1, z, \ldots, z^{p-1}$ sono distinti. Inoltre $(z^k)^p = (z^p)^k = 1$, dunque ciascun z^k è radice di $x^p - 1$. Ne segue che $x^p - 1 = \prod_{k=1}^{p}(x - z^k)$ in $E[x]$. Pertanto E è un campo di spezzamento di $x^p - 1$ su $\mathbb{Q}$ e $[E : \mathbb{Q}] = p - 1$.

Prima di passare al prossimo risultato principale, l'unicità dei campi di spezzamento a meno di isomorfismo, osserviamo che questi campi hanno dimensione finita sul campo base. Sia E/F un campo di spezzamento di $f(x)$. Allora $E = F(r_1, \ldots, r_n)$ con $f(r_i) = 0$ per $1 \le i \le n$. L'elemento r_i è algebrico su F, e quindi anche su $F(r_1, \ldots, r_{i-1})$. La dimensione $[F(r_1, \ldots, r_i) :$

$F(r_1, \ldots, r_{i-1})]$ è finita, essendo il grado di r_i su $F(r_1, \ldots, r_{i-1})$. Per applicazione ripetuta della formula delle dimensioni (teorema 4.2) otteniamo

$$[E : F] = \prod_{i=1}^{n} [F(r_1, \ldots, r_i) : F(r_1, \ldots, r_{i-1})] < \infty,$$

dove il primo fattore è $[F(r_1) : F]$.

Mostreremo che due campi di spezzamento su F di uno stesso polinomio $f(x) \in F[x]$ sono isomorfi, e otterremo informazioni importanti anche sul numero di isomorfismi fra essi. Per svolgere la dimostrazione per induzione è necessario generalizzare leggermente la situazione. Consideriamo campi isomorfi F e $\bar{F}$ e un isomorfismo $\eta : a \mapsto \bar{a}$ di F su $\bar{F}$. Esso si estende in modo unico a un isomorfismo $g(x) \mapsto \bar{g}(x)$ di $F[x]$ su $\bar{F}[x]$ che lascia fisso x. Sia $f(x) \in F[x]$ monico di grado positivo, sia E un suo campo di spezzamento su F e sia $\bar{E}$ un campo di spezzamento di $\bar{f}(x)$ su $\bar{F}$. Vale il seguente importante teorema.

Teorema 4.4. *Sia $\eta : a \mapsto \bar{a}$ un isomorfismo di F su $\bar{F}$. Siano $f(x) \in F[x]$ monico di grado positivo e $\bar{f}(x) \in \bar{F}[x]$ il corrispondente polinomio per l'isomorfismo che estende η e manda x in x. Siano E ed $\bar{E}$ campi di spezzamento di $f(x)$ e di $\bar{f}(x)$ su F e su $\bar{F}$ rispettivamente. Allora η si può estendere a un isomorfismo di E su $\bar{E}$. Il numero dei prolungamenti è al più $[E : F]$ ed è esattamente $[E : F]$ se $f(x)$ ha radici distinte in E.*

Isoliamo un lemma che servirà nel passo induttivo della dimostrazione.

Lemma. *Sia η un isomorfismo di F su $\bar{F}$ e siano E, $\bar{E}$ estensioni rispettivamente di F, $\bar{F}$. Se $r \in E$ è algebrico su F con polinomio minimo $g(x)$, allora η si estende a un monomorfismo ζ di $F(r)$ in $\bar{E}$ se e solo se $\bar{g}(x)$ ha una radice in $\bar{E}$. In tal caso il numero di prolungamenti è il numero di radici distinte di $\bar{g}(x)$ in $\bar{E}$.*

Dimostrazione. Se esiste un prolungamento ζ, applicandolo alla relazione $g(r) = 0$ otteniamo $\bar{g}(\zeta(r)) = 0$. Quindi $\zeta(r)$ è una radice di $\bar{g}(x)$ in $\bar{E}$. Viceversa, sia $\bar{r}$ una tale radice. Abbiamo un omomorfismo $h(x) \mapsto \bar{h}(\bar{r})$ di $F[x]$ in $\bar{E}$ (teorema 2.10, p. 122). Il nucleo contiene $(g(x))$ e si ha l'omomorfismo indotto $h(x) + (g(x)) \mapsto \bar{h}(\bar{r})$ di $F[x]/(g(x))$ in $\bar{E}$. Analogamente, $h(x) + (g(x)) \mapsto h(r)$ è un omomorfismo dello stesso quoziente su $F(r)$. L'irriducibilità di $g(x)$ implica che $F[x]/(g(x))$ è un campo; entrambi gli omomorfismi sono quindi iniettivi e il secondo è un isomorfismo. Se all'inverso di quest'ultimo facciamo seguire il monomorfismo del quoziente in $\bar{E}$, otteniamo il monomorfismo $h(r) \mapsto \bar{h}(\bar{r})$ di $F(r) = F[r]$ in $\bar{E}$. Poiché $F(r)$ è generato da F e da r, questo è l'unico monomorfismo che estende η e manda r in $\bar{r}$. Esistono perciò tanti prolungamenti quante sono le scelte possibili di $\bar{r}$, ossia le radici distinte di $\bar{g}(x)$ in $\bar{E}$. □

Dimostrazione del teorema 4.4. Procediamo per induzione su $[E : F]$. Se $[E : F] = 1$, allora $E = F$ e $f(x) = \prod_i (x - r_i)$ in $F[x]$. Applicando l'isomorfismo $h(x) \mapsto \bar{h}(x)$, otteniamo $\bar{f}(x) = \prod_i (x - \bar{r}_i)$ in $\bar{F}[x]$. Le radici di $\bar{f}(x)$ in $\bar{E}$ sono dunque gli $\bar{r}_i$ e, poiché queste radici generano $\bar{E}$ su $\bar{F}$, si ha $\bar{E} = \bar{F}$. Esiste così un solo prolungamento, e $1 = [E : F]$. Supponiamo $[E : F] > 1$. Allora $f(x)$ non è un prodotto di fattori lineari in $F[x]$. Sia $g(x)$ un suo fattore monico irriducibile di grado $m > 1$. Ne segue $\bar{g}(x) \mid \bar{f}(x)$ in $\bar{F}[x]$. Possiamo disporre le fattorizzazioni come segue:

$$g(x) = \prod_{i=1}^{m} (x - r_i), \quad f(x) = \prod_{j=1}^{n} (x - r_j) \quad \text{in } E[x],$$
$$\bar{g}(x) = \prod_{i=1}^{m} (x - s_i), \quad \bar{f}(x) = \prod_{j=1}^{n} (x - s_j) \quad \text{in } \bar{E}[x].$$

Poniamo $K = F(r_1)$. Poiché $g(x)$ è irriducibile, è il polinomio minimo di r_1 su F e $[K : F] = m$. Per il lemma vi sono k monomorfismi $\zeta_1, \ldots, \zeta_k$ di K in $\bar{E}$ che estendono η, dove k è il numero

di elementi distinti fra $s_1, \ldots, s_m$. In particolare $k = m$ se essi sono tutti distinti. È chiaro per definizione che E è un campo di spezzamento su K di $f(x) \in K[x]$ e $\bar{E}$ è un campo di spezzamento di $\bar{f}(x)$ su $\zeta_i(K)$. Poiché $[E : K] = [E : F]/[K : F] = [E : F]/m < [E : F]$, per induzione ogni ζ_i si estende a un isomorfismo di E su $\bar{E}$; il numero dei prolungamenti di ζ_i è al più $[E : K]$ ed è uguale a $[E : K]$ se $\bar{f}(x)$ ha radici distinte in $\bar{E}$. Ciascuno di questi isomorfismi estende η. Otteniamo così almeno un prolungamento di η a un isomorfismo di E su $\bar{E}$. Inoltre, i prolungamenti di ζ_i distinti sono distinti: si ottengono dunque al più $m[E : K] = [E : F]$ prolungamenti di η, ed esattamente $[E : F]$ se $\bar{f}(x)$ ha radici distinte. Resta soltanto da convincersi che il procedimento comprende ogni prolungamento da E su $\bar{E}$. Ma se ζ è uno di essi, la sua restrizione a K è un monomorfismo di K in $\bar{E}$ che estende η, e coincide quindi con uno degli ζ_i, $1 \leq i \leq k$. □

Nel caso particolare $\bar{F} = F$, con η identità di F, la prima parte del teorema dice che se $E, \bar{E}$ sono due campi di spezzamento di $f(x)$ su F, esiste un isomorfismo di E su $\bar{E}$ che si riduce all'identità su F. Un simile isomorfismo si dice *isomorfismo su* F di E su $\bar{E}$ (e analogamente per un monomorfismo). Dalla seconda parte segue che gli automorfismi di E/F, ossia di E su F, sono al più $[E : F]$ e sono esattamente $[E : F]$ se $f(x)$ ha n radici distinte.

Esercizi

1. Mostrare che la dimensione di un campo di spezzamento E/F di un polinomio di grado n è al più $n!$.

2. Costruire un campo di spezzamento di $x^5 - 2$ su $\mathbb{Q}$. Determinarne la dimensione su $\mathbb{Q}$.

3. Determinare un campo di spezzamento di $x^{p^e} - 1$ su $\mathbb{Z}/(p)$, con $e \in \mathbb{N}$.

4. Sia E/F un campo di spezzamento di $f(x)$ su F e sia K un sottocampo di E/F. Mostrare che ogni monomorfismo di K/F in E/F si estende a un automorfismo di E.

5. Sia E un'estensione di F con $[E : F] = n < \infty$, e sia K una qualunque estensione di F. Usare il metodo della dimostrazione del teorema 4.4 per mostrare che il numero dei monomorfismi di E/F in K/F è al più n.

4.4 Radici multiple

Sia $f(x)$ un polinomio monico di grado positivo in $F[x]$ e sia E/F un campo di spezzamento. Scriviamo la fattorizzazione in $E[x]$ come

$$f(x) = (x - r_1)^{k_1}(x - r_2)^{k_2} \cdots (x - r_s)^{k_s}, \tag{11}$$

con $r_i \in E$ e $r_i \neq r_j$ per $i \neq j$. Diciamo che r_i è una *radice di molteplicità* k_i dell'equazione $f(x) = 0$. Se $k_i = 1$ la radice si dice *semplice*, altrimenti *multipla*. In un secondo campo di spezzamento $\bar{E}/F$ si avrebbe $f(x) = \prod_{j=1}^{s}(x - \bar{r}_j)^{k_j}$ in $\bar{E}[x]$, dove $a \mapsto \bar{a}$ è un isomorfismo di E/F su $\bar{E}/F$. Le molteplicità k_i sono pertanto indipendenti dalla scelta del campo di spezzamento. In particolare, la proprietà che $f(x)$ abbia soltanto radici semplici è indipendente dalla scelta di E. Il teorema 4.4 mostra che vi è un netto vantaggio nel lavorare con siffatti polinomi, perché in questo caso abbiamo una formula esatta: il numero degli automorfismi di E/F è $[E : F]$.

Mostreremo che, se F ha caratteristica zero oppure è un campo finito, si può supporre senza perdita di generalità che tutte le radici siano semplici. Osserviamo anzitutto che, se in $F[x]$ si ha $f(x) = p_1(x)^{l_1} \cdots p_t(x)^{l_t}$, con $p_i(x)$ primi distinti, E/F è un campo di spezzamento di $f(x)$ se e solo se lo è di $f_0(x) = p_1(x) \cdots p_t(x)$. È immediato dalla definizione. Possiamo dunque

supporre fin dall'inizio che f sia prodotto di polinomi primi distinti di $F[x]$. Notiamo anche che, se $p(x)$ e $q(x)$ sono due primi monici distinti di $F[x]$, si ha $(p(x), q(x)) = 1$ in $F[x]$. Esistono quindi $a(x), b(x) \in F[x]$ tali che $a(x)p(x) + b(x)q(x) = 1$. Ciò esclude che $p(x) = 0$ e $q(x) = 0$ possano avere una radice comune in E. Pertanto, se f è prodotto di primi distinti, tutte le sue radici sono semplici se e solo se lo sono quelle dei suoi fattori primi.

Sviluppiamo ora un criterio per le radici multiple che si possa verificare in $F[x]$, senza ricorrere a un campo di spezzamento. Ci baseremo sulla derivazione formale dei polinomi. Aggiungiamo un'indeterminata h a $F[x]$, ottenendo l'anello $F[x, h]$ dei polinomi nelle due indeterminate x, h. Poiché $F[x, h] = F[x][h]$ e h è trascendente su $F[x]$, ogni suo elemento si scrive in un solo modo come $f_0(x) + f_1(x)h + \cdots + f_n(x)h^n$, con $f_i(x) \in F[x]$. In particolare, per $f(x) \in F[x]$ si ha $f(x+h) = f_0(x)+f_1(x)h+\cdots+f_n(x)h^n$. Ponendo $h = 0$, vale a dire applicando l'omomorfismo di $F[x, h]$ in $F[x]$ dato da $a \mapsto a$ per $a \in F$, $x \mapsto x$, $h \mapsto 0$, otteniamo $f(x) = f_0(x)$. La differenza $f(x+h)-f(x)$ è perciò divisibile per h. Dividendo per h si ottiene $f_1(x)+f_2(x)h+\cdots+f_n(x)h^{n-1}$; ponendo qui $h = 0$ rimane $f_1(x)$, che definiamo *derivata* $f'(x)$, o $f(x)'$, di $f(x)$.[5] Evidentemente $f'(x)$ soddisfa la congruenza

$$f(x + h) \equiv f(x) + f'(x)h \pmod{h^2}. \tag{12}$$

D'altra parte, se $g(x) \in F[x]$ soddisfa $f(x + h) \equiv f(x) + g(x)h \pmod{h^2}$, allora $f'(x)h \equiv g(x)h \pmod{h^2}$, quindi $f'(x) \equiv g(x) \pmod{h}$, da cui $g(x) = f'(x)$. La (12) caratterizza dunque $f'(x)$ e permette di stabilire rapidamente le proprietà fondamentali dell'applicazione $f \mapsto f'$, detta *derivazione standard* di $F[x]$. Esse sono:

(i) F-linearità: $(f + g)' = f' + g'$ e $(af)' = af'$ per $a \in F$.

(ii) Regola del prodotto:

$$(fg)' = f'g + fg'. \tag{13}$$

(iii) $x' = 1$.

La (i) segue immediatamente dalla (12). Per la regola del prodotto si moltiplica la (12) per la corrispondente congruenza per $g(x + h)$:

$$\begin{aligned}(fg)(x + h) &= f(x + h)g(x + h)\\ &\equiv [f(x) + f'(x)h][g(x) + g'(x)h]\\ &\equiv f(x)g(x) + [f'(x)g(x) + f(x)g'(x)]h \pmod{h^2}.\end{aligned}$$

La (13) deriva allora dalla proprietà caratteristica (12) applicata a fg. Da $x + h \equiv x + 1h \pmod{h^2}$ segue $x' = 1$, ossia (iii). Da ciò e dalla regola del prodotto si ha $(x^k)' = kx^{k-1}$ per $k = 1, 2, 3, \ldots$. Inoltre, da $1^2 = 1$ segue $1'\,1+1\,1' = 1'$, cioè $2(1') = 1'$, quindi $1' = 0$. La linearità dà infine, per $f(x) = a_0 + a_1x + \cdots + a_nx^n$,

$$f'(x) = a_1 + 2a_2x + \cdots + na_nx^{n-1},$$

come nel calcolo differenziale delle funzioni di una variabile reale.

Teorema 4.5. *Sia $f(x) \in F[x]$ monico di grado positivo. Le sue radici in un qualsiasi campo di spezzamento E/F sono tutte semplici se e solo se $(f, f') = 1$.*

[5] Il lettore noti l'analogia con la definizione di derivata di una funzione di variabile reale: lì h è una variabile e si prende il limite di $[f(x + h) - f(x)]/h$ per $h \to 0$.

Dimostrazione. Sia $d(x) = (f(x), f'(x))$ in $F[x]$. Se $f(x)$ ha una radice multipla in E, possiamo scrivere $f(x) = (x-r)^k g(x)$ con $k > 1$. Derivando in $E[x]$ si ottiene

$$f'(x) = (x-r)^k g'(x) + k(x-r)^{k-1} g(x),$$

che è divisibile per $x-r$, poiché $k-1 \geq 1$. Dunque $x-r$ è un fattore comune di $f(x)$ e $f'(x)$ in $E[x]$ e $d(x) \neq 1$. Supponiamo invece che tutte le radici siano semplici. Allora $f(x) = \prod_{i=1}^{n}(x-r_i)$ con $r_i \neq r_j$ per $i \neq j$. Estendendo la regola del prodotto a più di due fattori si ottiene

$$f'(x) = \sum_{j=1}^{n}(x-r_1)\cdots(x-r_{j-1})(x-r_{j+1})\cdots(x-r_n).$$

È chiaro che nessun $x - r_i$ divide $f'(x)$, dunque $(f, f') = 1$. □

Se $f(x)$ è irriducibile in $F[x]$, la condizione $(f, f') \neq 1$ implica $f \mid f'$. Poiché $\deg f' < \deg f$, si deve avere $f' = 0$. In caratteristica zero la formula della derivata mostra che $f' = 0$ se e solo se $f \in F$. Pertanto $f' \neq 0$ quando f è irriducibile e F ha caratteristica zero. Se la caratteristica è $p \neq 0$ e $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, allora $f'(x) = \sum_{i=1}^{n} i a_i x^{i-1}$ e $f' = 0$ se e solo se $ia_i = 0$ per $1 \leq i \leq n$. Ciò equivale a richiedere $a_i = 0$ per ogni i non divisibile per p, cioè a richiedere

$$f(x) = b_0 + b_1 x^p + b_2 x^{2p} + \cdots + b_m x^{mp} = g(x^p), \qquad g(x) = b_0 + b_1 x + \cdots + b_m x^m.$$

Costruiremo ora un polinomio irriducibile in caratteristica p che possieda radici multiple. Sia F un qualunque campo di caratteristica p. Si ha $1^p = 1$ e, per la commutatività del prodotto, $(ab)^p = a^p b^p$. Per il teorema binomiale,

$$(a+b)^p = a^p + b^p + \sum_{i=1}^{p-1} \binom{p}{i} a^i b^{p-i}.$$

Il coefficiente $\binom{p}{i} = p!/(i!(p-i)!)$ è intero e p compare nel numeratore, ma non nel denominatore della frazione: è quindi divisibile per p. La somma dei termini intermedi è zero in F e risulta $(a+b)^p = a^p + b^p$. Abbiamo dunque in F

$$(a+b)^p = a^p + b^p, \qquad (ab)^p = a^p b^p, \qquad 1^p = 1. \tag{14}$$

Ciò mostra che $a \mapsto a^p$ è un endomorfismo dell'anello F. Poiché F è un campo, è un monomorfismo e la sua immagine F^p, l'insieme delle potenze p-esime, è un sottocampo di F.

Lemma. *Se F ha caratteristica p e $a \in F$, il polinomio $x^p - a$ è irriducibile in $F[x]$ oppure è una potenza p-esima in $F[x]$.*

Dimostrazione. Supponiamo $x^p - a = g(x)h(x)$ in $F[x]$, con g monico di grado k, $1 \leq k \leq p-1$. Sia E un campo di spezzamento di $x^p - a$ su F e sia $b \in E$ una radice. Allora $b^p = a$ e $x^p - a = x^p - b^p = (x-b)^p = g(x)h(x)$. Dunque $g(x) = (x-b)^k$ e $b^k \in F$. Anche $b^p \in F$; poiché esistono interi u, v con $uk + vp = 1$, per $b \neq 0$ si ha $b = (b^k)^u (b^p)^v \in F$ (il caso $b = 0$ è immediato). Perciò $x^p - a = (x-b)^p$ in $F[x]$. □

Per l'esempio promesso prendiamo come campo base $F = (\mathbb{Z}/(p))(t)$, il campo delle espressioni razionali in un'indeterminata t sul campo fondamentale di p elementi, ossia il campo delle frazioni di $(\mathbb{Z}/(p))[t]$. Affermiamo che t non è una potenza p-esima in F. Supponiamo $t = (f(t)/g(t))^p$, con $f(t) = a_0 + a_1 t + \cdots + a_n t^n$ e $g(t) = b_0 + b_1 t + \cdots + b_m t^m$. Allora $f(t)^p = a_0^p + a_1^p t^p + \cdots + a_n^p t^{np}$ e $g(t)^p = b_0^p + b_1^p t^p + \cdots + b_m^p t^{mp}$, sicché

$$(b_0^p + b_1^p t^p + \cdots + b_m^p t^{mp})t = a_0^p + a_1^p t^p + \cdots + a_n^p t^{np}.$$

Per l'indipendenza lineare delle potenze $1, t, t^2, \dots$ su $\mathbb{Z}/(p)$, tutti i b_i devono essere zero, in contrasto con l'ipotesi sottintesa $g(t) \neq 0$. Il lemma mostra che $f(x) = x^p - t \in F[x]$ è irriducibile. Esso è però una potenza p-esima in $E[x]$, dove E è un campo di spezzamento. Si può verificare che ha radici multiple anche con il criterio della derivata, poiché $(x^p - t)' = px^{p-1} = 0$.

Diremo che un polinomio di $F[x]$ è *separabile* se i suoi fattori irriducibili hanno radici distinte. Abbiamo provato che in caratteristica zero ogni polinomio a coefficienti in F è separabile e che in caratteristica p vi sono, almeno per certi F, polinomi inseparabili. Esaminiamo più a fondo il secondo caso. Diremo che un campo F, di qualunque caratteristica, è *perfetto* se ogni polinomio di $F[x]$ è separabile. Abbiamo già visto che ogni campo di caratteristica zero è perfetto. In caratteristica $p \neq 0$ vale il seguente criterio.

Teorema 4.6. *Un campo F di caratteristica $p \neq 0$ è perfetto se e solo se coincide con F^p, il sottocampo delle potenze p-esime dei suoi elementi.*

Dimostrazione. Se $F^p \subsetneq F$, prendiamo $a \in F \setminus F^p$. Per il lemma, $x^p - a$ è irriducibile. La sua derivata è zero, quindi il polinomio è inseparabile e F non è perfetto. Viceversa, sia $f(x) \in F[x]$ un polinomio irriducibile inseparabile. Allora $(f, f') \neq 1$ e abbiamo visto che $f(x) = a_0 + a_p x^p + a_{2p} x^{2p} + \cdots$. Almeno uno di questi coefficienti non è una potenza p-esima. Infatti, se ogni $a_i = b_i^p$, allora

$$f(x) = b_0^p + b_p^p x^p + b_{2p}^p x^{2p} + \cdots = (b_0 + b_p x + b_{2p} x^2 + \cdots)^p,$$

contro l'irriducibilità di $f(x)$. Dunque $F \neq F^p$. □

Corollario. *Ogni campo finito è perfetto.*

Dimostrazione. La caratteristica di un campo finito F è un numero primo p. Il monomorfismo $a \mapsto a^p$ di F è un isomorfismo, poiché F è finito. Vale dunque $F = F^p$ e il teorema 4.6 dà la tesi. □

Esercizi

1. Sia F un campo di caratteristica zero e sia $f(x) \in F[x]$ monico di grado positivo. Posto $d(x) = (f(x), f'(x))$, mostrare che $g(x) = f(x)d(x)^{-1}$ ha le stesse radici di $f(x)$ e che sono tutte radici semplici di $g(x)$.

2. Sia $f(x)$ irriducibile in $F[x]$, con F di caratteristica p. Mostrare che $f(x)$ è della forma $g(x^{p^e})$, con g irriducibile e separabile. Dedurne che in un campo di spezzamento ogni radice di $f(x)$ ha la stessa molteplicità p^e.

3. Sia F di caratteristica p. Un polinomio $f(x) \in F[x]$ si dice *p-polinomio* se ha la forma $x^{p^m} + a_1 x^{p^{m-1}} + \cdots + a_m x$. Dimostrare che un polinomio monico di grado positivo è un p-polinomio se e solo se le sue radici formano un sottogruppo finito del gruppo additivo del campo di spezzamento e ciascuna ha la stessa molteplicità p^e.

4. Sia F imperfetto di caratteristica p. Mostrare che $x^{p^e} - a$ è irriducibile se $a \notin F^p$ ed $e = 0, 1, 2, \dots$.

5. Sia F di caratteristica p e sia $a \in F$. Mostrare che $f(x) = x^p - x - a$ non ha radici multiple e che è irriducibile in $F[x]$ se e solo se $a \neq c^p - c$ per ogni $c \in F$.

4.5 Il gruppo di Galois. La corrispondenza fondamentale di Galois

Dedurremo ora i risultati centrali della teoria di Galois. Se E è un campo di spezzamento di un polinomio separabile in $F[x]$, essi stabiliscono una corrispondenza biunivoca tra i sottocampi di

E/F e i sottogruppi del gruppo degli automorfismi di E/F. Le sue proprietà stanno alla base sia del criterio di Galois per la risolubilità di un'equazione per radicali sia del criterio di costruibilità con riga e compasso. Come si è osservato nell'introduzione al capitolo, questi risultati svolgono inoltre un ruolo fondamentale in molte altre questioni di algebra e di teoria dei numeri.

Cominciamo con alcune definizioni e notazioni. Data un'estensione E di un campo F, sia G l'insieme degli automorfismi di E/F, cioè degli automorfismi η di E tali che $\eta(a) = a$ per ogni $a \in F$. Si tratta di un gruppo di trasformazioni di E: $1 \in G$ e, se $\eta, \zeta \in G$, anche $\eta\zeta, \eta^{-1} \in G$. Lo chiameremo *gruppo di Galois di E su F* e lo indicheremo con $\mathrm{Gal}(E/F)$ quando vorremo precisare i campi.

Esempi.

1. Sia $E = F(u)$ con $u^2 = a \in F$, dove a non è un quadrato in F. Supponiamo inoltre $\mathrm{char}\, F \neq 2$. Il polinomio $x^2 - a$ è irriducibile in $F[x]$ ed è quindi il polinomio minimo di u su F. Dunque $[E : F] = 2$ e $(1, u)$ è una base di E/F. Le due applicazioni $c + du \mapsto c + du$ e $c + du \mapsto c - du$, con $c, d \in F$, sono evidentemente automorfismi di E/F. Sono i soli. Infatti, per $\eta \in \mathrm{Gal}(E/F)$, da $u^2 = a$ segue $\eta(u)^2 = a$. Le radici di $x^2 - a$ in E sono $u, -u$, sicché $\eta(u) = u$ oppure $\eta(u) = -u$. Nel primo caso η è l'identità, nel secondo manda $c + du$ in $c - du$. Il gruppo di Galois è perciò ciclico di ordine due.

2. Sia $E = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Si vede facilmente che $\mathrm{Gal}(E/\mathbb{Q})$ ha ordine quattro. Esso comprende $\eta_1 = 1, \eta_2, \eta_3, \eta_4$, che mandano la coppia $(\sqrt{2}, \sqrt{3})$ rispettivamente in $(\sqrt{2}, \sqrt{3})$, $(-\sqrt{2}, \sqrt{3})$, $(\sqrt{2}, -\sqrt{3})$, $(-\sqrt{2}, -\sqrt{3})$.

3. Sia F imperfetto di caratteristica p e sia $a \in F \setminus F^p$. Allora $x^p - a$ è irriducibile (lemma, p. 232 dell'originale). Aggiungendo una radice u di $x^p = a$ si ottiene un'estensione $E = F(u)$ con $[E : F] = p$. Poiché $x^p - a = (x - u)^p$ in $E[x]$, E è il campo di spezzamento su F del polinomio inseparabile $x^p - a$. Per ogni $\eta \in \mathrm{Gal}(E/F)$ si ha $\eta(u)^p = a$, quindi $\eta(u) = u$. Ne segue $\eta = 1$ e $\mathrm{Gal}(E/F) = 1$.

4. Sia $E = F(t)$ con t trascendente su F. Come si vedrà nell'esercizio 11, $u \in E$ genera E/F se e solo se è della forma

$$u = \frac{at + b}{ct + d}, \qquad ad - bc \neq 0. \tag{15}$$

   Poiché un automorfismo di E/F manda generatori in generatori, $\mathrm{Gal}(E/F)$ è l'insieme delle applicazioni $f(t)/g(t) \mapsto f(u)/g(u)$ con u come in (15). Si vede così che $\mathrm{Gal}(E/F)$ è isomorfo al quoziente $\mathrm{GL}_2(F)/F^*$, dove $\mathrm{GL}_2(F)$ è il gruppo delle matrici invertibili 2×2 su F e F^* è l'insieme delle matrici $\mathrm{diag}(a, a)$, $a \neq 0$.

Sia ora G un gruppo di automorfismi di un campo E, ossia un sottogruppo del gruppo degli automorfismi di E. Poniamo

$$\mathrm{Inv}\, G = \{a \in E \mid \eta(a) = a \text{ per ogni } \eta \in G\}.$$

Si tratta degli elementi di E lasciati fissi da tutti gli $\eta \in G$. Le proprietà

$$\eta(a + b) = \eta(a) + \eta(b), \qquad \eta(ab) = \eta(a)\eta(b), \qquad \eta(1) = 1, \qquad \eta(a^{-1}) = \eta(a)^{-1} \quad (a \neq 0)$$

mostrano che $\mathrm{Inv}\, G$ è un sottocampo di E. Lo chiamiamo *sottocampo degli invarianti di G* o *campo fisso di G* in E.

Fissato il campo E, le definizioni forniscono due applicazioni:

$$G \longmapsto \operatorname{Inv} G, \qquad F \longmapsto \operatorname{Gal}(E/F).$$

La prima manda i gruppi di automorfismi di E nei sottocampi di E, la seconda i sottocampi nei gruppi di automorfismi. Le loro proprietà fondamentali sono

$$\begin{aligned}
&\text{(i)} \quad G_1 \supseteq G_2 \implies \operatorname{Inv} G_1 \subseteq \operatorname{Inv} G_2,\\
&\text{(ii)} \quad F_1 \supseteq F_2 \implies \operatorname{Gal}(E/F_1) \subseteq \operatorname{Gal}(E/F_2),\\
&\text{(iii)} \quad \operatorname{Inv}(\operatorname{Gal}(E/F)) \supseteq F,\\
&\text{(iv)} \quad \operatorname{Gal}(E/\operatorname{Inv} G) \supseteq G.
\end{aligned}$$

Sono conseguenze immediate delle definizioni, la cui verifica è lasciata al lettore.

Applichiamo queste idee ai campi di spezzamento. Con la terminologia attuale possiamo riformulare le osservazioni successive al teorema 4.4. Se E è un campo di spezzamento di $f(x)$ su F, allora $\operatorname{Gal}(E/F)$ è finito e $|\operatorname{Gal}(E/F)| \leq [E:F]$. Vale l'uguaglianza quando $f(x)$ ha radici distinte. Nel §4.4 abbiamo visto che si può sostituire $f(x)$ con il prodotto $f_1(x)$ dei suoi fattori primi distinti e che, se $f(x)$ è separabile, $f_1(x)$ ha radici distinte. Otteniamo così il seguente risultato preliminare.

Lemma 1. *Se E/F è un campo di spezzamento di un polinomio separabile di $F[x]$, allora*

$$|\operatorname{Gal}(E/F)| = [E:F]. \tag{16}$$

Affrontiamo ora la questione dal lato dei gruppi. Partiamo da un campo arbitrario E e da un gruppo finito G di automorfismi di E.

Lemma 2 (Artin). *Se G è un gruppo finito di automorfismi di un campo E e $F = \operatorname{Inv} G$, allora*

$$[E:F] \leq |G|. \tag{17}$$

Dimostrazione. Sia $n = |G|$. Per dimostrare (17) basta provare che comunque si prendano $m > n$ elementi di E, essi sono linearmente dipendenti su F. Ci serviremo del noto risultato di algebra lineare secondo il quale ogni sistema di n equazioni lineari omogenee in $m > n$ incognite, con coefficienti in un campo E, ammette una soluzione non banale in E. Il lettore probabilmente conosce questo teorema, spesso usato per dimostrare l'invarianza della dimensione di uno spazio vettoriale di dimensione finita. Per completezza ne daremo una dimostrazione subito dopo questa.

Siano $G = \{\eta_1 = 1, \eta_2, \ldots, \eta_n\}$ e $u_1, \ldots, u_m$ elementi di E, con $m > n$. Il teorema sulle equazioni lineari assicura una soluzione non banale $(a_1, \ldots, a_m)$ del sistema di n equazioni

$$\sum_{j=1}^{m} \eta_i(u_j) x_j = 0 \qquad (1 \leq i \leq n) \tag{18}$$

nelle m incognite x_j. Non banale significa diversa da $(0, \ldots, 0)$. Tra queste soluzioni scegliamone una $(b_1, \ldots, b_m)$ con il minor numero di componenti diverse da zero. Riordinando le incognite si può supporre $b_1 \neq 0$. Anche $b_1^{-1}(b_1, \ldots, b_m)$ è soluzione, perciò possiamo supporre $b_1 = 1$. Sosteniamo che ogni b_j appartiene a $F = \operatorname{Inv} G$. Questo dimostrerà la dipendenza degli u_i su F, dato che, essendo $\eta_1 = 1$, la prima delle (18) è $\sum_j u_j x_j = 0$.

Se qualche b_j non appartenesse a F, potremmo supporre che sia b_2 e scegliere $\eta_k \in G$ con $\eta_k(b_2) \neq b_2$. Applichiamo η_k alle equazioni $\sum_j \eta_i(u_j)b_j = 0$. Otteniamo $\sum_j (\eta_k\eta_i)(u_j)\eta_k(b_j) = 0$. Poiché $(\eta_k\eta_1, \ldots, \eta_k\eta_n)$ è una permutazione di $(\eta_1, \ldots, \eta_n)$, si ha anche $\sum_j \eta_i(u_j)\eta_k(b_j) = 0$ per ogni i. Pertanto $(1, \eta_k(b_2), \ldots, \eta_k(b_m))$ è un'altra soluzione di (18). Sottraendola da $(1, b_2, \ldots, b_m)$ si ottiene la soluzione $(0, b_2 - \eta_k(b_2), \ldots, b_m - \eta_k(b_m))$. Essa è non banale perché $b_2 - \eta_k(b_2) \neq 0$, ma ha meno componenti non nulle della soluzione prescelta. La contraddizione conclude la dimostrazione, salvo il lemma che segue. □

Lemma sulle equazioni lineari. *Un sistema $\sum_{j=1}^{m} a_{ij}x_j = 0$ $(1 \leq i \leq n)$ di $n < m$ equazioni lineari omogenee a coefficienti a_{ij} in un campo E ammette una soluzione $(a_1, \ldots, a_m) \neq (0, \ldots, 0)$ con $a_i \in E$.*

Dimostrazione. Il risultato è banale se tutti gli a_{ij} sono nulli. Altrimenti, riordinando equazioni e incognite, possiamo supporre $a_{nm} \neq 0$. Dall'i-esima equazione, per $1 \leq i \leq n-1$, sottraiamo l'ultima moltiplicata per $a_{im}a_{nm}^{-1}$. Il sistema equivalente così ottenuto ha la forma

$$\sum_{j=1}^{m-1} a'_{ij}x_j = 0 \quad (1 \leq i \leq n-1), \qquad \sum_{j=1}^{m} a_{nj}x_j = 0.$$

Supponendo vero il risultato per $n-1$, le prime $n-1$ equazioni hanno una soluzione non banale $(a_1, \ldots, a_{m-1})$. Ponendo $a_m = -a_{nm}^{-1}\sum_{j=1}^{m-1} a_{nj}a_j$ otteniamo una soluzione non banale del secondo sistema e quindi del primo. Il caso $n = 1$ è immediato e il lemma segue per induzione. □

Introduciamo due nozioni relative ai campi, legate alle analoghe nozioni già introdotte per i polinomi. Un'estensione E/F si dice algebrica se ogni elemento di E è algebrico su F. Ciò avviene certamente se E ha dimensione finita su F, poiché per t trascendente $F(t)$ ha dimensione infinita. Diremo *separabile* un'estensione algebrica E/F se il polinomio minimo di ciascun elemento di E è separabile. L'estensione algebrica E/F si dice *normale* se ogni polinomio irriducibile in $F[x]$ che abbia una radice in E è prodotto di fattori lineari in $E[x]$. Equivalentemente, E contiene un campo di spezzamento del polinomio minimo di ciascuno dei suoi elementi. Normalità e separabilità insieme significano che ogni irriducibile di $F[x]$ con una radice in E si spezza in fattori lineari distinti in $E[x]$. Per i risultati della sezione precedente, ogni estensione algebrica di F è separabile se la caratteristica è zero, oppure se è $p \neq 0$ e $F^p = F$.

Il primo dei nostri risultati principali dà due caratterizzazioni astratte dei campi di spezzamento dei polinomi separabili e ulteriori informazioni importanti.

Teorema 4.7. *Sia E un'estensione di un campo F. Le seguenti condizioni su E/F sono equivalenti.*

1. *E è un campo di spezzamento su F di un polinomio separabile $f(x)$.*
2. *$F = \operatorname{Inv} G$ per qualche gruppo finito G di automorfismi di E.*
3. *E ha dimensione finita su F ed è normale e separabile su F.*

Inoltre, nelle ipotesi di (1), se $G = \operatorname{Gal}(E/F)$ si ha $F = \operatorname{Inv} G$; nelle ipotesi di (2) si ha $G = \operatorname{Gal}(E/F)$.

Dimostrazione. (1) ⇒ (2). Poniamo $G = \operatorname{Gal}(E/F)$ e $F' = \operatorname{Inv} G$. Allora $F \subseteq F' \subseteq E$. Il campo E è campo di spezzamento di $f(x)$ anche su F' e $G = \operatorname{Gal}(E/F')$. Dal lemma 1 segue $|G| = [E : F] = [E : F']$. Poiché $[E : F] = [E : F'][F' : F]$, si ha $[F' : F] = 1$, cioè $F' = F$. Abbiamo così provato anche la prima affermazione supplementare.

(2) $\Rightarrow$ (3). Per il lemma di Artin $[E : F] \leq |G|$, quindi E ha dimensione finita su F. Sia $f(x)$ irriducibile monico in $F[x]$ con una radice r in E. Sia $\{r_1 = r, r_2, \ldots, r_m\}$ l'orbita di r sotto G, ossia l'insieme degli elementi distinti della forma $\eta(r)$, $\eta \in G$. Per $\zeta \in G$, la successione $(\zeta(r_1), \ldots, \zeta(r_m))$ è una permutazione degli r_i. Da $f(r) = 0$ segue $f(r_i) = 0$. Quindi $x - r_i$ divide $f(x)$ e, poiché gli r_i sono distinti, anche $g(x) = \prod_{i=1}^{m}(x - r_i)$ divide $f(x)$. Applichiamo a $g(x)$ l'automorfismo di $E[x]$ che fissa x e manda $a \in E$ in $\zeta(a)$. Si ottiene $\prod_i(x - \zeta(r_i)) = \prod_i(x - r_i) = g(x)$. Poiché ciò vale per ogni $\zeta \in G$, i coefficienti di $g(x)$ sono invarianti e $g(x) \in F[x]$. Dall'irriducibilità di $f(x)$ in $F[x]$ segue $f(x) = g(x) = \prod_i(x - r_i)$, prodotto di fattori lineari distinti in $E[x]$. Dunque E è separabile e normale su F.

(3) $\Rightarrow$ (1). Poiché $[E : F] < \infty$, scriviamo $E = F(r_1, \ldots, r_k)$ con ciascun r_i algebrico su F. Per ipotesi il suo polinomio minimo $f_i(x)$ è prodotto di fattori lineari distinti in $E[x]$. Pertanto $f(x) = \prod_{i=1}^{k} f_i(x)$ è separabile e $E = F(r_1, \ldots, r_k)$ ne è campo di spezzamento su F. Vale (1).

Resta la seconda affermazione supplementare. Nelle ipotesi di (2) abbiamo visto che $[E : F] \leq |G|$. Dall'equivalenza già provata e dal lemma 1 segue $|\operatorname{Gal}(E/F)| = [E : F]$. Poiché $G \subseteq \operatorname{Gal}(E/F)$ e $|G| \geq [E : F] = |\operatorname{Gal}(E/F)|$, si ha $G = \operatorname{Gal}(E/F)$. $\square$

Possiamo ora stabilire la corrispondenza fondamentale di Galois fra gruppi e campi.

Teorema fondamentale della teoria di Galois. *Sia E/F un'estensione che soddisfa una, e quindi tutte, le condizioni equivalenti del teorema 4.7. Sia G il gruppo di Galois di E su F. Indichiamo con $\Gamma = \{H\}$ l'insieme dei sottogruppi di G e con Σ quello dei campi intermedi tra E e F, cioè dei sottocampi di E/F. Le applicazioni $H \mapsto \operatorname{Inv} H$, $K \mapsto \operatorname{Gal}(E/K)$, per $H \in \Gamma$, $K \in \Sigma$, sono inverse l'una dell'altra e quindi biiezioni tra Γ e Σ. La corrispondenza ha le seguenti proprietà:*

$$
\begin{aligned}
&(\alpha) \quad H_1 \supseteq H_2 \iff \operatorname{Inv} H_1 \subseteq \operatorname{Inv} H_2;\\
&(\beta) \quad |H| = [E : \operatorname{Inv} H], \qquad [G : H] = [\operatorname{Inv} H : F];\\
&(\gamma) \quad H \text{ è normale in } G \iff \operatorname{Inv} H \text{ è normale su } F.
\end{aligned}
$$

In quest'ultimo caso $\operatorname{Gal}((\operatorname{Inv} H)/F) \cong G/H$.

Dimostrazione. Sia H un sottogruppo di $G = \operatorname{Gal}(E/F)$. Poiché $F = \operatorname{Inv} G$, il sottocampo $K = \operatorname{Inv} H$ contiene F. La seconda affermazione supplementare del teorema 4.7, applicata a H, dà $\operatorname{Gal}(E/\operatorname{Inv} H) = H$. Parimenti $|H| = |\operatorname{Gal}(E/\operatorname{Inv} H)| = [E : \operatorname{Inv} H]$. Viceversa, per ogni campo intermedio K, il gruppo $H = \operatorname{Gal}(E/K)$ è un sottogruppo di G. Poiché E è campo di spezzamento di un polinomio separabile su F, lo è anche su K. La prima affermazione supplementare del teorema 4.7, applicata a E/K, dà $K = \operatorname{Inv} H = \operatorname{Inv}(\operatorname{Gal}(E/K))$. Abbiamo provato che le due applicazioni sono inverse.

Sappiamo già che $H_1 \supseteq H_2$ implica $\operatorname{Inv} H_1 \subseteq \operatorname{Inv} H_2$. Viceversa, da quest'ultima inclusione segue $H_1 = \operatorname{Gal}(E/\operatorname{Inv} H_1) \supseteq \operatorname{Gal}(E/\operatorname{Inv} H_2) = H_2$, il che prova (α). La prima parte di (β) è stata osservata sopra. Poiché

$$|G| = [E : F] = [E : \operatorname{Inv} H][\operatorname{Inv} H : F] = |H|[\operatorname{Inv} H : F] = |H|[G : H],$$

si ha anche $[\operatorname{Inv} H : F] = [G : H]$.

Se $K = \operatorname{Inv} H$ corrisponde al sottogruppo H, al sottogruppo coniugato $\eta H \eta^{-1}$ corrisponde $\eta(K)$: infatti $\zeta(k) = k$ equivale a $(\eta\zeta\eta^{-1})(\eta(k)) = \eta(k)$. Pertanto H è normale in G se e solo se $\eta(K) = K$ per ogni $\eta \in G$. Supponiamo ciò verificato. Allora la restrizione $\bar{\eta} = \eta|_K$ è un automorfismo di K/F e abbiamo l'omomorfismo di restrizione $\eta \mapsto \bar{\eta}$ da G in $\operatorname{Gal}(K/F)$. L'immagine $\bar{G}$ ha campo fisso F e dunque $\bar{G} = \operatorname{Gal}(K/F)$. Il nucleo è costituito dagli $\eta \in G$

tali che $\eta|_K = 1_K$, ossia, per la corrispondenza, è $\mathrm{Gal}(E/K) = H$. Segue $\mathrm{Gal}(K/F) \cong G/H$. Poiché $F = \mathrm{Inv}\, \bar{G}$, il teorema 4.7 assicura inoltre che K è normale su F.

Viceversa, supponiamo K normale su F. Per $a \in K$ il polinomio minimo $f(x)$ di a su F si scrive $(x - a_1) \cdots (x - a_m)$ in $K[x]$, con $a_1 = a$. Per ogni $\eta \in G$, $f(\eta(a)) = 0$, perciò $\eta(a) = a_i$ per qualche i, e $\eta(a) \in K$. Così $\eta(K) \subseteq K$ per ogni η; applicando lo stesso fatto a η^{-1} si ottiene l'uguaglianza. Per quanto visto sopra, H è normale in G. Con ciò è provato (γ). □

Come primo esempio consideriamo il campo delle radici diciassettesime dell'unità. Si chiariranno così i calcoli di $\cos(2\pi/17)$ delle pp. 222–224 dell'originale e la ragione del loro successo.

Esempio. Siano $F = \mathbb{Q}$ ed E il campo delle radici diciassettesime dell'unità, ossia un campo di spezzamento di $x^{17} - 1$ su $\mathbb{Q}$. Poiché la derivata $17x^{16}$ è prima con $x^{17} - 1$, le radici sono distinte. Esse costituiscono un sottogruppo ciclico U del gruppo moltiplicativo di E. Se z ne è un generatore, $U = \{z, z^2, \ldots, z^{17} = 1\}$ ed $E = \mathbb{Q}(z)$. Il polinomio minimo di z su $\mathbb{Q}$ è $x^{16} + x^{15} + \cdots + 1$, che è irriducibile. Dunque $[E : \mathbb{Q}] = 16$ e $G = \mathrm{Gal}(E/\mathbb{Q})$ ha ordine 16. Ogni $\eta \in G$ manda U in sé e induce un automorfismo di $U = \langle z \rangle$. L'omomorfismo $\eta \mapsto \bar{\eta} = \eta|_U$ da G in $\mathrm{Aut}\, U$ è un monomorfismo: se $\bar{\eta} = 1$, allora $\eta(z) = z$ e quindi $\eta = 1$ perché $E = \mathbb{Q}(z)$. Sappiamo che il gruppo degli automorfismi del gruppo ciclico di ordine n è isomorfo al gruppo delle unità di $\mathbb{Z}/(n)$ e ha ordine $\varphi(n)$. Nel nostro caso $\mathrm{Aut}\, U$ ha ordine 16 ed è isomorfo al gruppo moltiplicativo del campo $\mathbb{Z}/(17)$, ciclico con generatore $\bar{3} = 3 + (17)$. Dal confronto degli ordini segue che $\mathrm{Gal}(E/\mathbb{Q})$ è isomorfo a quest'ultimo gruppo. L'automorfismo η di $E/\mathbb{Q}$ che manda z in z^3 genera pertanto $G = \{\eta, \eta^2, \ldots, \eta^{16} = 1\}$. I sottogruppi sono

$$G = G_1 = \langle \eta \rangle \supset G_2 = \langle \eta^2 \rangle \supset G_3 = \langle \eta^4 \rangle \supset G_4 = \langle \eta^8 \rangle \supset G_5 = \{1\},$$

dove $\langle \eta^i \rangle$ indica il sottogruppo generato da η^i. I rispettivi ordini sono $16, 8, 4, 2, 1$. Per la corrispondenza di Galois a questi sottogruppi corrispondono nel medesimo ordine i campi

$$F = F_1 \subset F_2 \subset F_3 \subset F_4 \subset F_5 = E. \tag{19}$$

Qual è F_i? Riprendiamo le notazioni introdotte alla fine del §4.2. Come si è osservato, $(z, z^2, \ldots, z^{16})$ è una base di $E/\mathbb{Q}$, poiché $x^{16} + x^{15} + \cdots + 1$ è il polinomio minimo di z. Si ha $\eta(z) = z^3$, $\eta^i(z) = z^{3^i}$. Ponendo $x_1 = \sum_{i=1}^{8} \eta^{2i}(z)$ si ha $\eta^2(x_1) = x_1$, $\eta(x_1) \neq x_1$, per cui $x_1 \in F_2 \setminus F_1$. Poiché $[G : G_2] = [G_1 : G_2] = 2$, si ha $[F_2 : F_1] = 2$ e $F_2 = F_1(x_1)$. Analogamente, se $y_1 = \sum_{i=1}^{4} \eta^{4i}(z)$ e $z_1 = \sum_{i=1}^{2} \eta^{8i}(z)$, si ottiene $F_3 = F_2(y_1)$, $F_4 = F_3(z_1)$. La catena (19) è dunque

$$F \subset F(x_1) \subset F(x_1, y_1) \subset F(x_1, y_1, z_1) \subset E. \tag{19'}$$

I calcoli fatti in precedenza consistevano nel determinare i polinomi minimi di x_1, y_1, z_1, z rispettivamente su $F_1, F_2, \ldots$ e nel trovarne le radici mediante equazioni di secondo grado. Inoltre il gruppo G è abeliano, tutti i suoi sottogruppi sono normali e quindi ogni sottocampo di E/F è normale su F.

Come seconda illustrazione della corrispondenza di Galois ricaviamo la teoria delle espressioni razionali simmetriche, affine ai risultati sui polinomi simmetrici del §2.13. Dato un campo F, consideriamo il campo delle frazioni $F(x_1, \ldots, x_n)$ dell'anello $F[x_1, \ldots, x_n]$ nelle indeterminate x_i. Ricordiamo che per ogni permutazione π di $\{1, \ldots, n\}$ vi è un unico automorfismo $\zeta(\pi)$ di $F[x_1, \ldots, x_n]$ che fissa gli elementi di F e manda x_i in $x_{\pi(i)}$ $(1 \leq i \leq n)$ (teorema 2.12, p. 125 dell'originale). Esso si prolunga in un unico modo a un automorfismo del campo $F(x_1, \ldots, x_n)$, che indicheremo ancora con $\zeta(\pi)$. Per due permutazioni π_1, π_2 si ha $\zeta(\pi_1\pi_2) = \zeta(\pi_1)\zeta(\pi_2)$ nell'anello dei polinomi e quindi anche nel campo delle frazioni. Gli automorfismi $\zeta(\pi)$ costituiscono

perciò un gruppo G di automorfismi del campo, isomorfo al gruppo simmetrico S_n. Gli elementi che G lascia fissi si chiamano *espressioni razionali simmetriche* e $\operatorname{Inv} G$ è il loro campo. Per determinarlo mediante la corrispondenza di Galois, poniamo $E = F(x_1, \dots, x_n)$ e introduciamo in $E[x]$ il polinomio

$$g(x) = (x - x_1)(x - x_2) \cdots (x - x_n), \tag{20}$$

che si può scrivere

$$g(x) = x^n - p_1 x^{n-1} + p_2 x^{n-2} - \cdots + (-1)^n p_n, \tag{21}$$

dove

$$p_1 = \sum_{i=1}^{n} x_i, \quad p_2 = \sum_{i<j} x_i x_j, \quad \dots, \quad p_n = x_1 x_2 \cdots x_n. \tag{22}$$

Prolunghiamo $\zeta(\pi)$ all'automorfismo $\zeta'(\pi)$ di $E[x]$ che fissa x. Esso manda $g(x)$ nel prodotto $\prod_i (x - x_{\pi(i)})$, uguale a $g(x)$ poiché π permuta gli indici. Quindi per ogni π si ha $\zeta'(\pi)(g(x)) = g(x)$ e $\zeta(\pi)(p_i) = p_i$ per $1 \le i \le n$. I p_i appartengono a $\operatorname{Inv} G$ e così pure il campo $F(p_1, \dots, p_n)$ che essi generano su F. D'altra parte, da $E = F(x_1, \dots, x_n) = F(p_1, \dots, p_n, x_1, \dots, x_n)$ e da (20) risulta che E è campo di spezzamento su $F(p_1, \dots, p_n)$ di $g(x)$, le cui radici sono distinte. Se $\zeta \in \operatorname{Gal}(E/F(p_1, \dots, p_n))$, applicando ζ a $g(x_i) = 0$ si ottiene $g(\zeta(x_i)) = 0$. Perciò ζ permuta gli x_i e coincide con uno degli $\zeta(\pi)$. Dunque $\operatorname{Gal}(E/F(p_1, \dots, p_n)) = G$. La corrispondenza di Galois associa ai sottogruppi di G i sottocampi di E che contengono $F(p_1, \dots, p_n)$. In particolare associa G a $F(p_1, \dots, p_n)$; siccome al sottogruppo H corrisponde $\operatorname{Inv} H$, ne segue $\operatorname{Inv} G = F(p_1, \dots, p_n)$. Abbiamo provato l'analogo, per i campi, della prima parte del teorema 2.20: ogni espressione razionale simmetrica nelle indeterminate x_i è un'espressione razionale nei polinomi simmetrici elementari $p_1, \dots, p_n$. Il risultato si deduce facilmente anche dal teorema 2.20.

Esercizi.

1. Nella corrispondenza del teorema fondamentale, dimostrare che all'intersezione di due sottogruppi H_1, H_2 corrisponde il campo generato dai campi intermedi $\operatorname{Inv} H_1, \operatorname{Inv} H_2$ e che all'intersezione di due campi intermedi K_1, K_2 corrisponde il sottogruppo generato da $\operatorname{Gal}(E/K_1)$ e $\operatorname{Gal}(E/K_2)$.

2. Siano E, F, G come nel teorema fondamentale. Supponiamo che E sia generato da due campi intermedi K, L con $K \cap L = F$ ed L/F normale. Posto $N = \operatorname{Gal}(E/L)$, $H = \operatorname{Gal}(E/K)$, mostrare che N è normale in G, $H \cap N = 1$ e $G = HN$, sicché G è il prodotto semidiretto di N e H (esercizio 9, p. 79 dell'originale). Mostrare inoltre che se K/F ed L/F sono normali allora $G = H \times N$.

3. Sia $E = \mathbb{Q}(r)$, dove $r^3 + r^2 - 2r - 1 = 0$. Verificare che anche $r' = r^2 - 2$ è una radice di $x^3 + x^2 - 2x - 1 = 0$. Determinare $\operatorname{Gal}(E/\mathbb{Q})$. Mostrare che E è normale su $\mathbb{Q}$.

4. Sia E un campo di spezzamento di $x^5 - 2$ su $\mathbb{Q}$. Determinare il gruppo di Galois di $E/\mathbb{Q}$ e dimostrare che è isomorfo all'olomorfo di un gruppo ciclico di ordine 5 (p. 63 dell'originale). Determinarne i sottogruppi e i sottocampi corrispondenti.

5. Sia F un campo di caratteristica p e sia $a \in F$ un elemento che non sia della forma $b^p - b$, $b \in F$. Determinare il gruppo di Galois su F di un campo di spezzamento di $x^p - x - a$.

6. Sia $E = \mathbb{C}(t)$, con t trascendente su $\mathbb{C}$, e sia $w \in \mathbb{C}$ tale che $w^3 = 1$, $w \neq 1$. Siano σ, τ gli automorfismi di $E/\mathbb{C}$ tali che $\sigma(t) = wt$, $\tau(t) = t^{-1}$. Mostrare che $\sigma^3 = 1 = \tau^2$ e $\tau\sigma = \sigma^{-1}\tau$. Mostrare che il gruppo G generato da σ, τ ha ordine 6 e che $F = \operatorname{Inv} G = \mathbb{C}(u)$, dove $u = t^3 + t^{-3}$.

7. Sia $E = (\mathbb{Z}/(p))(t)$ con t trascendente su $\mathbb{Z}/(p)$. Sia G il gruppo generato dall'automorfismo di E dato da $t \mapsto t + 1$. Determinare $F = \operatorname{Inv} G$ e $[E : F]$.

8. Ripetere l'esercizio 7 per il gruppo degli automorfismi $t \mapsto at + b$, con $a, b \in \mathbb{Z}/(p)$, $a \neq 0$.

9. Mostrare che $E = \mathbb{Q}(\sqrt{2}, \sqrt{3}, u)$, dove $u^2 = (9-5\sqrt{3})(2-\sqrt{2})$, è normale su $\mathbb{Q}$. Determinare $\operatorname{Gal}(E/\mathbb{Q})$.

10. Con il metodo della dimostrazione del lemma di Artin (p. 236 dell'originale) provare il seguente risultato sulle equazioni differenziali. Se $y_1, \ldots, y_{n+1}$ sono funzioni analitiche reali che soddisfano l'equazione differenziale lineare $y^{(n)} + a_1 y^{(n-1)} + \cdots + a_n y = 0$ con coefficienti costanti $a_i \in \mathbb{R}$, allora esse sono linearmente dipendenti su $\mathbb{R}$.

11. Sia $E = F(t)$ con t trascendente su F e si scriva ogni elemento non nullo u nella forma $f(t)/g(t)$ con $(f(t), g(t)) = 1$. Si chiami grado di u il massimo dei gradi di f e g. Per u non costante e per indeterminate x, y, mostrare che $f(x) - yg(x)$ è irriducibile in $F[x, y]$ e quindi in $F(y)[x]$. Mostrare che t è algebrico su $F(u)$ e che il suo polinomio minimo è il multiplo monico di $f(x) - ug(x)$ per uno scalare di $F(u)$. Concludere che $[F(t) : F(u)] = \deg u$ e che $F(u) = F(t)$ se e solo se $\deg u = 1$. Osservare che in tal caso $u = (at + b)/(ct + d)$ con $ad - bc \neq 0$. Dedurne che $\operatorname{Gal}(E/F)$ è l'insieme delle applicazioni $h(t) \mapsto h(u)$ con u della forma indicata.

12. Sia $E = F(x_1, \ldots, x_n)$, dove gli x_i sono indeterminate, e sia $\zeta(\pi)$ definito come nel testo per ogni permutazione π di $\{1, \ldots, n\}$. Si scriva un elemento di E nella forma $f(x_1, \ldots, x_n)/g(x_1, \ldots, x_n)$ con f, g privi di fattori comuni. Mostrare che, se questo elemento è simmetrico, sono simmetrici sia il numeratore sia il denominatore. Servirsene per dedurre dal teorema 2.20 che $f/g \in F(p_1, \ldots, p_n)$ con i p_i definiti sopra.

13. Sia $\operatorname{char} F \neq 2$ e sia H il sottogruppo di $G = \operatorname{Gal}(E/F(p_1, \ldots, p_n))$ corrispondente al gruppo alterno, cioè costituito dagli $\zeta(\pi)$ con $\pi \in A_n$. Mostrare che $\operatorname{Inv} H = F(p_1, \ldots, p_n, \Delta)$, dove $\Delta = \prod_{i<j}(x_i - x_j)$.

4.6 Alcuni risultati sui gruppi finiti

Apriamo una breve parentesi per stabilire alcuni risultati sui gruppi finiti necessari alla teoria delle equazioni. Essi riguardano soprattutto una classe di gruppi, detti risolubili, che, come vedremo nella prossima sezione, corrispondono alle equazioni risolubili per radicali.

Dato un gruppo G, introduciamo la notazione consueta $G \triangleright H$ o $H \triangleleft G$ per indicare che H è un sottogruppo normale di G. Una successione di sottogruppi

$$G = G_1 \triangleright G_2 \triangleright \cdots \triangleright G_s \triangleright G_{s+1} = 1 \tag{23}$$

si chiama *serie normale* di G. La notazione indica che G_{i+1} è normale in G_i, senza che sia necessariamente normale in G. Per esempio, $S_3 \triangleright A_3 \triangleright 1$ e $S_4 \triangleright A_4 \triangleright V \triangleright W \triangleright 1$ dove

$$V = \{1, (12)(34), (13)(24), (14)(23)\}, \qquad W = \{1, (12)(34)\}.$$

Lasciamo al lettore verificare che V è normale in A_4. Poiché V è abeliano, W è normale in V. Alla serie normale (23) associamo la *successione dei fattori*

$$G_1/G_2,\ G_2/G_3,\ \ldots,\ G_s/G_{s+1} \cong G_s. \tag{24}$$

Un gruppo si dice *risolubile* se ammette una serie normale i cui fattori siano tutti abeliani. Le serie date per S_3 e S_4 mostrano che questi gruppi sono risolubili. Infatti S_3/A_3 è ciclico di ordine 2, A_3 ciclico di ordine 3, S_4/A_4 ciclico di ordine 2, A_4/V ciclico di ordine 3 e V/W, W ciclici di ordine 2. Naturalmente ogni gruppo abeliano è risolubile, poiché $G \triangleright 1$ è una serie normale di fattore abeliano $G = G/1$. Il teorema seguente fornisce un'altra classe importante.

Teorema 4.8. *Ogni gruppo finito il cui ordine sia potenza di un primo è risolubile.*

Dimostrazione. Sia G un p-gruppo, ossia un gruppo di ordine p^n con $n \geq 1$ e p primo. Abbiamo visto che il centro C di G non è banale. Se $G \neq C$, poniamo $C_1 = C$. Il quoziente G/C_1 è un p-gruppo, dunque ha centro non banale, della forma C_2/C_1 con C_2 normale in G. Se $C_2 \neq G$, sia C_3 il sottogruppo tale che C_3/C_2 sia il centro di G/C_2. Procedendo così otteniamo la successione di sottogruppi normali $1 \subsetneq C_1 \subsetneq C_2 \subsetneq C_3 \subsetneq \cdots$. Poiché G è finito, si giunge a $G = C_{s+1}$. Allora $G = C_{s+1} \triangleright C_s \triangleright \cdots \triangleright C_1 \triangleright 1$ è una serie normale con fattori abeliani C_{i+1}/C_i. □

Stabiliamo alcune proprietà fondamentali dei gruppi risolubili. Cominciamo da un criterio che usa una particolare serie di sottogruppi normali, la *serie derivata*. Per $g, h \in G$ si definisce *commutatore* di g e h l'elemento $(g,h) = g^{-1}h^{-1}gh$. Poiché $gh = hg(g,h)$, il commutatore misura quanto g e h si discostino dalla commutatività. Il *sottogruppo derivato* (o *dei commutatori*) G' è il sottogruppo generato da tutti i commutatori (g,h), $g, h \in G$. Siccome $(g,h)^{-1} = (g^{-1}h^{-1}gh)^{-1} = h^{-1}g^{-1}hg = (h,g)$, G' è l'insieme dei prodotti

$$(g_1,h_1)(g_2,h_2)\cdots(g_k,h_k), \qquad g_i, h_i \in G. \tag{25}$$

Per un omomorfismo $\eta : G \to \bar{G}$ si ha $\eta((g,h)) = \eta(g)^{-1}\eta(h)^{-1}\eta(g)\eta(h) = (\eta(g),\eta(h))$. Dunque $\eta(G') \subseteq \bar{G}'$. Se η è suriettivo, questa formula mostra che ogni commutatore $(\bar{g},\bar{h})$, $\bar{g},\bar{h} \in \bar{G}$, appartiene a $\eta(G')$, sicché $\eta(G') = \bar{G}'$. In particolare tali osservazioni valgono per ogni endomorfismo di G. Sia ora $K \triangleleft G$. Ogni automorfismo interno $I_a : x \mapsto axa^{-1}$ di G induce un endomorfismo di K. Quindi per ogni $a \in G$ si ha $I_a(K') \subseteq K'$, ossia K' è normale in G. In simboli,

$$K \triangleleft G \quad \Longrightarrow \quad K' \triangleleft G. \tag{26}$$

Poiché $G \triangleleft G$, abbiamo $G' \triangleleft G$.

Definiamo il secondo gruppo derivato $G'' = (G')'$ e, iterando, $G^{(k)} = (G^{(k-1)})'$ per $k \geq 1$, con $G^{(0)} = G$. Per induzione, da (26) segue $G^{(k)} \triangleleft G$. In particolare $G \triangleright G' \triangleright G'' \triangleright \cdots$. Mostreremo che G è risolubile se e solo se, per qualche $k \geq 1$,

$$G^{(k)} = 1. \tag{27}$$

Ciò segue dalla seguente caratterizzazione del gruppo derivato.

Lemma. *G/G' è abeliano e G' è contenuto in ogni sottogruppo normale K tale che G/K sia abeliano.*

Dimostrazione. Per definizione G è abeliano se e solo se $G' = 1$. Per $g, h \in G$ e $K \triangleleft G$, $(gK, hK) = (gK)^{-1}(hK)^{-1}gKhK = (g,h)K$. Quindi $(gK,hK) = 1$ (cioè K) se e solo se $(g,h) \in K$. Dunque G/K è abeliano se e solo se K contiene ogni commutatore (g,h), ossia se e solo se $K \supseteq G'$. Ne seguono entrambe le conclusioni. □

Teorema 4.9. *Un gruppo G è risolubile se e solo se $G^{(k)} = 1$ per qualche $k \geq 1$.*

Dimostrazione. Se la condizione vale, la serie normale $G \triangleright G' \triangleright G'' \triangleright \cdots \triangleright G^{(k)} = 1$ ha fattori $G^{(i)}/G^{(i+1)}$ abeliani per il lemma. Dunque G è risolubile. Viceversa, sia $G = G_1 \triangleright G_2 \triangleright \cdots \triangleright G_s \triangleright G_{s+1} = 1$ una serie normale con fattori abeliani. Dal lemma $G_{i+1} \supseteq G_i'$, perché G_i/G_{i+1} è abeliano. In particolare $G_2 \supseteq G_1' = G'$. Se $G_k \supseteq G^{(k-1)}$, si ha $G_{k+1} \supseteq G_k' \supseteq (G^{(k-1)})' = G^{(k)}$. Per induzione $G_{i+1} \supseteq G^{(i)}$ e pertanto $G^{(s)} = 1$. □

Dal criterio appena stabilito si ottiene facilmente il seguente risultato.

Teorema 4.10. *Ogni sottogruppo e ogni immagine omomorfa di un gruppo risolubile sono risolubili. Se $K \triangleleft G$ e K, G/K sono risolubili, anche G è risolubile.*

Dimostrazione. Per un sottogruppo $H \subseteq G$ si ha evidentemente $H^{(i)} \subseteq G^{(i)}$. Quindi $G^{(k)} = 1$ implica $H^{(k)} = 1$, e dalla risolubilità di G segue quella di H. Sia $\eta : G \to H$ un omomorfismo suriettivo. Poiché $\eta(G') = (\eta(G))'$, la restrizione di η manda G' suriettivamente su $(\eta(G))'$. Quindi $\eta(G'') = \eta((G')') = (\eta(G'))' = (\eta(G))''$. Per induzione $\eta(G^{(i)}) = (\eta(G))^{(i)}$. Da $G^{(k)} = 1$ segue $(\eta(G))^{(k)} = 1$: ogni immagine omomorfa di un gruppo risolubile è risolubile. Infine, supponiamo $K \triangleleft G$ e G/K risolubile. Per l'omomorfismo naturale suriettivo $\nu : G \to G/K$ si ha $\nu(G^{(i)}) = (G/K)^{(i)}$. Per un opportuno $k \geq 1$ si ha dunque $\nu(G^{(k)}) = 1$, ossia $G^{(k)} \subseteq K$. Se anche K è risolubile, esiste $l \geq 1$ tale che $K^{(l)} = 1$. Allora $G^{(k+l)} \subseteq K^{(l)} = 1$ e G è risolubile. □

Un gruppo G si dice *semplice* se G e 1 sono i suoi soli sottogruppi normali. Poiché in un gruppo abeliano ogni sottogruppo è normale, un gruppo abeliano è semplice se e solo se non possiede sottogruppi diversi da sé stesso e da 1. Per un gruppo non banale ciò equivale a essere ciclico di ordine primo. Questa classe di gruppi semplici è generalmente considerata banale. Il seguente teorema dà la classe più elementare di gruppi semplici non abeliani.

Teorema 4.11. *A_n è semplice per $n \geq 5$.*[6]

Dimostrazione. Mostriamo che da $A_n \triangleright K \neq 1$ segue $K = A_n$. Basta dimostrare che K contiene un 3-ciclo, per esempio (123). Infatti, per qualsiasi 3-ciclo (ijk) scegliamo una permutazione della forma

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \cdots \\ i & j & k & l & m & \cdots \end{pmatrix}.$$

Possiamo sceglierla pari: se la prima scelta è dispari basta sostituirla con $(lm)\gamma$. Allora $(ijk) = \gamma(123)\gamma^{-1} \in K$. Dunque K contiene ogni 3-ciclo e, poiché i 3-cicli generano A_n (esercizio 2, p. 51 dell'originale), $K = A_n$.

Fra gli elementi di K diversi da 1 sia α uno con il maggior numero di punti fissi. Un punto fisso di α è un indice i, $1 \leq i \leq n$, tale che $\alpha(i) = i$. Affermiamo che α è un 3-ciclo. In caso contrario, scrivendolo come prodotto di cicli disgiunti e omettendo quelli di lunghezza uno, si avrebbe, rinominando gli indici,

$$\alpha = (123\cdots)\cdots, \tag{28}$$

oppure un prodotto di trasposizioni disgiunte

$$\alpha = (12)(34)\cdots. \tag{29}$$

Nel primo caso α muove almeno altri due punti, diciamo $4, 5$, poiché non è una delle permutazioni dispari $(123k)$. Siano $\beta = (345)$ e $\alpha_1 = \beta\alpha\beta^{-1}$. Nel caso (28) $\alpha_1 = (124\cdots)\cdots$, nel caso (29) $\alpha_1 = (12)(45)\cdots$. In entrambi i casi $\alpha_1 \neq \alpha$, quindi $\alpha_2 = \alpha_1\alpha^{-1} \neq 1$ e $\alpha_2 \in K$. Ogni indice

[6] Questo risultato è dovuto a Galois.

> 5 è fissato da β: se lo fissa α, lo fissa anche $\alpha_2 = \beta\alpha\beta^{-1}\alpha^{-1}$. Nel caso (28) inoltre $\alpha_2(2) = 2$. Siccome α muove $1, 2, 3, 4, 5$, α_2 ha più punti fissi di α, contro la scelta di α. Nel caso (29) si ha $\alpha_2(1) = 1$, $\alpha_2(2) = 2$, e di nuovo α_2 ha più punti fissi di α (anche se α fissava 5). La contraddizione dimostra che α è un 3-ciclo e $K = A_n$. □

Corollario. *Se $n \geq 5$, S_n non è risolubile.*

Dimostrazione. Altrimenti il suo sottogruppo A_n sarebbe risolubile. Avremmo allora $A_n' \subsetneq A_n$. Poiché A_n è semplice e $A_n' \triangleleft A_n$, ne seguirebbe $A_n' = 1$. Ma A_n non è abeliano, già per $n \geq 4$, perché (123) e (234) non commutano. □

Un altro criterio di risolubilità di un gruppo finito si basa su un'importante nozione della teoria dei gruppi finiti, quella di *serie di composizione*. Si chiama così una serie normale $G = G_1 \triangleright G_2 \triangleright \cdots \triangleright G_{s+1} = 1$ in cui G_{i+1} sia un sottogruppo normale massimale di G_i: non esiste un sottogruppo normale H di G_i tale che $G_i \supsetneq H \supsetneq G_{i+1}$. Se $K \triangleleft G$, l'applicazione $H \mapsto H/K$ è una biiezione tra i sottogruppi di G contenenti K e i sottogruppi di G/K, e in essa si corrispondono i sottogruppi normali. Pertanto K è normale massimale in G se e solo se G/K è semplice e diverso da 1. Una serie normale è dunque una serie di composizione se e solo se ogni suo fattore G_i/G_{i+1} è semplice e non banale. Questi fattori si chiamano *fattori di composizione* determinati dalla serie.

Ogni gruppo finito non banale $G = G_1$ contiene un sottogruppo normale massimale G_2 e, se questo è non banale, esso contiene un sottogruppo normale massimale G_3. Continuando, si vede che ogni gruppo finito ha una serie di composizione. I fattori sono determinati a meno di isomorfismo nel seguente senso preciso.

Teorema di Jordan–Hölder. *Siano $G = G_1 \triangleright G_2 \triangleright \cdots \triangleright G_{s+1} = 1$ e $G = H_1 \triangleright H_2 \triangleright \cdots \triangleright H_{t+1} = 1$ due serie di composizione di un gruppo finito G. Allora $s = t$ ed esiste una permutazione $i \mapsto i'$ di $1, \ldots, s$ tale che $G_i/G_{i+1} \cong H_{i'}/H_{i'+1}$ per ogni $1 \leq i \leq s$.*

Dimostrazione. Procediamo per induzione su $|G|$, distinguendo i casi (I) $G_2 = H_2$ e (II) $G_2 \neq H_2$. Nel caso I, le due serie che iniziano con G_2 e H_2 sono serie di composizione dello stesso gruppo, di ordine minore di $|G|$. Dunque $s - 1 = t - 1$ ed esiste una permutazione $i \mapsto i'$ di $2, \ldots, s$ tale che $G_i/G_{i+1} \cong H_{i'}/H_{i'+1}$. Poiché $G_1/G_2 = H_1/H_2$, segue la tesi.

Nel caso II, da $G_2, H_2 \triangleleft G$ segue $G_2H_2 \triangleleft G$ (esercizio 5, p. 57 dell'originale). Poiché il prodotto contiene entrambi i sottogruppi normali massimali distinti, dev'essere $G_2H_2 = G$. Per il secondo teorema di isomorfismo per i gruppi (p. 65 dell'originale), $G/G_2 = G_2H_2/G_2 \cong H_2/(G_2 \cap H_2)$ e, analogamente, $G/H_2 \cong G_2/(G_2 \cap H_2)$. Quindi $K_3 = G_2 \cap H_2$ è normale massimale sia in H_2 sia in G_2 e valgono gli isomorfismi

$$G/G_2 \cong H_2/K_3, \qquad G/H_2 \cong G_2/K_3. \tag{30}$$

Sia $K_3 \triangleright K_4 \triangleright \cdots \triangleright K_{u+1} = 1$ una serie di composizione di K_3. Per la massimalità appena provata si hanno le quattro serie di composizione

$$\begin{aligned}
&\text{(i)} && G = G_1 \triangleright G_2 \triangleright G_3 \triangleright \cdots \triangleright G_{s+1} = 1,\\
&\text{(ii)} && G = G_1 \triangleright G_2 \triangleright K_3 \triangleright \cdots \triangleright K_{u+1} = 1,\\
&\text{(iii)} && G = H_1 \triangleright H_2 \triangleright K_3 \triangleright \cdots \triangleright K_{u+1} = 1,\\
&\text{(iv)} && G = H_1 \triangleright H_2 \triangleright H_3 \triangleright \cdots \triangleright H_{t+1} = 1.
\end{aligned}$$

Confrontando (i) e (ii) per il caso I, si ha $s = u$ e, ponendo $K_1 = G_1$, $K_2 = G_2$, esiste una permutazione $i \mapsto i'$ di $1, \ldots, s$ per cui $G_i/G_{i+1} \cong K_{i'}/K_{i'+1}$. Lo stesso vale per le ultime due

serie. Quanto a (ii) e (iii), i primi due fattori sono rispettivamente $G/G_2, G_2/K_3$ e $G/H_2, H_2/K_3$: per (30) il primo dell'una è isomorfo al secondo dell'altra e viceversa. I fattori successivi K_i/K_{i+1} coincidono. Componendo le corrispondenze, il teorema risulta provato anche nel caso II. □

Concludiamo la discussione con un criterio di risolubilità per i gruppi finiti.

Teorema 4.12. *Un gruppo finito è risolubile se e solo se ogni fattore di composizione G_i/G_{i+1} di una sua serie di composizione è ciclico di ordine primo.*

Dimostrazione. Se G è risolubile, ogni fattore G_i/G_{i+1} è risolubile. Essendo semplice è abeliano e dunque ciclico di ordine primo. Viceversa, se G ha una serie di composizione con fattori ciclici di ordine primo, tutti questi fattori sono abeliani e G è risolubile per definizione. □

Esercizi.

1. Mostrare che un gruppo abeliano possiede una serie di composizione se e solo se è finito.

2. Sia G ciclico di ordine finito n e sia $G = G_1 \triangleright \cdots \triangleright G_{s+1} = 1$ una serie di composizione. Posto $n_i = |G_i|$, dimostrare che $p_i = n_i/n_{i+1}$ è primo. Viceversa, data una successione di interi $n = n_1, n_2, \ldots, n_{s+1} = 1$ tale che n_i/n_{i+1} sia primo, mostrare che esiste una serie di composizione con $|G_i| = n_i$. Servirsene per dedurre il teorema fondamentale dell'aritmetica per $\mathbb{Z}$.

3. Per g, h in un gruppo scriviamo $g^h = h^{-1}gh$. Allora $g^{hk} = (g^h)^k$ e per definizione di commutatore $g^h = g(g, h)$. Verificare le identità

$$(\alpha) \quad (g, hk) = (g, k)(g, h)^k,$$
$$(\beta) \quad (gh, k) = (g, k)^h(h, k),$$
$$(\gamma) \quad (g^h, (h, k))(h^k, (k, g))(k^g, (g, h)) = 1.$$

4. Se $H, K \triangleleft G$, sia (H, K) il sottogruppo generato dai commutatori (h, k) con $h \in H$, $k \in K$. Mostrare che $(H, K) = (K, H) \triangleleft G$.

5. Mostrare che, per $H, K, L \triangleleft G$, vale $(H, (K, L)) \subseteq (K, (L, H))(L, (H, K))$.

6. Definiamo $G^1 = G$, $G^i = (G^{i-1}, G)$. La successione di sottogruppi normali $G^1 \supseteq G^2 \supseteq G^3 \supseteq \cdots$ si chiama *serie centrale discendente* di G. Il gruppo si dice *nilpotente* se $G^k = 1$ per qualche intero k. Mostrare che un gruppo nilpotente è risolubile. Dare un esempio in cui non vale il viceversa.

7. Mostrare che ogni elemento h di un gruppo nilpotente ha la *proprietà di Engel*: esiste un intero k tale che per ogni $g \in G$ il commutatore iterato $(\cdots((g, h), h), \ldots, h)$ con k occorrenze di h sia uguale a 1.

8. Dimostrare che se H è un sottogruppo proprio di un gruppo nilpotente G, il normalizzatore $N(H)$ contiene propriamente H (per la definizione si veda p. 81 dell'originale).

9. Mostrare che in un gruppo nilpotente finito ogni sottogruppo di Sylow è normale. Mostrare che per ogni primo p che divide $|G|$ esiste un unico p-sottogruppo di Sylow.

10. Per un gruppo G si definisca la *serie centrale ascendente* $1 \subseteq C_1 \subseteq C_2 \subseteq \cdots$ ponendo $C_1 = C(G)$, centro di G, e scegliendo C_i normale tale che C_i/C_{i-1} sia il centro di G/C_{i-1}. Mostrare che un gruppo finito G è nilpotente se e solo se tale serie raggiunge G in un numero finito di passi ($G = C_k$ per qualche k). Osservare che questo risultato e la dimostrazione del teorema 4.8 implicano che ogni p-gruppo è nilpotente.

11. Dimostrare che un gruppo finito è nilpotente se e solo se è prodotto diretto di p-gruppi.

4.7 Il criterio di Galois per la risolubilità per radicali

Siano F un campo e $f(x)$ un polinomio a coefficienti in F. È essenziale precisare che cosa significhi dire che l'equazione $f(x) = 0$ è risolubile per radicali su F.

Definizione 4.2. *Sia $f(x) \in F[x]$ monico di grado positivo. L'equazione $f(x) = 0$ si dice risolubile per radicali su F se esiste un'estensione K/F che possieda una torre di sottocampi*

$$F = F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{r+1} = K, \tag{31}$$

dove $F_{i+1} = F_i(d_i)$ e $d_i^{n_i} = a_i \in F_i$, e tale che K contenga un campo di spezzamento di $f(x)$ su F. Una torre come (31) si chiama torre di radicali di K su F.

Ogni F_{i+1} si ottiene aggiungendo una radice dell'equazione $x^{n_i} = a_i$, e K contiene tutte le radici di $f(x)$. Ciò significa che si può ottenere ogni radice di $f(x)$ partendo da elementi del campo base, eseguendo una successione finita di operazioni razionali e risolvendo equazioni della forma $x^n = a$.

Supponiamo ora che $f(x)$ abbia radici distinte in un campo di spezzamento E/F. Chiamiamo *gruppo di Galois del polinomio* $f(x)$, o *dell'equazione* $f(x) = 0$, il gruppo di Galois di E/F. Poiché due campi di spezzamento sono isomorfi su F, esso è essenzialmente indipendente dalla scelta di E. Sia $f(x) = \prod_{i=1}^{n}(x - r_i)$ in $E[x]$, sicché $E = F(r_1, \ldots, r_n)$ e $R = \{r_1, \ldots, r_n\}$ è l'insieme delle radici distinte di $f(x)$ in E. Si può identificare $G = \mathrm{Gal}(E/F)$ con un gruppo di permutazioni dell'insieme R.[7]

Infatti ogni $\eta \in G$ manda R in sé e induce quindi una permutazione di R. La restrizione $\eta \mapsto \eta|_R$ è un omomorfismo di G nel gruppo simmetrico S_n delle permutazioni di R. Poiché gli r_i generano E/F, l'omomorfismo è iniettivo e la sua immagine, che indicheremo con G_f, è un sottogruppo di S_n isomorfo a G. Spesso identificheremo G e G_f. Per esempio, se $G_f = S_n$ diremo che il gruppo di Galois dell'equazione $f(x) = 0$ è S_n.

Siamo ora in grado di raggiungere il risultato che corona la teoria di Galois, il seguente criterio per la risolubilità per radicali.

Un'equazione $f(x) = 0$ è risolubile per radicali su un campo F di caratteristica zero se e solo se il suo gruppo di Galois è risolubile.

Oltre alla corrispondenza fondamentale tra gruppi e campi ci occorrono alcuni risultati specifici sui campi delle radici dell'unità e sui campi ciclici, ossia campi di spezzamento con gruppo di Galois ciclico. Chiameremo *campo ciclotomico di ordine n su F* un campo di spezzamento di $x^n - 1$ su F.

[7] Era questo il punto di vista di Galois: egli definì il suo gruppo come un particolare gruppo di permutazioni delle radici di $f(x) = 0$. A Dedekind si deve l'osservazione che questo gruppo si può identificare con quello degli automorfismi del campo di spezzamento, osservazione che, come abbiamo visto, sta alla base della teoria di Galois moderna.

Lemma 1. *Il gruppo di Galois del campo ciclotomico di ordine n su un campo F di caratteristica zero è abeliano.*

Dimostrazione. Poiché $x^n - 1$ e la sua derivata nx^{n-1} sono primi tra loro, $x^n - 1$ ha n radici distinte $z_1, \dots, z_n$. Queste costituiscono un sottogruppo U del gruppo moltiplicativo del campo ciclotomico; sappiamo che U è ciclico. La restrizione $\eta \mapsto \eta|_U$ è un monomorfismo del gruppo di Galois G nel gruppo $\operatorname{Aut} U$ degli automorfismi del gruppo ciclico di ordine n. Dunque G è isomorfo a un sottogruppo di $\operatorname{Aut} U$, il quale è isomorfo al gruppo abeliano delle unità dell'anello $\mathbb{Z}/(n)$. Perciò G è abeliano. □

Osservazione. Naturalmente G può essere più piccolo del gruppo delle unità di $\mathbb{Z}/(n)$. Per esempio, F stesso può contenere le radici n-esime dell'unità; allora coincide con il campo ciclotomico e $G = 1$.

Un'estensione E/F si dice *di Galois* o *galoisiana* se soddisfa le condizioni equivalenti del teorema 4.7, per esempio se è finita, normale e separabile. Per una tale estensione si ha la corrispondenza fondamentale tra i sottogruppi di $\operatorname{Gal}(E/F)$ e i sottocampi di E/F. L'estensione si dice *abeliana* (rispettivamente *ciclica*) se è di Galois e il suo gruppo è abeliano (rispettivamente ciclico). Il lemma 1 afferma che ogni estensione ciclotomica di caratteristica zero è abeliana. Stabiliamo due risultati sulle estensioni cicliche, supponendo che il campo base contenga certe radici dell'unità.

Lemma 2. *Se F contiene n radici n-esime distinte dell'unità, il gruppo di Galois di $x^n - a$ su F è ciclico di ordine divisore di n.*

Dimostrazione. Sia U l'insieme di queste radici e sia E il campo di spezzamento di $x^n - a$ su F. Per $a = 0$ il gruppo è banale. Altrimenti, se r è una radice di $x^n = a$ in E, le n radici sono zr, $z \in U$, e $E = F(r)$. Dati $\eta, \zeta \in G = \operatorname{Gal}(E/F)$ si ha $\eta(r) = zr$, $\zeta(r) = z'r$ con $z, z' \in U$, quindi $\zeta\eta(r) = zz'r$. L'applicazione $\eta \mapsto z$ è un monomorfismo di G nel gruppo ciclico U: G è isomorfo a un suo sottogruppo. □

Vale il seguente parziale viceversa.

Lemma 3. *Sia p primo e supponiamo che F contenga p radici p-esime distinte dell'unità. Se E/F è ciclica di grado p, si ha $E = F(d)$ con $d^p \in F$.*

Dimostrazione. Sia $c \in E \setminus F$. Allora $E = F(c)$. Indichiamo con $U = \{z_1, \dots, z_p\}$ le radici p-esime dell'unità e con η un generatore del gruppo di Galois di E/F. Poniamo $c_i = \eta^{i-1}(c)$ per $1 \le i \le p$. Si ha $c_1 = c$, $\eta(c_i) = c_{i+1}$ per $i < p$ e $\eta(c_p) = c_1$. Introduciamo la *risolvente di Lagrange*

$$(z_i, c) = c_1 + c_2 z_i + \cdots + c_p z_i^{p-1}. \tag{32}$$

Allora $\eta((z_i, c)) = c_2 + c_3 z_i + \cdots + c_1 z_i^{p-1} = z_i^{-1}(z_i, c)$. Elevando alla p-esima potenza otteniamo $\eta((z_i, c)^p) = (z_i, c)^p$, quindi $(z_i, c)^p \in F$. Possiamo esprimere $c_1, \dots, c_p$ come combinazioni lineari delle (z_i, c): leggendo (32), per $1 \le i \le p$, come sistema nelle c_j, il determinante dei coefficienti è il determinante di Vandermonde, di valore $\prod_{i<j}(z_j - z_i) \neq 0$. Poiché $E = F(c)$, si ha pertanto $E = F(d_1, \dots, d_p)$ con $d_i = (z_i, c)$. Almeno un d_i è esterno a F e, ponendo $d = d_i$, si ottiene $E = F(d)$ con $d^p \in F$. □

Ci occorre anche sapere che cosa accada al gruppo di Galois di un'equazione quando si estende il campo base. Nella letteratura meno recente tale operazione prende il nome di «aggiunzione di irrazionalità accessorie».

Lemma 4. *Siano $f(x) \in F[x]$ e K un'estensione di F. Il gruppo di Galois di $f(x)$ su K è isomorfo a un sottogruppo del gruppo di Galois di $f(x)$ su F.*

Dimostrazione. Sia L un campo di spezzamento di $f(x)$ su K. Poiché $K \supseteq F$, L contiene un campo di spezzamento E di $f(x)$ su F: se $f(x) = \prod_i (x - r_i)$ in $L[x]$, allora $L = K(r_1, \dots, r_n)$ ed $E = F(r_1, \dots, r_n)$. Un $\eta \in \mathrm{Gal}(L/K)$ manda $R = \{r_1, \dots, r_n\}$ in sé e quindi anche E in sé. Poiché η è determinato dalla sua azione su R, la restrizione $\eta \mapsto \eta|_E$ è un monomorfismo di $\mathrm{Gal}(L/K)$ in $\mathrm{Gal}(E/F)$, come volevamo. □

Sia ora E/F un'estensione di dimensione finita. Essa è generata da un insieme finito di elementi algebrici $a_1, \dots, a_n$. Sia $f_i(x)$ il polinomio minimo di a_i su F e poniamo $f(x) = \prod_i f_i(x)$. Costruiamo un campo di spezzamento K di $f(x)$ su E. Poiché $K \supseteq F(a_1, \dots, a_n)$ e $f_i(a_i) = 0$, K è campo di spezzamento di $f(x)$ anche su F. Il campo di spezzamento di un polinomio è sempre normale, come indicheremo in un esercizio. Qui ci basta il caso separabile, che comprende quello di caratteristica zero: la normalità e la separabilità di K/F seguono allora dal teorema 4.7. Ogni estensione di F normale su F che contenga E contiene un campo di spezzamento di $f(x)$ e quindi un sottocampo isomorfo a K. Ne segue che, a meno di isomorfismo, K è determinato da E/F e non dipende dai generatori scelti. Chiameremo K/F la *chiusura normale* di E/F.

Sempre supponendo $f(x)$ separabile, sia $G = \mathrm{Gal}(K/F)$. Per ogni $\eta \in G$, il sottocampo $\eta(E)$ è isomorfo a E su F. I campi $\eta(E)$ si chiamano *coniugati* di E/F in K e generano K. Infatti, se K' è il sottocampo da essi generato, G lo manda in sé e induce in K' un gruppo finito di automorfismi G' di campo fisso F. Dal teorema 4.7 segue che K' è normale su F e quindi $K' = K$.

Nella definizione di risolubilità per radicali, nel caso separabile possiamo supporre che il campo K sia normale e separabile su F. È una conseguenza del seguente lemma.

Lemma 5. *Sia $F = F_1 \subseteq F_2 \subseteq \cdots \subseteq F_{r+1} = E$ una torre di radicali di E/F, con $F_{i+1} = F_i(d_i)$, $d_i^{n_i} \in F_i$, e supponiamo che E sia generato su F da un insieme finito di elementi di polinomio minimo separabile. Allora la chiusura normale K/F di E/F ammette una torre di radicali in cui gli interi distinti che compaiono come esponenti sono gli stessi n_i della torre data.*

Dimostrazione. La chiusura normale K/F è generata dai coniugati $\eta(E)$, $\eta \in \mathrm{Gal}(K/F)$. Applicando η alla torre data si ottiene una torre di radicali per $\eta(E)$ su F, con gli stessi esponenti. Se $\mathrm{Gal}(K/F) = \{\eta_1, \eta_2, \dots\}$, si ha quindi

$$K = F(\eta_1(d_1), \dots, \eta_1(d_r), \eta_2(d_1), \dots, \eta_2(d_r), \dots).$$

Aggiungendo i generatori nell'ordine indicato si ottiene la torre richiesta. □

Dimostriamo ora il criterio di Galois. Supponiamo dapprima che $f(x) = 0$ sia risolubile per radicali su un campo F di caratteristica zero. Esiste allora un campo K contenente un campo di spezzamento E di $f(x)$ e dotato di una torre di radicali su F come in (31). Per il lemma 5 possiamo supporre K normale su F. Essendo automaticamente separabile, è di Galois su F. Sia n il minimo comune multiplo degli esponenti n_i della torre e sia z una radice primitiva n-esima dell'unità. Prolunghiamo la torre da K a $K(z)$. Se K è il campo di spezzamento su F di $g(x)$, $K(z)$ lo è di $g(x)(x^n - 1)$, quindi anch'esso è di Galois su F. Poiché $z^n = 1 \in F$, possiamo riordinare la torre ponendo $F(z)$ al secondo posto:

$$F = F_1 \subseteq F_2 = F_1(z) \subseteq F_3 = F_2(d_1) \subseteq \cdots \subseteq K(z). \tag{33}$$

Siano $G = \mathrm{Gal}(E/F)$ e $H = \mathrm{Gal}(K(z)/F)$. Nella torre (33) ogni F_{i+1} è abeliano su F_i: per $i = 1$ si applica il lemma 1, per $i > 1$ il lemma 2, dato che F_i contiene le radici dell'unità necessarie. Sia H_i il sottogruppo di H corrispondente a F_i, ossia $H_i = \mathrm{Gal}(K(z)/F_i)$. Poiché F_{i+1} è normale

su F_i, si ha $H_{i+1} \triangleleft H_i$. Inoltre H_i/H_{i+1} è isomorfo al gruppo di Galois di F_{i+1}/F_i, dunque è abeliano. Abbiamo così una serie normale a fattori abeliani e H è risolubile. Poiché il campo di spezzamento E/F è contenuto in $K(z)/F$, il suo gruppo di Galois G è isomorfo a un gruppo quoziente di H. Dunque anche G è risolubile.

Viceversa, sia risolubile il gruppo di Galois G di $f(x) = 0$ su F. Sia $n = |G| = [E : F]$, dove E è un campo di spezzamento di $f(x)$ su F. Poniamo $F_1 = F$, $F_2 = F(z)$, $K = E(z)$, con z radice primitiva n-esima dell'unità. Per il lemma 4 il gruppo di Galois di K/F_2 è isomorfo a un sottogruppo H di G. Perciò H è risolubile e ha una serie di composizione $H = H_1 \triangleright H_2 \triangleright \cdots \triangleright H_{r+1} = 1$ in cui H_i/H_{i+1} è ciclico di ordine primo p_i, $1 \leq i \leq r$. Le corrisponde una catena $F_2 \subseteq F_3 \subseteq \cdots \subseteq F_{r+2} = K$ con $H_i = \mathrm{Gal}(K/F_{i+1})$. Quindi F_{i+2} è normale su F_{i+1}, con gruppo ciclico di ordine p_i. Poiché p_i divide $n = |G|$ e F_{i+1} contiene una radice primitiva n-esima dell'unità, esso contiene le p_i radici p_i-esime dell'unità. Per il lemma 3, $F_{i+2} = F_{i+1}(d_i)$ con $d_i^{p_i} \in F_{i+1}$. Dunque K ha una torre di radicali su F e, contenendo il campo di spezzamento, mostra che $f(x) = 0$ è risolubile per radicali su F. □

Esercizi.

1. Sia p un primo diverso dalla caratteristica di F. Mostrare che, per $a \in F$, il polinomio $x^p - a$ è irriducibile in $F[x]$ oppure ha una radice in F.

2. Supponiamo che $x^p - a$, $a \in \mathbb{Q}$, sia irriducibile in $\mathbb{Q}[x]$ (p primo). Mostrare che il suo gruppo di Galois su $\mathbb{Q}$ è isomorfo al gruppo delle trasformazioni di $\mathbb{Z}/(p)$ della forma $y \mapsto ky + l$ con $k, l \in \mathbb{Z}/(p)$, $k \neq 0$.

3. Sia E/F il campo ciclotomico delle radici p-esime dell'unità su un campo F di caratteristica zero. Mostrare che E si può immergere in un campo K dotato di una torre di radicali su F nella quale gli esponenti n_i siano primi e $[F_{i+1} : F_i] = n_i$. Chiameremo *normalizzata* una tale torre.

4. Determinare torri di radicali normalizzate su $\mathbb{Q}$ per i campi ciclotomici delle radici quinte e settime dell'unità.

5. Dimostrare che, se $f(x) = 0$ ha gruppo di Galois risolubile su un campo F di caratteristica zero, il suo campo di spezzamento si può immergere in un'estensione dotata di una torre di radicali normalizzata su F.

6. Sia E un campo di spezzamento su F di $f(x) \in F[x]$. Mostrare che E è normale su F. (*Suggerimento.* Sia $g(x)$ irriducibile con una radice s in E e si formi un suo campo di spezzamento $K = E(s_1 = s, \ldots, s_m)$ su E, con $g(x) = \prod_j (x - s_j)$ in $K[x]$. Per $2 \leq i \leq m$, gli elementi s_1, s_i hanno lo stesso polinomio minimo $g(x)$ su F, perciò esiste un isomorfismo su F da $F(s_1)$ a $F(s_i)$ che manda s_1 in s_i. Esso si prolunga a un isomorfismo $\eta : E(s_1) \to E(s_i)$, poiché questi sono campi di spezzamento di $f(x)$ rispettivamente su $F(s_1)$ e $F(s_i)$. Allora $\eta(E(s_1)) = E(s_i)$ e $E(s_1) = E$ perché $s_1 \in E$. Ne segue $E(s_i) = E$, che contiene tutti gli s_i. Quindi la fattorizzazione $g(x) = \prod_j (x - s_j)$ ha luogo in $E[x]$.)

4.8 Il gruppo di Galois come gruppo di permutazioni delle radici

Sfrutteremo l'idea introdotta all'inizio del §4.7: identificare il gruppo di Galois di un'equazione con un gruppo di permutazioni delle sue radici. Consideriamo, come prima, un polinomio monico $f(x) \in F[x]$ di grado positivo, con radici distinte $r_1, \ldots, r_n$ in un campo di spezzamento $E = F(r_1, \ldots, r_n)$. Il gruppo G_f, isomorfo a $G = \mathrm{Gal}(E/F)$, è il sottogruppo del gruppo

simmetrico S_n su $R = \{r_1, \ldots, r_n\}$ indotto da G. Identificando G con G_f, la corrispondenza di Galois associa ai sottogruppi di G_f i sottocampi di E/F. Qual è il sottocampo corrispondente a $G_f \cap A_n$, dove A_n è il sottogruppo alterno di S_n? Per semplicità supponiamo $\operatorname{char} F \neq 2$, lasciando a un esercizio il caso di caratteristica due.

Teorema 4.13. *Sia F un campo di caratteristica diversa da due e sia $f(x) \in F[x]$ monico di grado positivo con radici distinte r_i in un campo di spezzamento E/F. Posto*

$$D = \prod_{1 \leq i < j \leq n} (r_i - r_j), \tag{34}$$

il sottocampo di E/F corrispondente a $G_f \cap A_n$ nella corrispondenza di Galois è $F(D)$.

Dimostrazione. Consideriamo anzitutto l'anello $F[x_1, \ldots, x_n]$ nelle indeterminate x_i. Ricordiamo che a ogni permutazione π di $1, \ldots, n$ corrisponde un unico automorfismo $\zeta(\pi)$ dell'anello che fissa F e manda x_i in $x_{\pi(i)}$ per $1 \leq i \leq n$ (teorema 2.12, p. 125 dell'originale). Per $\pi_1, \pi_2 \in S_n$ si ha $\zeta(\pi_1\pi_2) = \zeta(\pi_1)\zeta(\pi_2)$. Poniamo

$$\Delta = \prod_{1 \leq i < j \leq n} (x_i - x_j) \tag{35}$$

e consideriamo l'effetto su Δ dell'automorfismo corrispondente alla trasposizione (kl), $k < l$. Affermiamo che $\zeta(kl)(\Delta) = -\Delta$. Anzitutto $\zeta(kl)(x_k - x_l) = x_l - x_k = -(x_k - x_l)$. Se $k < l < i$, si scambiano $x_k - x_i$ e $x_l - x_i$, entrambi presenti in Δ. Se $i < k < l$, si scambiano $x_i - x_k$ e $x_i - x_l$. Se invece $k < i < l$, il fattore $x_k - x_i$ diventa $x_l - x_i = -(x_i - x_l)$ e il fattore $x_i - x_l$ diventa $x_i - x_k = -(x_k - x_i)$. Infine tutti i fattori $x_i - x_j$ con $i, j \notin \{k, l\}$ rimangono fissi. Ne segue $\zeta(kl)(\Delta) = -\Delta$. Dalla moltiplicatività di ζ segue quindi che $\zeta(\pi)(\Delta)$ è Δ o $-\Delta$ secondo che π sia pari o dispari.

Sia $\eta \in \operatorname{Gal}(E/F)$ e sia π la corrispondente permutazione delle radici. Applicando a $\zeta(\pi)(\Delta)$ l'omomorfismo $F[x_1, \ldots, x_n] \to E$ che è l'identità su F e manda x_i in r_i, otteniamo $\eta(D)$, con D come in (34). Quindi $\eta(D) = D$ se π è pari e $\eta(D) = -D$ se π è dispari. Poiché le radici sono distinte e la caratteristica è diversa da due, gli elementi di $\operatorname{Gal}(E/F)$ che fissano $F(D)$ sono esattamente quelli che inducono permutazioni pari. Identificando G con G_f, il sottogruppo corrispondente a $F(D)$ è $G_f \cap A_n$; il campo corrispondente a tale sottogruppo è dunque $F(D)$. □

La dimostrazione mostra che $\eta(D) = \pm D$ per ogni $\eta \in G$. Posto $d = D^2$, si ha $\eta(d) = d$ per ogni η e quindi $d \in F$. Inoltre $F(D) = F$ se e solo se $G_f \subseteq A_n$. Le due radici quadrate di d in E sono $D, -D$, quindi $G_f \subseteq A_n$ se e solo se d è un quadrato in F.

Corollario. *Nelle ipotesi del teorema 4.13, il gruppo di Galois di $f(x)$ su F è un sottogruppo del gruppo alterno se e solo se*

$$d = \prod_{1 \leq i < j \leq n} (r_i - r_j)^2 \tag{36}$$

è il quadrato di un elemento di F.

L'elemento $d \in F$ si chiama *discriminante* di $f(x)$. Per calcolarlo scriviamo

$$f(x) = x^n - a_1 x^{n-1} + \cdots + (-1)^n a_n = \prod_{i=1}^{n} (x - r_i). \tag{37}$$

Allora

$$a_1 = \sum_i r_i, \quad a_2 = \sum_{i<j} r_i r_j, \quad \ldots, \quad a_n = r_1 r_2 \cdots r_n. \tag{38}$$

Vale inoltre la nota formula del determinante di Vandermonde

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ r_1 & r_2 & \cdots & r_n \\ \vdots & \vdots & & \vdots \\ r_1^{n-1} & r_2^{n-1} & \cdots & r_n^{n-1} \end{vmatrix} = \prod_{i>j} (r_i - r_j). \tag{39}$$

Moltiplicando a destra la matrice per la sua trasposta e prendendo il determinante otteniamo

$$\begin{vmatrix} n & s_1 & s_2 & \cdots & s_{n-1} \\ s_1 & s_2 & s_3 & \cdots & s_n \\ s_2 & s_3 & s_4 & \cdots & s_{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ s_{n-1} & s_n & s_{n+1} & \cdots & s_{2n-2} \end{vmatrix} = \prod_{i<j} (r_i - r_j)^2 = d, \tag{40}$$

dove $s_i = r_1^i + \cdots + r_n^i$. Queste somme di potenze si possono esprimere come polinomi negli a_i a coefficienti interi (teorema 2.20, p. 139 dell'originale). La (40) fornisce dunque per il discriminante $d = d(f)$ una formula polinomiale a coefficienti interi nei coefficienti a_i. Dalla (36) risulta che f ha radici multiple se e solo se $d = 0$.

Calcoliamo d per $n = 2, 3$. Se $n = 2$, $f(x) = x^2 - a_1 x + a_2 = (x - r_1)(x - r_2)$ con $a_1 = r_1 + r_2$, $a_2 = r_1 r_2$. Quindi $s_2 = r_1^2 + r_2^2 = (r_1 + r_2)^2 - 2r_1 r_2 = a_1^2 - 2a_2$. Dalla (40) viene la consueta formula per il discriminante del polinomio quadratico:

$$d = 2s_2 - s_1^2 = a_1^2 - 4a_2. \tag{41}$$

Se $n = 3$, $f(x) = x^3 - a_1 x^2 + a_2 x - a_3 = (x - r_1)(x - r_2)(x - r_3)$, quindi $s_1 = r_1 + r_2 + r_3 = a_1$, $r_1 r_2 + r_1 r_3 + r_2 r_3 = a_2$ e $r_1 r_2 r_3 = a_3$. Segue $s_2 = (r_1 + r_2 + r_3)^2 - 2(r_1 r_2 + r_1 r_3 + r_2 r_3) = a_1^2 - 2a_2$. Per s_3, s_4 usiamo le relazioni $r_k^3 = a_1 r_k^2 - a_2 r_k + a_3$, $r_k^4 = a_1 r_k^3 - a_2 r_k^2 + a_3 r_k$. Allora

$$\begin{aligned} s_3 &= r_1^3 + r_2^3 + r_3^3 = a_1(r_1^2 + r_2^2 + r_3^2) - a_2(r_1 + r_2 + r_3) + 3a_3 \\ &= a_1(a_1^2 - 2a_2) - a_2 a_1 + 3a_3 = a_1^3 - 3a_1 a_2 + 3a_3, \\ s_4 &= a_1 s_3 - a_2 s_2 + a_3 s_1 \\ &= a_1(a_1^3 - 3a_1 a_2 + 3a_3) - a_2(a_1^2 - 2a_2) + a_3 a_1 \\ &= a_1^4 - 4a_1^2 a_2 + 4a_1 a_3 + 2a_2^2. \end{aligned}$$

Sostituendo nella (40) otteniamo

$$\begin{aligned} d &= 3s_2 s_4 + 2s_1 s_2 s_3 - s_2^3 - 3s_3^2 - s_1^2 s_4 \\ &= -4a_1^3 a_3 + a_1^2 a_2^2 + 18a_1 a_2 a_3 - 4a_2^3 - 27a_3^2. \end{aligned} \tag{42}$$

Ricaviamo ora dal gruppo di Galois, visto come gruppo di permutazioni delle radici, un criterio perché $f(x)$ sia irriducibile in $F[x]$.

Teorema 4.14. *Sia $f(x) \in F[x]$ privo di radici multiple. Allora $f(x)$ è irriducibile in $F[x]$ se e solo se G_f è un gruppo transitivo di permutazioni delle radici r_i.*

Dimostrazione. Ricordiamo che un gruppo G di trasformazioni di un insieme M è transitivo se a ogni coppia (x, y) di elementi di M si può associare un $\eta \in G$ tale che $\eta(x) = y$. Supponiamo $f(x)$ irriducibile e siano r_i, r_j due sue radici. Poiché $f(r_i) = f(r_j) = 0$, esiste un isomorfismo di $F(r_i)/F$ su $F(r_j)/F$ che manda r_i in r_j (p. 227 dell'originale). Il campo $E = F(r_1, \dots, r_n)$ è campo di spezzamento di $f(x) = \prod_k (x - r_k)$ sia su $F(r_i)$ sia su $F(r_j)$, quindi tale isomorfismo si prolunga a un automorfismo $\eta \in \mathrm{Gal}(E/F)$. Si ha $\eta(r_i) = r_j$: l'azione sulle radici è transitiva. Viceversa, se G_f è transitivo, sia f_1 un fattore irriducibile monico di grado positivo di f, e r_i una sua radice. Per ogni altra radice r_j di f esiste $\eta \in G_f$ con $\eta(r_i) = r_j$. Da $f_1(r_i) = 0$ segue $0 = \eta(f_1(r_i)) = f_1(r_j)$. Quindi ogni radice di f è una radice di f_1. Essendo distinte, $f = f_1$ è irriducibile. □

Questi risultati rendono agevole il calcolo dei gruppi di Galois delle equazioni di secondo e terzo grado. Idee analoghe valgono per quelle di quarto grado, che tratteremo negli esercizi.[8] Supponiamo $\mathrm{char}\, F \neq 2$ e che f abbia radici distinte. Per $f(x) = x^2 - a_1 x + a_2$ il gruppo è S_2 oppure $A_2 = 1$ secondo che $d = a_1^2 - 4a_2$ non sia o sia un quadrato in F. Passiamo a $f(x) = x^3 - a_1 x^2 + a_2 x - a_3$. Se $f(x) = (x - r)g(x)$ in $F[x]$, il gruppo coincide con quello del polinomio quadratico $g(x)$. Possiamo dunque supporre f irriducibile. I soli sottogruppi transitivi di S_3 sono S_3, A_3, quindi G_f è uno di questi. Per il corollario al teorema 4.13, $G_f = A_3$ se $d = -4a_1^3 a_3 + a_1^2 a_2^2 + 18 a_1 a_2 a_3 - 4a_2^3 - 27 a_3^2$ è un quadrato in F. Altrimenti $G_f = S_3$.

Esercizi.

1. Supponiamo $d(f) \neq 0$. Se $f(x) = f_1(x) \cdots f_r(x)$, con f_i irriducibile di grado n_i in $F[x]$, mostrare che l'insieme R delle radici di f si decompone in orbite di G_f di cardinalità n_i, $1 \leq i \leq r$. Se il gruppo di Galois è ciclico, generato da η, dedurne che si può ordinare R in modo che la permutazione indotta da η abbia decomposizione $(1\,2 \dots n_1)(n_1 + 1 \dots n_1 + n_2)(n_1 + n_2 + 1 \dots n_1 + n_2 + n_3) \cdots$.

2. Sia $F = \mathbb{R}$ e sia $f(x)$ cubico di discriminante d. Mostrare che f ha radici multiple, tre radici reali distinte oppure una radice reale e due non reali secondo che $d = 0$, $d > 0$ oppure $d < 0$.

3. Sia F di caratteristica arbitraria, anche due, e siano $r_1, \dots, r_n$ le radici distinte di f. Posto $D' = \sum_{\pi \in A_n} r_{\pi(1)}^0 r_{\pi(2)}^1 \cdots r_{\pi(n)}^{n-1}$, mostrare che per ogni permutazione dispari σ vale

$$D' - \sum_{\pi \in A_n} r_{\sigma\pi(1)}^0 r_{\sigma\pi(2)}^1 \cdots r_{\sigma\pi(n)}^{n-1} = \prod_{i>j} (r_i - r_j).$$

   Mostrare che il sottocampo degli invarianti di $G_f \cap A_n$ è $F(D')$. Determinare un'equazione di secondo grado a coefficienti in F avente D' per radice.

   Negli esercizi seguenti sul quarto grado supponiamo $\mathrm{char}\, F \neq 2$, che $f(x) = x^4 - a_1 x^3 + a_2 x^2 - a_3 x + a_4$ abbia radici distinte r_1, r_2, r_3, r_4, che $E = F(r_1, r_2, r_3, r_4)$, $G = \mathrm{Gal}(E/F)$ e che G_f sia il corrispondente gruppo di permutazioni delle radici.

4. Mostrare che $V = \{1, (12)(34), (13)(24), (14)(23)\}$ è normale in S_4.

[8] Per le equazioni di quinto grado si vedano E. Dehn, *Algebraic Equations*, New York, Columbia University Press, 1930, p. 195, oppure H. Weber, *Lehrbuch der Algebra*, vol. I, 1898, p. 670. Altre indicazioni si trovano negli esercizi alla fine del §4.16.

5. Mostrare che il sottocampo degli invarianti di $G_f \cap V$ è $F(t_1, t_2, t_3)$, dove $t_1 = r_1r_2 + r_3r_4$, $t_2 = r_1r_3 + r_2r_4$, $t_3 = r_1r_4 + r_2r_3$.

6. Sia $g(x) = (x - t_1)(x - t_2)(x - t_3)$, detto *risolvente cubica* della quartica $f(x)$. Verificare che

$$g(x) = x^3 - b_1x^2 + b_2x - b_3, \tag{43}$$

dove

$$b_1 = a_2, \qquad b_2 = a_1a_3 - 4a_4, \qquad b_3 = a_1^2a_4 + a_3^2 - 4a_2a_4, \tag{44}$$

e che f e g hanno lo stesso discriminante.

7. Mostrare che i sottogruppi transitivi di S_4 sono (i) S_4, (ii) A_4, (iii) V, (iv) $C = \{1, (1234), (13)(24), (1432)\}$ e i suoi coniugati, (v) $D = V \cup \{(12), (34), (1423), (1324)\}$, che è un 2-sottogruppo di Sylow di S_4 (di ordine 8), e i suoi coniugati.

8. Mostrare che il gruppo di Galois G_g di $g(x) = 0$ è isomorfo a $G_f/(G_f \cap V)$. Supponendo f irriducibile, verificare che (i) se $G_f = S_4$, G_g ha ordine 6; (ii) se $G_f = A_4$, ha ordine 3; (iii) se $G_f = V$, $G_g = 1$; (iv) se $G_f = C$ o uno dei suoi coniugati, ossia un qualunque sottogruppo ciclico di ordine 4, G_g ha ordine 2; (v) se $G_f = D$ o uno dei suoi coniugati, ossia un qualunque sottogruppo di Sylow di ordine 8 in S_4, G_g ha ordine 2. Si osservi che da G_g si riconosce G_f, salvo l'alternativa tra (iv) e (v).

9. Dimostrare che, se G_g ha ordine 2, allora $G_f \cong D$ oppure $G_f \cong C$ secondo che $f(x)$ sia o non sia irriducibile su $F(\sqrt{d})$, dove d è il discriminante di f.

10. Determinare il gruppo di Galois su $\mathbb{Q}$ di $x^4 + 3x^3 - 3x - 2 = 0$.

I quattro esercizi seguenti hanno lo scopo di mostrare che ogni sottogruppo transitivo risolubile di S_p, p primo, è equivalente a un sottogruppo del gruppo delle trasformazioni $x \mapsto ax + b$, $a \neq 0$, di $\mathbb{Z}/(p)$ che contenga tutte le traslazioni $x \mapsto x + b$.

11. Sia $H \neq 1$ un sottogruppo normale di un sottogruppo transitivo G di S_n su $\{1, \ldots, n\}$. Mostrare che tutte le orbite di H hanno la stessa cardinalità. Dedurne che se $n = p$ è primo allora H è transitivo.

12. Sia L il gruppo di tutte le trasformazioni $x \mapsto ax + b$, $a \neq 0$, di $\mathbb{Z}/(p)$, p primo, comprese tutte le traslazioni $x \mapsto x + b$. Mostrare che le traslazioni con $b \neq 0$ sono le sole trasformazioni di L prive di punti fissi e quindi le sole che si scrivono come un p-ciclo.

13. Sia G un gruppo di trasformazioni di $\mathbb{Z}/(p)$ contenente come sottogruppo normale il gruppo H delle traslazioni. Mostrare che G è un sottogruppo di L. (*Suggerimento.* Sia $\tau : x \mapsto x + 1$, $\eta \in G$. Allora $\eta\tau\eta^{-1}$ è della forma $x \mapsto x + k$ (esercizio 12). Quindi $\eta(x + 1) = \eta\tau(x) = \eta(x) + k$, da cui si può concludere che $\eta(x) = kx + b$.)

14. Usare l'induzione e l'esercizio 13 per dimostrare che ogni sottogruppo risolubile transitivo di S_p, p primo, è equivalente a un sottogruppo di L contenente le traslazioni.

15. (Galois.) Sia $f(x) \in F[x]$ irriducibile di grado primo su un campo F di caratteristica zero e sia E un suo campo di spezzamento. Mostrare che f è risolubile per radicali su F se e solo se $E = F(r_i, r_j)$ per ogni coppia di radici distinte r_i, r_j.

16. Sia E un campo di spezzamento su F di $f(x) \in F[x]$ e $G = \mathrm{Gal}(E/F)$. Per indeterminate $x_1, \dots, x_n$ poniamo $\widetilde{E} = E(x_1, \dots, x_n)$, $\widetilde{F} = F(x_1, \dots, x_n)$. Osservare che $\widetilde{E}$ è campo di spezzamento di f su $\widetilde{F}$ e mostrare che la restrizione $\widetilde{\eta} \mapsto \eta = \widetilde{\eta}|_E$ è un isomorfismo da $\widetilde{G} = \mathrm{Gal}(\widetilde{E}/\widetilde{F})$ a G. Supponiamo che f abbia grado n e n radici distinte $r_1, \dots, r_n$ in E. Per ogni permutazione π di $1, \dots, n$ poniamo

$$u_\pi = \sum_{i=1}^{n} r_{\pi(i)} x_i = \sum_{i=1}^{n} r_i x_{\pi^{-1}(i)}.$$

Osservare che $u_{\pi_1} \neq u_{\pi_2}$ per permutazioni distinte e quindi l'orbita di u_π sotto $\widetilde{G}$ ha $|\widetilde{G}|$ elementi. Concludere che u_π è un elemento primitivo di $\widetilde{E}$ su $\widetilde{F}$, cioè $\widetilde{E} = \widetilde{F}(u_\pi)$, e che il suo polinomio minimo è $\varphi_\pi(x) = \prod_{\sigma \in G_f} (x - \sum_i r_{\sigma\pi(i)} x_i)$. Poniamo $\varphi(x) = \prod_{\pi \in S_n} (x - u_\pi)$. Mostrare che $\varphi(x) \in F[x_1, \dots, x_n, x]$ e che i suoi fattori irriducibili in tale anello hanno la forma $\varphi_\pi(x)$. Dedurne che G è isomorfo al sottogruppo delle permutazioni $\sigma \in S_n$ per le quali l'automorfismo $\eta(\sigma)$ di $F[x_1, \dots, x_n, x]$ che fissa F e x e manda x_i in $x_{\sigma(i)}$ fissa i fattori irriducibili di $\varphi(x)$.

4.9 L'equazione generale di grado n

Per equazione generale intendiamo un'equazione i cui coefficienti siano indeterminate distinte. Più precisamente, dati un campo F e indeterminate distinte $t_1, \dots, t_n$, chiamiamo *equazione generale di grado n su F* l'equazione

$$f(x) = x^n - t_1 x^{n-1} + t_2 x^{n-2} - \cdots + (-1)^n t_n = 0. \tag{45}$$

Essa si dice risolubile per radicali se lo è sul campo $F(t_1, \dots, t_n)$, campo delle frazioni di $F[t_1, \dots, t_n]$. Per esempio la formula $x = \frac{1}{2} t_1 \pm \frac{1}{2}\sqrt{t_1^2 - 4t_2}$ dimostra, in caratteristica diversa da due, la risolubilità dell'equazione generale di secondo grado: le radici appartengono a $F(t_1, t_2, d)$ con $d^2 = t_1^2 - 4t_2 \in F(t_1, t_2)$. Per decidere la questione mediante il criterio di Galois dobbiamo determinare il gruppo G_f di $f(x)$ su $F(t_1, \dots, t_n)$. Sia E il campo di spezzamento e sia $f(x) = (x - y_1) \cdots (x - y_n)$ in $E[x]$. Il confronto con (45) dà $t_1 = \sum_i y_i$, $t_2 = \sum_{i>j} y_i y_j$, $\dots$, $t_n = y_1 \cdots y_n$. Pertanto

$$E = F(t_1, \dots, t_n, y_1, \dots, y_n) = F(y_1, \dots, y_n). \tag{46}$$

Per ottenere G_f applichiamo un risultato del §4.5 sulle espressioni razionali simmetriche. Introduciamo nuove indeterminate $x_1, \dots, x_n$ e formiamo $F(x_1, \dots, x_n)$ e il suo sottocampo delle espressioni razionali simmetriche. Quest'ultimo coincide con $F(p_1, \dots, p_n)$, dove i p_i sono i polinomi simmetrici elementari negli x_i. Inoltre $F(x_1, \dots, x_n)$ è campo di spezzamento, su $F(p_1, \dots, p_n)$, di $g(x) = \prod_{i=1}^{n} (x - x_i)$, con gruppo di Galois $G_g = S_n$.

Trasportiamo questo risultato dalla coppia $F(x_1, \dots, x_n) \supseteq F(p_1, \dots, p_n)$ alla coppia $F(y_1, \dots, y_n) \supseteq F(t_1, \dots, t_n)$, nella quale sono i t_i a essere indeterminate. A tale scopo costruiremo un isomorfismo di $F(y_1, \dots, y_n)$ su $F(x_1, \dots, x_n)$ che mandi $F(t_1, \dots, t_n)$ su $F(p_1, \dots, p_n)$. Poiché i t_i sono indeterminate, esiste un omomorfismo $\sigma : F[t_1, \dots, t_n] \to F[p_1, \dots, p_n]$ che fissa F e manda t_i in p_i. Affermiamo che è iniettivo. Infatti, essendo anche gli x_i indeterminate, esiste un omomorfismo $\tau : F[x_1, \dots, x_n] \to F[y_1, \dots, y_n]$ che fissa F e manda x_i in y_i. Abbiamo il diagramma

$$\begin{array}{ccc} F[x_1, \dots, x_n] & \xrightarrow{\tau} & F[y_1, \dots, y_n] \\ \cup\mathrm{I} & & \cup\mathrm{I} \\ F[p_1, \dots, p_n] & \xleftarrow{\sigma} & F[t_1, \dots, t_n]. \end{array}$$

La composta $\tau\sigma$ è definita e, per le formule che legano i p_i agli x_i e i t_i agli y_i,

$$\tau\sigma(t_i) = \tau(p_i) = \tau\left(\sum_{j_1<\cdots<j_i} x_{j_1}\cdots x_{j_i}\right) = \sum_{j_1<\cdots<j_i} y_{j_1}\cdots y_{j_i} = t_i.$$

Per ogni $h \in F[t_1,\ldots,t_n]$ si ha dunque $\tau\sigma(h) = h$. Ne segue l'iniettività di σ, perché da $\sigma(h) = 0$ segue $h = \tau\sigma(h) = 0$. Inoltre σ è suriettivo e quindi è un isomorfismo, che si prolunga in modo unico a un isomorfismo, ancora denotato con σ, di $F(t_1,\ldots,t_n)$ su $F(p_1,\ldots,p_n)$. Facendo corrispondere x a x ne otteniamo un isomorfismo σ' di $F(t_1,\ldots,t_n)[x]$ su $F(p_1,\ldots,p_n)[x]$ che manda $f(x) = x^n - t_1x^{n-1} + \cdots + (-1)^n t_n$ in $g(x) = x^n - p_1x^{n-1} + \cdots + (-1)^n p_n$. Poiché $F(y_1,\ldots,y_n)$ e $F(x_1,\ldots,x_n)$ sono i rispettivi campi di spezzamento, σ si prolunga a un isomorfismo ρ tra questi campi. Esso manda $F(t_1,\ldots,t_n)$ su $F(p_1,\ldots,p_n)$ e induce l'isomorfismo $\eta \mapsto \rho\eta\rho^{-1}$ tra G_f e G_g. Poiché G_g è il gruppo simmetrico, $G_f = S_n$. È chiaro inoltre che $f(x)$ è irriducibile in $F(t_1,\ldots,t_n)[x]$ e che le sue radici y_i sono distinte. Riuniamo i risultati ottenuti.

Teorema 4.15. *L'equazione generale di grado n, $f(x) = 0$ nella forma (45), è irriducibile in $F(t_1,\ldots,t_n)[x]$ e ha radici distinte. Il suo gruppo di Galois è il gruppo simmetrico S_n.*

Per il corollario al teorema 4.11, S_n non è risolubile per $n > 4$. Ne segue il celebre

Teorema di Ruffini–Abel. *In caratteristica zero l'equazione generale di grado $n > 4$ non è risolubile per radicali.*

Il criterio di Galois implica anche che le equazioni generali di terzo e quarto grado sono risolubili per radicali. La sua dimostrazione suggerisce un procedimento per risolverle. Applichiamolo alle cubiche: ritroveremo così le formule di Cardano. Il risultato corrispondente per le quartiche sarà indicato in un esercizio.

Supponiamo soltanto che la caratteristica di F sia diversa da due e da tre e consideriamo la cubica generale $x^3 - t_1x^2 + t_2x - t_3 = (x-x_1)(x-x_2)(x-x_3)$, con i t_i indeterminate e gli x_i radici nel campo di spezzamento $F(x_1,x_2,x_3)$. La dimostrazione del criterio di Galois mostra l'utilità di disporre di una radice cubica primitiva dell'unità. Le due radici primitive $w, w^2 = w^{-1}$ risolvono $x^2 + x + 1 = 0$, perciò sono, per esempio, $w = -\frac{1}{2} + \frac{1}{2}\sqrt{-3}$ e $w^{-1} = -\frac{1}{2} - \frac{1}{2}\sqrt{-3}$. Supponiamo che appartengano a F. Per semplificare i calcoli sostituiamo gli x_i con $y_i = x_i - \frac{1}{3}(x_1+x_2+x_3) = x_i - \frac{1}{3}t_1$. L'equazione diviene $y^3 + py + q = 0$ e le sue radici soddisfano $y_1+y_2+y_3 = 0$. Il suo gruppo è S_3, con serie di composizione $S_3 \triangleright A_3 \triangleright 1$. Il sottocampo corrispondente ad A_3 è $K(\sqrt{d})$, con $K = F(p,q)$ e discriminante $d = -4p^3 - 27q^2$ per (42). Il campo di spezzamento cercato $K(y_1,y_2,y_3)$ è ciclico di grado tre su $K(\sqrt{d})$. Lo si può quindi ottenere mediante estrazione di una radice cubica, usando le risolventi di Lagrange definite dagli y_i, permutati ciclicamente da A_3, e da una radice cubica dell'unità. Le tre risolventi sono

$$z_1 = y_1 + wy_2 + w^2y_3, \quad z_2 = y_1 + w^2y_2 + wy_3, \quad z_3 = y_1 + y_2 + y_3 = 0$$

con $w = -\frac{1}{2} + \frac{1}{2}\sqrt{-3}$. Allora

$$\begin{aligned} z_1^3 = {} & \sum_i y_i^3 + 3w(y_1^2y_2 + y_2^2y_3 + y_3^2y_1) \\ & + 3w^2(y_1y_2^2 + y_2y_3^2 + y_3y_1^2) + 6y_1y_2y_3. \end{aligned} \tag{47}$$

Per ottenere z_2^3 basta scambiare w e w^2. Ora

$$\begin{aligned} \sqrt{d} &= (y_1-y_2)(y_2-y_3)(y_1-y_3) \\ &= y_1^2y_2 + y_2^2y_3 + y_3^2y_1 - (y_1y_2^2 + y_2y_3^2 + y_3y_1^2). \end{aligned}$$

Ponendo $u = y_1^2y_2 + y_2^2y_3 + y_3^2y_1 + y_1y_2^2 + y_2y_3^2 + y_3y_1^2$ e utilizzando $w + w^2 = -1$, $w - w^2 = \sqrt{-3}$, dalla (47) si ottiene

$$z_1^3 = \sum_i y_i^3 - \tfrac{3}{2}u + \tfrac{3}{2}\sqrt{-3}\sqrt{d} + 6y_1y_2y_3. \tag{48}$$

D'altra parte $0 = (y_1+y_2+y_3)^3 = \sum_i y_i^3 + 3u + 6y_1y_2y_3$ e $0 = (y_1+y_2+y_3)(y_1y_2+y_2y_3+y_1y_3) = u + 3y_1y_2y_3$. Poiché $y_1y_2y_3 = -q$, segue $z_1^3 = -\frac{9}{2}u + \frac{3}{2}\sqrt{-3}\sqrt{d} = -\frac{27}{2}q + \frac{3}{2}\sqrt{-3d}$. Nel passare a z_2^3 scambiamo w con w^2. Poiché $w^2 - w = -\sqrt{-3}$ basta sostituire $\sqrt{-3}$ con $-\sqrt{-3}$ nella formula precedente. Dunque

$$z_1^3 = -\tfrac{27}{2}q + \tfrac{3}{2}\sqrt{-3d}, \qquad z_2^3 = -\tfrac{27}{2}q - \tfrac{3}{2}\sqrt{-3d}, \tag{49}$$

con la medesima determinazione di $\sqrt{-3d}$ in entrambe le formule. Ciascuna estrazione di radice cubica offre tre determinazioni, che si devono abbinare tenendo conto che

$$z_1z_2 = \sum_i y_i^2 - \sum_{i<j} y_iy_j = \left(\sum_i y_i\right)^2 - 3\sum_{i<j} y_iy_j = -3p.$$

Ne segue

$$\begin{aligned} z_1 &= \sqrt[3]{-\tfrac{27}{2}q + \tfrac{3}{2}\sqrt{-3d}}, \qquad d = -4p^3 - 27q^2, \\ z_2 &= \sqrt[3]{-\tfrac{27}{2}q - \tfrac{3}{2}\sqrt{-3d}}, \end{aligned} \tag{50}$$

dove si usa la stessa radice quadrata e si scelgono le radici cubiche in modo che $z_1z_2 = -3p$. Usando le espressioni delle risolventi e la relazione $w + w^2 + 1 = 0$ si ottiene

$$y_1 = \tfrac{1}{3}(z_1 + z_2), \qquad y_2 = \tfrac{1}{3}(w^2z_1 + wz_2), \qquad y_3 = \tfrac{1}{3}(wz_1 + w^2z_2). \tag{51}$$

Le (50) e (51) sono le *formule di Cardano* per risolvere la cubica $x^3 + px + q = 0$ a coefficienti indeterminati. Si applicano anche alle cubiche a coefficienti in un campo qualsiasi di caratteristica diversa da due e da tre.

Esercizi.

1. Risolvere su $\mathbb{Q}$ mediante le formule di Cardano (a) $x^3 - 2x + 4 = 0$ e (b) $x^3 - 15x + 4 = 0$.

2. Supponiamo $\operatorname{char} F \neq 2, 3$ e consideriamo la quartica generale $x^4 - t_1x^3 + t_2x^2 - t_3x + t_4 = \prod_{i=1}^4 (x - x_i)$. Sostituendo x_i con $y_i = x_i - \frac{1}{4}t_1$ si ottiene $f(y) = y^4 + py^2 + qy + r = 0$, con $\sum_i y_i = 0$. Mostrare che la risolvente cubica di $f(y) = 0$ è $g(z) = z^3 - pz^2 - 4rz + (4pr - q^2) = 0$ (esercizio 6, p. 261 dell'originale). Siano z_1, z_2, z_3 le sue radici. Mostrare che il gruppo di Galois del campo di spezzamento $F(x_1, x_2, x_3, x_4) = F(y_1, y_2, y_3, y_4)$ sul campo delle radici della risolvente $F(z_1, z_2, z_3)$, intendendo aggiunti al campo base i coefficienti della quartica, è $V = \{1, (12)(34), (13)(24), (14)(23)\}$. Esprimere gli y_i mediante gli z_j e radici quadrate di elementi di $F(z_1, z_2, z_3)$. Insieme alla soluzione di Cardano per la risolvente, ciò dà una soluzione per radicali della quartica.

3. Applicare il metodo dell'esercizio 2 per risolvere $x^4 - 2x^3 - 8x - 3 = 0$ su $\mathbb{Q}$.

4. Sfruttare il fatto che ogni gruppo finito è isomorfo a un sottogruppo di S_n per dimostrare che, dato un qualsiasi gruppo finito G, esistono campi F ed E/F tali che $\operatorname{Gal}(E/F) \cong G$.

5. Sia $E = \mathbb{C}(t,u)$ un'estensione di $\mathbb{C}$ con t trascendente su $\mathbb{C}$ e $u^2 + t^2 = 1$ su $\mathbb{C}(t)$. Determinare il gruppo di Galois di $\mathbb{C}(t,u)$ su $\mathbb{C}(t^n, u^n)$ per ogni $n \in \mathbb{N}$. Mostrare che $u_n = \frac{1}{2}[(t+iu)^n + (t-iu)^n]$, $i = \sqrt{-1}$, appartiene a $\mathbb{C}(t^n, u^n)$. Dedurne che la funzione $\cos(nx)$ si può esprimere razionalmente, a coefficienti complessi, mediante $\cos^n x$ e $\sin^n x$. Vale lo stesso per $\sin(nx)$?

6. Sia F un sottocampo di $\mathbb{R}$ e sia $f(x) \in F[x]$ cubico irriducibile di discriminante $d > 0$ e con $G_f = A_3$. Mostrare che le radici di f in $\mathbb{C}$ sono reali. Sia p primo e sia $K = F(r)$, dove r è reale e $r^p \in F$. Mostrare che K non può contenere un campo di spezzamento di f su F.

7. Nelle ipotesi dell'esercizio 6 le radici di f sono reali, ma le formule di Cardano le esprimono mediante numeri non reali. Dimostrare che ciò è inevitabile: non esiste un sottocampo K/F di $\mathbb{R}/F$ che ammetta una torre di radicali su F e contenga un campo di spezzamento di f su F. È il cosiddetto *casus irreducibilis* delle equazioni reali.

8. Mostrare che $H = \{1, (12345), (13524), (14253), (15432), (14)(23), (15)(24), (25)(34), (12)(35), (13)(45)\}$ è un sottogruppo risolubile di A_5. Sia $f(x) = x^5 - t_1 x^4 + \cdots$ la quintica generale con radici $x_1, \ldots, x_5$ su $F(t_1, \ldots, t_5)$. Sia d il discriminante e sia $K = F(t_1, \ldots, t_5, \sqrt{d})$. Poniamo

$$\chi_1 = x_1x_2 + x_2x_3 + x_3x_4 + x_4x_5 + x_5x_1,$$
$$\chi_1' = x_1x_3 + x_1x_4 + x_2x_4 + x_2x_5 + x_3x_5$$

   e $\omega_1 = \chi_1 - \chi_1'$. Mostrare che H fissa ω_1 e determinarne i coniugati sotto il gruppo di Galois di $F(x_1, \ldots, x_5)/K$. Mostrare che $[K(\omega_1) : K] = 6$.

9. Mostrare che il discriminante di $f(x) = x^5 + px + q$ è $d = 2^8p^5 + 5^5q^4$. Siano $\rho_1, \ldots, \rho_5$ le sue radici e supponiamo $d \neq 0$. Poniamo

$$\lambda_1 = \rho_1\rho_2 + \rho_2\rho_3 + \rho_3\rho_4 + \rho_4\rho_5 + \rho_5\rho_1,$$
$$\lambda_1' = \rho_1\rho_3 + \rho_1\rho_4 + \rho_2\rho_4 + \rho_2\rho_5 + \rho_3\rho_5 = -\lambda_1$$

   e $\mu_1 = \lambda_1 - \lambda_1' = 2\lambda_1$. Mostrare che μ_1 è una radice della *risolvente sestica* di $f(x)$,

$$g(x) = (x^3 - 5px^2 + 15p^2x + 5p^3)^2 - dx,$$

   e che, in caratteristica zero, $f(x)$ è risolubile per radicali su $F(\sqrt{d}, \mu_1)$.

   Nota alla traduzione. Nell'enunciato originale compare μ_1. Con la normalizzazione del polinomio qui riportata, la quantità da sostituire in g è $\lambda_1^2 = \mu_1^2/4$. Si vede già il problema nell'esempio $f = x^5 - x$, con le radici ordinate come $0, 1, -1, i, -i$: si ha $\lambda_1 = -i$, $g(-1) = 0$ e $g(-2i) = 141 + 588i$.

4.10 Equazioni a coefficienti razionali aventi per gruppo di Galois il gruppo simmetrico

Il teorema di Ruffini–Abel afferma che le equazioni generali di grado $n \geq 5$ non sono risolubili per radicali. In termini approssimativi, è impossibile dare per $n \geq 5$ una formula generale per radicali che, sostituendo ai coefficienti valori di un campo F, fornisca le radici di qualsiasi equazione di grado n su F. È tuttavia concepibile che tutte le equazioni a coefficienti in un dato campo F siano risolubili per radicali su F. In certi casi ciò avviene, banalmente ad esempio per $F = \mathbb{R}$. Mostreremo che per $F = \mathbb{Q}$ e per ogni primo p esistono polinomi $f(x) \in \mathbb{Q}[x]$ con gruppo di Galois S_p. Per $p \geq 5$ non sono risolubili per radicali. Proviamo dapprima il seguente risultato sui gruppi di permutazioni.

Lemma. *Se G è un gruppo di permutazioni di un numero primo p di elementi e contiene un elemento di ordine p e una trasposizione, allora $G = S_p$.*

Dimostrazione. L'ordine di un ciclo $(1\,2\,\ldots\,m)$ è m e l'ordine di un prodotto di cicli disgiunti è il minimo comune multiplo dei loro ordini (p. 48 dell'originale). Quindi G contiene un p-ciclo $\sigma = (i_1\,i_2\,\ldots\,i_p)$ con $\{i_1,\ldots,i_p\} = \{1,\ldots,p\}$. Riordinando gli elementi si può supporre che la trasposizione in G sia (12). Una potenza opportuna di σ ha la forma $(12\cdots)$; riordinando ulteriormente gli altri elementi si può dunque supporre che G contenga (12) e $\sigma = (1\,2\,3\,\ldots\,p)$. Allora contiene $\sigma(12)\sigma^{-1} = (23)$, $\sigma(23)\sigma^{-1} = (34)$, $\ldots$, $\sigma(p-2,\,p-1)\sigma^{-1} = (p-1,\,p)$. Si vede facilmente che queste trasposizioni generano S_p. Pertanto $G = S_p$. □

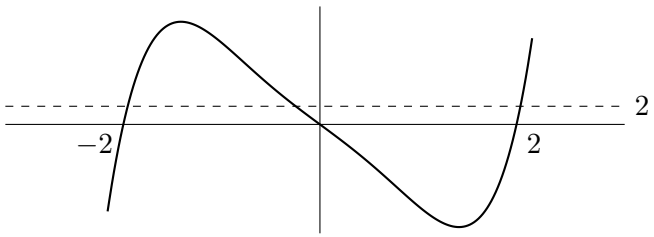
Teorema 4.16. *Sia $f(x)$ un polinomio di grado primo p a coefficienti razionali, irriducibile su $\mathbb{Q}$. Se $f(x) = 0$ ha esattamente due radici non reali in $\mathbb{C}$, il suo gruppo G_f su $\mathbb{Q}$ è S_p.*

Dimostrazione. Ammettiamo il risultato classico, che proveremo nel §5.1, per cui, normalizzato f a essere monico, $f(x) = \prod_{i=1}^{p}(x - r_i)$ in $\mathbb{C}[x]$. Quindi $E = \mathbb{Q}(r_1,\ldots,r_p)$ è un campo di spezzamento di f su $\mathbb{Q}$ contenuto in $\mathbb{C}$. Siccome $E \supseteq \mathbb{Q}(r_1)$ e $[\mathbb{Q}(r_1) : \mathbb{Q}] = \deg f = p$, p divide $[E : \mathbb{Q}]$. Per il teorema di Sylow G_f contiene un elemento di ordine p. Consideriamo in $\mathbb{C}$ l'automorfismo di coniugio $u = a + b\sqrt{-1} \mapsto \bar{u} = a - b\sqrt{-1}$, con a, b reali. Esso fissa $f(x)$ e ne permuta quindi le radici. Siano r_1, r_2 quelle non reali. Poiché $f(\bar{r}_1) = 0$ e $\bar{r}_1 \neq r_1$, si deve avere $r_2 = \bar{r}_1$. Dunque il coniugio scambia r_1, r_2 e fissa tutte le altre radici. La sua restrizione a E dà una trasposizione in G_f. Per il lemma $G_f = S_p$. □

Vediamo come costruire polinomi che soddisfino le ipotesi del teorema.[9] Siano m un intero positivo pari e $n_1 < \cdots < n_{k-2}$ interi pari, con $k > 3$ dispari. Consideriamo il polinomio

$$g(x) = (x^2 + m)(x - n_1)(x - n_2)\cdots(x - n_{k-2}). \tag{52}$$

Le radici reali di g sono $n_1,\ldots,n_{k-2}$. Il grafico di $y = g(x)$ ha l'andamento illustrato.



$$g(x) = x(x^2 + 2)(x + 2)(x - 2)$$

Esso presenta $(k-3)/2$ massimi relativi e, poiché $|g(h)| > 2$ per ogni intero dispari h, i valori di tali massimi sono maggiori di 2. Quindi $f(x) = g(x) - 2$ ha $(k-3)/2$ massimi relativi positivi fra n_1 e n_{k-2}. Ne segue che f ha $k-3$ radici reali nell'intervallo (n_1, n_{k-2}). Siccome $f(n_{k-2}) = -2$ e $f(x) \to +\infty$ per $x \to +\infty$, vi è anche una radice maggiore di n_{k-2}. Otteniamo così $k-2$ radici reali di f. Scriviamo $f(x) = \prod_{i=1}^{k}(x - r_i)$ in $\mathbb{C}[x]$. Confrontando i coefficienti di x^{k-1}, x^{k-2} con quelli di $(x^2 + m)(x - n_1)\cdots(x - n_{k-2}) - 2$ risulta

$$\sum_{i=1}^{k} r_i = \sum_{l=1}^{k-2} n_l, \qquad \sum_{i<j} r_i r_j = \sum_{l<q} n_l n_q + m. \tag{53}$$

Dunque

$$\sum_i r_i^2 = \left(\sum_i r_i\right)^2 - 2\sum_{i<j} r_i r_j = \sum_l n_l^2 - 2m. \tag{54}$$

[9] La costruzione che segue è dovuta a R. Brauer. Una costruzione di un polinomio con gruppo di Galois su $\mathbb{Q}$ uguale a un qualsiasi S_n sarà data nell'esercizio 5, p. 305 dell'originale.

Scegliendo m sufficientemente grande, la (54) dà $\sum_i r_i^2 < 0$, sicché non tutti gli r_i sono reali. Se r_1 non è reale, $\bar{r}_1 \neq r_1$ è un'altra radice non reale. Dato che abbiamo già trovato $k-2$ radici reali, le radici non reali sono precisamente due. Scriviamo $f(x) = x^k + a_1 x^{k-1} + \cdots + a_k$: tutti gli a_i sono pari. Inoltre il termine costante di g è divisibile per 4, perciò quello di $f = g - 2$ non lo è. Per il criterio di Eisenstein applicato al primo 2 (esercizio 2, p. 154 dell'originale) f è irriducibile in $\mathbb{Q}[x]$. Le ipotesi del teorema 4.16 sono così soddisfatte per ogni primo $p = k \geq 5$. Possiamo quindi costruire equazioni a coefficienti razionali con gruppo S_p per ogni primo $p \geq 5$. Per $p = 2, 3$ la costruzione è facile e il risultato vale per tutti i primi.

Ciò suggerisce una domanda interessante. Dato un gruppo finito G, esiste un'equazione a coefficienti razionali il cui gruppo di Galois su $\mathbb{Q}$ sia isomorfo a G? Si tratta di un problema difficilissimo, che, pur essendo stato posto per la prima volta circa un secolo fa, resta irrisolto. I primi risultati generali danno una risposta affermativa per $G = S_n$ e $G = A_n$, qualunque sia n. Nel 1954, usando profondi risultati aritmetici, I. R. Šafarevič dimostrò che la risposta è affermativa per ogni gruppo finito risolubile. Più di recente si sono ottenuti altri risultati di questo tipo.

D. Hilbert avviò un metodo generale per affrontare il problema, sviluppato poi da Emmy Noether.[10] Il metodo di Hilbert–Noether conduce al seguente teorema: per un gruppo G la risposta al problema è affermativa se è affermativa la risposta alla domanda che segue. Realizziamo G come sottogruppo di S_n e sia F il sottocampo degli invarianti di G in $\mathbb{Q}(x_1, \ldots, x_n)$, dove gli x_i sono indeterminate e S_n agisce mediante gli automorfismi che le permutano in tutti i modi. Il campo F è isomorfo a $\mathbb{Q}(x_1, \ldots, x_n)$? Il risultato della sezione precedente mostra che per $G = S_n$ la risposta è sì: in questo caso $F = \mathbb{Q}(p_1, \ldots, p_n)$, dove i polinomi simmetrici elementari p_i sono algebricamente indipendenti. Si sa molto poco su questa domanda relativa ai sottocampi di $\mathbb{Q}(x_1, \ldots, x_n)$. Per esempio, non si conosce la risposta generale per $G = A_n$. R. Swan ha mostrato recentemente che per certi gruppi ciclici (per esempio il gruppo ciclico di ordine 47) la risposta è negativa.[11] Il risultato negativo di Swan non risponde negativamente al problema originario sulle equazioni razionali con gruppo di Galois assegnato. Mostra però che il metodo di Hilbert–Noether non può fornire la risposta affermativa in tutti i casi. Non discuteremo qui questi risultati. Li abbiamo ricordati soprattutto per dissipare l'impressione che la teoria di Galois, per il fatto di avere una lunga storia, sia un argomento concluso.[12]

Esercizio.

1. (Masuda.) Sia F un campo e n un intero positivo. Supponiamo che F contenga n radici n-esime distinte di 1 (quindi la caratteristica di F non divide n). Sia $K = F(x_1, \ldots, x_n)$ con gli x_i indeterminate e sia σ l'automorfismo di K/F che le permuta ciclicamente: $\sigma x_i = x_{i+1}$ per $1 \leq i < n$, $\sigma x_n = x_1$. Poniamo $G = \langle \sigma \rangle$, $E = \operatorname{Inv} G$, sicché K/E è di Galois e $\operatorname{Gal}(K/E) = G$ (teorema 4.7, p. 238 dell'originale). Sia ζ una radice primitiva n-esima dell'unità e poniamo $y_j = \sum_{i=1}^{n} \zeta^{ij} x_i$ (cfr. lemma 3, p. 253 dell'originale). Posto $c_{jk} = y_j y_k y_{j+k}^{-1}$, mostrare che $E = F(c_{11}, c_{12}, \ldots, c_{1n})$.

[10] Se ne tratta in N. Tschebotarëw, *Grundzüge der Galoischen Theorie* (traduzione dal russo al tedesco di H. Schwerdtfeger), Groningen, 1950, p. 399.

[11] R. Swan, «Invariant rational functions and a problem of Steenrod», *Inventiones Mathematicae* 7 (1969), pp. 148–158.

[12] Il teorema fondamentale della teoria di Galois è stato generalizzato in varie direzioni e continua a essere oggetto di ricerca. La bibliografia è troppo vasta per essere indicata qui. Il lettore può consultare la rivista di recensioni *Mathematical Reviews*. Si veda anche la rassegna di Swan, «Noether's problem in Galois theory», nel volume *Emmy Noether in Bryn Mawr*, a cura di J. Sally e B. Srinivasan, Springer-Verlag, New York, 1982.

4.11 Poligoni regolari costruibili. Campi ciclotomici su $\mathbb{Q}$

Torniamo al problema della costruibilità euclidea esaminato nel §4.2. Il risultato principale era il seguente criterio: il numero complesso z è costruibile con riga e compasso a partire da $z_1, \dots, z_n$ se e solo se è contenuto in un campo K che ammette una torre di radici quadrate su $F = \mathbb{Q}(z_1, \dots, z_n, \bar{z}_1, \dots, \bar{z}_n)$. Possiamo ora precisarlo.

Criterio B. *Il numero complesso z è costruibile con riga e compasso a partire da $z_1, \dots, z_n$ se e solo se z è algebrico su $F = \mathbb{Q}(z_1, \dots, z_n, \bar{z}_1, \dots, \bar{z}_n)$ e la chiusura normale K/F di $F(z)/F$ ha grado una potenza di due.*

Dimostrazione. Se z è costruibile, è contenuto in un sottocampo L/F di $\mathbb{C}$ dotato di una torre di radici quadrate su F. Per il lemma 5 del §4.7 possiamo supporre che L sia di Galois su F. Contiene allora la chiusura normale K/F di $F(z)/F$. Dall'esistenza della torre segue $[L : F] = 2^s$. Pertanto $[K : F]$, divisore di $[L : F]$, è della forma 2^t. Viceversa, se $[K : F] = 2^t$, il gruppo $G = \mathrm{Gal}(K/F)$ ha ordine 2^t, è risolubile e possiede una serie di composizione $G = G_1 \triangleright G_2 \triangleright \cdots \triangleright G_{t+1} = 1$ con ciascun fattore G_i/G_{i+1} ciclico di ordine due. Le corrisponde la torre $F = F_1 \subset F_2 \subset \cdots \subset F_{t+1} = K$ con $[F_{i+1} : F_i] = 2$. Allora $F_{i+1} = F_i(u_i)$ con $u_i^2 - a_i u_i + b_i = 0$, $a_i, b_i \in F_i$. Sostituendo u_i con $v_i = u_i - \frac{1}{2}a_i$ si ottiene $F_{i+1} = F_i(v_i)$, $v_i^2 \in F_i$. Quindi K possiede una torre di radici quadrate e, poiché $z \in K$, il primo criterio dà la costruibilità di z. □

Applichiamo questo risultato per determinare gli n per cui il poligono regolare di n lati è costruibile con riga e compasso. Occorre calcolare $[\Lambda^{(n)} : \mathbb{Q}]$, dove $\Lambda^{(n)}$ indica il campo ciclotomico delle radici n-esime dell'unità su $\mathbb{Q}$. Il loro insieme U è un gruppo ciclico di ordine n rispetto alla moltiplicazione. Quindi il numero delle radici primitive, ossia dei generatori di U, è $\varphi(n)$ (esercizio 4, p. 47 dell'originale). Per ciascuna di esse, z, si ha $\Lambda^{(n)} = \mathbb{Q}(z)$ e il grado cercato è quello del suo polinomio minimo su $\mathbb{Q}$. Poniamo

$$\lambda_n(x) = \prod_{z \text{ radice primitiva } n\text{-esima di } 1} (x - z). \tag{55}$$

Se $\eta \in \mathrm{Gal}(\Lambda^{(n)}/\mathbb{Q})$ e z è primitiva, anche $\eta(z)$ lo è. Perciò $\eta(\lambda_n(x)) = \lambda_n(x)$ e $\lambda_n(x) \in \mathbb{Q}[x]$. È chiaro che $\lambda_n(x)$ divide $x^n - 1$; più precisamente, poiché ogni radice n-esima dell'unità ha un ordine d che divide n,

$$x^n - 1 = \prod_{d|n} \lambda_d(x). \tag{56}$$

Teorema 4.17. *Il polinomio $\lambda_n(x)$ è irriducibile in $\mathbb{Q}[x]$.*

Dimostrazione. Osserviamo anzitutto che λ_n ha coefficienti interi. Ciò vale per $n = 1$. Supponendolo vero per $d < n$, scriviamo $x^n - 1 = \lambda_n(x)g(x)$ con $g(x) = \prod_{d|n,\, d<n} \lambda_d(x)$ monico a coefficienti interi. Dividendo per il polinomio monico g si ottengono polinomi interi q, r con $\deg r < \deg g$ e $x^n - 1 = qg + r$. Poiché quoziente e resto sono unici anche in $\mathbb{Q}[x]$, risulta $\lambda_n = q \in \mathbb{Z}[x]$. Supponiamo ora

$$\lambda_n(x) = h(x)k(x), \tag{57}$$

con $h, k \in \mathbb{Z}[x]$ e h irriducibile in $\mathbb{Z}[x]$, quindi in $\mathbb{Q}[x]$ (p. 153 dell'originale). Possiamo supporre h, k monici e $\deg h \geq 1$. Sia p un primo che non divide n e sia z una radice di h. Poiché $(p, n) = 1$, z^p è ancora una radice primitiva n-esima. Se non fosse radice di h, lo sarebbe di k, e z sarebbe radice di $k(x^p)$. Da $(h(x), k(x^p)) \neq 1$ e dall'irriducibilità di h seguirebbe $h(x) \mid k(x^p)$.

Come all'inizio della dimostrazione, $k(x^p) = h(x)l(x)$ con l monico a coefficienti interi. Inoltre $x^n - 1 = hkg$. Passando alle congruenze modulo p, ossia all'anello $(\mathbb{Z}/(p))[x]$, otteniamo

$$x^n - \bar{1} = \bar{h}(x)\bar{k}(x)\bar{g}(x), \tag{58}$$

dove per $f(x) = a_0x^m + a_1x^{m-1} + \cdots + a_m \in \mathbb{Z}[x]$ si pone $\bar{f}(x) = \bar{a}_0x^m + \bar{a}_1x^{m-1} + \cdots + \bar{a}_m$, $\bar{a}_i = a_i + (p)$. Parimenti $\bar{k}(x^p) = \bar{h}(x)\bar{l}(x)$. Poiché $\bar{a}^p = \bar{a}$ per ogni $a \in \mathbb{Z}$, si ha

$$\bar{f}(x)^p = \bar{a}_0^p x^{pm} + \bar{a}_1^p x^{p(m-1)} + \cdots + \bar{a}_m^p = \bar{a}_0 x^{pm} + \bar{a}_1 x^{p(m-1)} + \cdots + \bar{a}_m = \bar{f}(x^p).$$

Dunque $\bar{k}(x)^p = \bar{k}(x^p) = \bar{h}(x)\bar{l}(x)$, da cui $(\bar{h}, \bar{k}) \neq 1$. Per (58), $x^n - \bar{1}$ avrebbe radici multiple nel suo campo di spezzamento su $\mathbb{Z}/(p)$. Ma la derivata è $\bar{n}x^{n-1}$ e $\bar{n} \neq 0$, sicché il polinomio è primo con la derivata, in contrasto con il criterio per le radici multiple. Dunque z^p è radice di h per ogni primo $p \nmid n$. Iterando, z^r è radice di h per ogni intero positivo r primo con n. Ogni radice primitiva n-esima ha la forma z^r, $(r, n) = 1$, perciò h è divisibile per ogni $x - z'$ con z' primitiva. Ne segue $h = \lambda_n$, che è irriducibile in $\mathbb{Q}[x]$. □

Pertanto λ_n è il polinomio minimo su $\mathbb{Q}$ di qualsiasi radice primitiva n-esima e $\varphi(n) = \deg \lambda_n = [\Lambda^{(n)} : \mathbb{Q}]$. Il teorema generalizza il risultato già provato per p primo: $\lambda_p(x) = (x^p - 1)/(x - 1) = x^{p-1} + x^{p-2} + \cdots + 1$ è irriducibile in $\mathbb{Q}[x]$.

Scriviamo $n = 2^{e_1}p_2^{e_2} \cdots p_s^{e_s}$ con i p_i primi dispari distinti, $e_1 \geq 0$ ed $e_i > 0$ per $i > 1$. Allora

$$\varphi(n) = \varphi(2^{e_1}) \prod_{i=2}^{s} \varphi(p_i^{e_i}) = \begin{cases} 2^{e_1-1} \prod_{i=2}^{s} p_i^{e_i-1}(p_i - 1), & e_1 > 0, \\ \prod_{i=2}^{s} p_i^{e_i-1}(p_i - 1), & e_1 = 0. \end{cases} \tag{59}$$

Ne risulta che $\varphi(n)$ è una potenza di due se e solo se i fattori primi dispari di n sono primi di Fermat e ciascuno compare con molteplicità uno. La costruibilità con riga e compasso del poligono regolare di n lati equivale a quella della radice primitiva $z = e^{2\pi i/n}$. Poiché $\mathbb{Q}(z)$ è di Galois e $[\mathbb{Q}(z) : \mathbb{Q}] = \varphi(n)$, il criterio B fornisce il seguente risultato, dovuto a Gauss.

Teorema 4.18. *Il poligono regolare di n lati è costruibile con riga e compasso se e solo se $n = 2^e p_2 \cdots p_s$, dove $e \geq 0$ e i p_i sono primi di Fermat distinti.*

La formula (56) dà un algoritmo per calcolare λ_n, che chiameremo *polinomio ciclotomico*. Si comincia con $\lambda_1(x) = x - 1$. Conoscendo già i λ_d per i divisori propri d di n, si ricava dalla (56) λ_n. Per esempio,

$$\begin{aligned} \lambda_2(x) &= (x^2 - 1)/\lambda_1(x) = x + 1, \\ \lambda_3(x) &= (x^3 - 1)/\lambda_1(x) = x^2 + x + 1, \\ \lambda_4(x) &= (x^4 - 1)/(\lambda_1(x)\lambda_2(x)) = x^2 + 1, \\ \lambda_6(x) &= (x^6 - 1)/(\lambda_1(x)\lambda_2(x)\lambda_3(x)) = x^2 - x + 1, \\ \lambda_{12}(x) &= (x^{12} - 1)/(\lambda_1(x)\lambda_2(x)\lambda_3(x)\lambda_4(x)\lambda_6(x)) = x^4 - x^2 + 1. \end{aligned}$$

Completiamo la trattazione dei campi ciclotomici su $\mathbb{Q}$ determinando i loro gruppi di Galois, ossia i gruppi di Galois dei $\lambda_n(x)$ su $\mathbb{Q}$. L'ordine è il grado dell'irriducibile λ_n, cioè $\varphi(n)$. Dalla dimostrazione del lemma 1 del §4.7 segue quindi che il gruppo del campo ciclotomico di ordine n su $\mathbb{Q}$ è isomorfo al gruppo moltiplicativo U_n delle unità di $\mathbb{Z}/(n)$. Se $n = p$ è primo, sappiamo che è ciclico di ordine $\varphi(p) = p - 1$. Determiniamone la struttura per n composto.

Come indicato negli esercizi 9 e 10, p. 110 dell'originale, per $n = p_1^{e_1} \cdots p_s^{e_s}$ con i p_i primi distinti, U_n è isomorfo al prodotto diretto dei $U_{p_i^{e_i}}$. Basta pertanto considerare U_{p^e}, p primo. Cominciamo dal caso dispari.

Teorema 4.19. *Se p è un primo dispari, il gruppo moltiplicativo U_{p^e} delle unità di $\mathbb{Z}/(p^e)$ è ciclico.*

Dimostrazione. L'ordine di $G = U_{p^e}$ è $p^{e-1}(p-1)$. Quindi G è prodotto diretto del sottogruppo H di ordine p^{e-1}, formato dagli elementi con $x^{p^{e-1}} = 1$, e del sottogruppo K di ordine $p-1$, formato da quelli con $x^{p-1} = 1$.[13] Basta mostrare che H e K sono ciclici, poiché il prodotto diretto di gruppi ciclici di ordini coprimi è ciclico. Essendo U_p ciclico, possiamo scegliere un intero a con $a + (p), a^2 + (p), \ldots, a^{p-1} + (p)$ distinti in $\mathbb{Z}/(p)$. Poniamo $b = a^{p^{e-1}}$. Poiché $(a, p) = 1$, si ha $(b, p^e) = 1$, e $b + (p^e), a + (p^e) \in G$. Inoltre $b^{p-1} = a^{\varphi(p^e)} \equiv 1 \pmod{p^e}$, quindi $b + (p^e) \in K$. Da $b = a^{p^{e-1}} \equiv a \pmod{p}$ segue che le prime $p-1$ potenze di b sono distinte modulo p, quindi anche modulo p^e. Allora $b + (p^e)$ ha ordine $p-1$ e genera K.

Resta da mostrare che H è ciclico. Possiamo supporre $e \geq 2$, poiché altrimenti $H = 1$. Per il teorema di struttura, H è prodotto diretto di $k \geq 1$ gruppi ciclici di ordini p^{e_i}, $e_i \geq 1$. Il numero di soluzioni di $x^p = 1$ in H è p^k. Basta dunque mostrare che vi sono al più p interi n, $0 < n < p^e$, per cui $n^p \equiv 1 \pmod{p^e}$. Un tale n soddisfa $n \equiv 1 \pmod{p}$ perché $n^p \equiv n \pmod{p}$. Se $n \neq 1$, scriviamo $n = 1 + yp^f + zp^{f+1}$ con $1 \leq f \leq e-1$, $0 < y < p$ e z intero non negativo. Allora

$$\begin{aligned} n^p &= 1 + \binom{p}{1}(y + zp)p^f + \binom{p}{2}(y + zp)^2 p^{2f} + \cdots + (y + zp)^p p^{pf} \\ &\equiv 1 + yp^{f+1} \pmod{p^{f+2}}. \end{aligned}$$

Se $f < e-1$, dalla congruenza $n^p \equiv 1 \pmod{p^e}$ seguirebbe $yp^{f+1} \equiv 0 \pmod{p^{f+2}}$ e quindi $y \equiv 0 \pmod{p}$, contro $0 < y < p$. Dunque, per $1 < n < p^e$, le sole possibilità sono $n = 1 + yp^{e-1}$ con $0 < y < p$. Includendo 1 si ottengono al più p soluzioni, come richiesto. □

Passiamo ora al primo 2.

Teorema 4.20. *U_2 e U_4 sono ciclici. Per $e \geq 3$, U_{2^e} è prodotto diretto di un gruppo ciclico di ordine 2 e di uno di ordine 2^{e-2}.*

Dimostrazione. L'ordine di $G = U_{2^e}$ è $\varphi(2^e) = 2^{e-1}$. Se $e = 1$ si ha $|G| = 1$, se $e = 2$ si ha $|G| = 2$, e in entrambi i casi G è ciclico. Supponiamo $e \geq 3$. Mostriamo dapprima che $x^2 = 1$ ha quattro soluzioni distinte in G. Ne seguirà che G è prodotto diretto di almeno due gruppi ciclici non banali. Poniamo $a_1 = 1$, $a_2 = -1$, $a_3 = 1 + 2^{e-1}$, $a_4 = -1 + 2^{e-1}$ e $x_i = a_i + (2^e)$. Gli x_i sono quattro elementi distinti di G con $x_i^2 = 1$. Poiché $|G| = 2^{e-1}$ e i fattori ciclici non banali sono almeno due, per ogni $x \in G$ si ha $x^{2^{e-2}} = 1$; equivalentemente, per ogni intero dispari a, $a^{2^{e-2}} \equiv 1 \pmod{2^e}$. Basterà esibire un x per cui $x^{2^{e-3}} \neq 1$. Avremo allora un sottogruppo ciclico di ordine 2^{e-2} e questo, nella decomposizione di G, richiede un fattore ciclico di ordine 2^{e-2} e uno di ordine 2.

Prendiamo $x = 5 + (2^e)$. Per $e = 3$, $5^{2^{e-3}} = 5 \not\equiv 1 \pmod{8}$ ma $5 \equiv 1 \pmod{4}$. Per $f \geq 3$ sia $k(f)$ il massimo intero k tale che $5^{2^{f-3}} \equiv 1 \pmod{2^k}$. Allora $k(3) = 2$. Scrivendo $5^{2^{f-3}} = 1 + y2^{k(f)}$ con y dispari, otteniamo

$$5^{2^{f+1-3}} = (5^{2^{f-3}})^2 = 1 + y2^{k(f)+1} + y^2 2^{2k(f)}.$$

Di qui $k(f+1) \geq k(f)$ e quindi $k(f) \geq 2$ per $f \geq 3$. La stessa relazione dà $5^{2^{f+1-3}} = 1 + z2^{k(f)+1}$ con $z = y + 2^{k(f)-1}y^2$ dispari. Dunque $k(f+1) = k(f) + 1$, perciò $k(f) = f - 1$ per ogni $f \geq 3$. In particolare $5^{2^{e-3}} \not\equiv 1 \pmod{2^e}$ per $e \geq 3$, come richiesto. □

I due teoremi precedenti descrivono il gruppo di Galois del campo delle radici p^e-esime dell'unità sui razionali.

[13] Qui e nel resto della sezione occorre il teorema di struttura dei gruppi abeliani finiti, dimostrato nel §3.10, p. 195 dell'originale.

Teorema 4.21. *Il gruppo di Galois del campo ciclotomico delle radici p^e-esime dell'unità su $\mathbb{Q}$, ossia il gruppo di Galois di $\lambda_{p^e}(x) = 0$, è ciclico, salvo che per $p = 2$ ed $e \geq 3$. In quest'ultimo caso è prodotto diretto di un gruppo ciclico di ordine 2 e di uno di ordine 2^{e-2}.*

Esercizi.

1. Con la formula di inversione di Möbius (esercizio 18, p. 151 dell'originale) dimostrare che $\lambda_n(x) = \prod_{d|n}(x^d - 1)^{\mu(n/d)}$.

2. Siano $r_1, \ldots, r_n$ le radici distinte di un polinomio monico $f(x)$. Mostrare che il discriminante è
$$d = \prod_{1 \leq i < j \leq n} (r_i - r_j)^2 = (-1)^{n(n-1)/2} \prod_{i=1}^{n} f'(r_i),$$
dove f' è la derivata. Sia $\lambda_p(x) = x^{p-1} + x^{p-2} + \cdots + 1$, p primo. Derivando $x^p - 1 = (x-1)\lambda_p(x)$ si ricava $px^{p-1} = \lambda_p(x) + (x-1)\lambda_p'(x)$. Usare questa relazione e la formula precedente per provare che il discriminante di λ_p è $d = (-1)^{p(p-1)/2}p^{p-2}$ per p dispari.

3. Sia Λ_p il campo delle radici p-esime dell'unità su $\mathbb{Q}$, con p primo dispari. Mostrare che Λ_p ha un unico sottocampo quadratico $E/\mathbb{Q}$ e che E è reale (sottocampo di $\mathbb{R}$) oppure non reale secondo che p abbia la forma $4n + 1$ oppure $4n + 3$.

4. Usare l'esercizio 2, p. 243 dell'originale, e il teorema 4.21 per mostrare che per ogni gruppo ciclico finito G esiste un sottocampo di un campo ciclotomico su $\mathbb{Q}$ avente G come gruppo di Galois.

Osservazione. Un teorema classico di Dirichlet afferma che ogni progressione aritmetica $a + kd$, dove a, d sono interi positivi coprimi e $k = 0, 1, 2, \ldots$, contiene infiniti primi. Il caso particolare $a = 1$, $d = p$ primo, insieme al teorema di struttura dei gruppi abeliani finiti (teorema 3.13), permette di mostrare che ogni gruppo abeliano finito è immagine omomorfa del gruppo delle unità U_n di un opportuno $\mathbb{Z}/(n)$ con n privo di fattori quadrati. Da questo risultato e dal teorema 4.21 si può dedurre l'esistenza di un'estensione di Galois $E/\mathbb{Q}$ con gruppo di Galois abeliano finito assegnato. La dimostrazione del teorema di Dirichlet richiede la teoria delle funzioni. Il caso particolare delle progressioni $1 + kn$ ammette però una dimostrazione elementare (si veda T. Nagell, *Introduction to Number Theory*, Wiley, New York, 1951, p. 118).

4.12 Trascendenza di e e di π. Il teorema di Lindemann–Weierstrass

Dimostreremo che π è trascendente, ossia non è algebrico su $\mathbb{Q}$. Ne seguirà che π e $\sqrt{\pi}$ non sono numeri costruibili. Sarà dunque impossibile costruire con riga e compasso un segmento lungo quanto la circonferenza di un cerchio di raggio dato, oppure lungo quanto il lato di un quadrato avente la stessa area di un cerchio dato. Con poco sforzo in più possiamo dimostrare un risultato assai più generale.

Teorema di Lindemann–Weierstrass. *Se $u_1, \ldots, u_n$ sono numeri algebrici (ossia numeri complessi algebrici su $\mathbb{Q}$) linearmente indipendenti su $\mathbb{Q}$, gli esponenziali complessi $e^{u_1}, \ldots, e^{u_n}$ sono algebricamente indipendenti sul campo dei numeri algebrici.*

Per $u \in \mathbb{C}$ si può definire $e^u = 1 + u + u^2/2! + u^3/3! + \cdots$, serie assolutamente convergente per ogni u. Dalla definizione mediante serie di potenze segue facilmente l'equazione funzionale $e^u e^v = e^{u+v}$. Osserviamo inoltre che i numeri algebrici costituiscono un sottocampo di $\mathbb{C}$ (tra

poco dimostreremo un risultato più generale). Prendendo $n = 1$ nel teorema vediamo che, per u algebrico non nullo, e^u è trascendente. In particolare e è trascendente e, poiché $e^{\pi i} = \cos\pi + i\sin\pi = -1$, anche πi è trascendente. Dall'algebricità di i segue quindi la trascendenza di π. Analogamente, se u è un numero algebrico reale diverso da $0, 1$, da $e^{\log u} = u$ segue che $\log u$ è trascendente (per $u > 0$ si può prendere il logaritmo reale).

Mostreremo che il teorema di Lindemann–Weierstrass equivale a un altro teorema, talvolta anch'esso detto di Lindemann–Weierstrass o *teorema generalizzato di Lindemann*.

Teorema 4.22. *Se $u_1, \ldots, u_n$ sono numeri algebrici distinti, gli esponenziali $e^{u_1}, \ldots, e^{u_n}$ sono linearmente indipendenti sul campo dei numeri algebrici.*

Supponiamo vero questo teorema e siano $u_1, \ldots, u_n$ numeri algebrici linearmente indipendenti su $\mathbb{Q}$. Date due successioni distinte $(k_1, \ldots, k_n)$ e $(l_1, \ldots, l_n)$ di interi non negativi, i prodotti $\prod_i (e^{u_i})^{k_i} = e^{\sum_i k_i u_i}$ e $\prod_i (e^{u_i})^{l_i} = e^{\sum_i l_i u_i}$ hanno esponenti distinti. Per il teorema 4.22, date r successioni distinte $(k_{1j}, \ldots, k_{nj})$, i corrispondenti r numeri $\prod_i (e^{u_i})^{k_{ij}}$ sono quindi linearmente indipendenti sui numeri algebrici. Questo significa precisamente che gli e^{u_i} sono algebricamente indipendenti sui numeri algebrici. Il teorema 4.22 implica dunque quello di Lindemann–Weierstrass.

Viceversa, supponiamo vero il teorema di Lindemann–Weierstrass e siano $u_1, \ldots, u_n$ numeri algebrici distinti. Il sottogruppo del gruppo additivo di $\mathbb{C}$ da essi generato è uno $\mathbb{Z}$-modulo libero: è finitamente generato e privo di torsione. Possiede dunque una base di r numeri algebrici $v_1, \ldots, v_r$, $r \leq n$, linearmente indipendenti su $\mathbb{Q}$, tali che $u_i = \sum_{j=1}^r a_{ij} v_j$, $a_{ij} \in \mathbb{Z}$. (Se il sottogruppo è nullo, la tesi è immediata.) Allora $e^{u_i} = \prod_{j=1}^r (e^{v_j})^{a_{ij}}$ e, poiché gli u_i sono distinti, anche i vettori $(a_{i1}, \ldots, a_{ir})$ sono distinti. Se gli e^{u_i} soddisfacessero una relazione lineare non banale a coefficienti algebrici, moltiplicandola per un'opportuna potenza $(e^{v_1} \cdots e^{v_r})^a$, con a intero positivo, otterremmo una relazione algebrica non banale tra gli e^{v_j}, ancora a coefficienti algebrici. Ciò contraddirebbe il teorema di Lindemann–Weierstrass, poiché i v_j sono linearmente indipendenti su $\mathbb{Q}$. I due teoremi sono dunque equivalenti.

Dimostreremo il teorema 4.22. Ci servono alcuni risultati sugli elementi algebrici dei campi e sugli interi algebrici complessi. Possiamo ricavare insieme la maggior parte di essi considerando la situazione seguente. Sia E un campo e sia R un suo sottoanello. Un elemento $u \in E$ si dice *intero su R* se esiste un polinomio monico $f(x) \in R[x]$ con $f(u) = 0$. Se $R = F$ è un sottocampo, ciò significa che u è algebrico su F. Per $E = \mathbb{C}$, i complessi algebrici su $\mathbb{Q}$ si chiamano *numeri algebrici*; quelli interi su $\mathbb{Z}$ si chiamano *interi algebrici*. Questi ultimi sono naturalmente numeri algebrici.

Se $u \in E$ è intero su R, esiste una relazione, con gli a_i in R,

$$u^n = a_0 + a_1 u + \cdots + a_{n-1} u^{n-1}. \tag{60}$$

Sia $M = R1 + Ru + \cdots + Ru^{n-1}$ l'R-sottomodulo di E generato da $1, u, \ldots, u^{n-1}$ (consideriamo E come R-modulo con l'addizione e lo zero del campo e con il prodotto in E come azione degli scalari). Abbiamo $uM \subseteq Ru + Ru^2 + \cdots + Ru^n$. Da (60) $u^n \in M$, dunque $uM \subseteq M$. Ciò dimostra la necessità nel seguente criterio.

Lemma 1. *Un elemento $u \in E$ è intero su R se e solo se esiste un R-sottomodulo finitamente generato M di E che contiene 1 e soddisfa $uM \subseteq M$.*

Dimostrazione. Per la sufficienza, sia $M = Ru_1 + \cdots + Ru_n$ con $1 \in M$ e $uM \subseteq M$. Allora $uu_i = \sum_{j=1}^n a_{ij}u_j$ per $1 \le i \le n$, con $a_{ij} \in R$. Il sistema lineare omogeneo

$$\begin{aligned}(a_{11} - u)x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0,\\ a_{21}x_1 + (a_{22} - u)x_2 + \cdots + a_{2n}x_n &= 0,\\ &\vdots\\ a_{n1}x_1 + a_{n2}x_2 + \cdots + (a_{nn} - u)x_n &= 0\end{aligned} \tag{61}$$

ha dunque la soluzione $(u_1, \ldots, u_n)$. Essa è non banale perché $1 \in M$ e quindi qualche $u_i \neq 0$. Per un noto risultato di algebra lineare $\det(A - uI) = 0$, dove $A = (a_{ij})$. Quindi u è radice del polinomio caratteristico $\det(xI - A)$, monico a coefficienti in R. Ne segue la tesi. □

Per due R-sottomoduli M, N di E indichiamo con MN il sottomodulo generato dai prodotti uv, con $u \in M$, $v \in N$. Esso consiste delle somme finite $\sum_i u_i v_i$ con $u_i \in M$, $v_i \in N$. Se $M = \sum_{i=1}^m Ru_i$ e $N = \sum_{j=1}^n Rv_j$ sono finitamente generati, anche $MN = \sum_{i,j} Ru_i v_j$ lo è. Inoltre, se $wM \subseteq M$, si ha $w(MN) = (wM)N \subseteq MN$. Per induzione, se $M_1, \ldots, M_r$ sono finitamente generati, lo è anche il sottomodulo $M_1 \cdots M_r$ generato da tutti i prodotti $x_1 \cdots x_r$, $x_i \in M_i$; e se $wM_i \subseteq M_i$ per uno degli M_i, allora $wM_1 \cdots M_r \subseteq M_1 \cdots M_r$. Con queste osservazioni e il lemma 1 proviamo il seguente risultato.

Teorema 4.23. *Se E è un campo e R un suo sottoanello, l'insieme A degli elementi di E interi su R è un sottoanello di E contenente R. Inoltre ogni elemento di E intero su A è intero su R, e dunque appartiene ad A.*

Dimostrazione. Siano $u, v \in A$. Esistono R-sottomoduli finitamente generati M, N contenenti 1, con $uM \subseteq M$, $vN \subseteq N$. Allora $(u \pm v)MN \subseteq u(MN) + v(MN) \subseteq MN$ e $(uv)MN \subseteq MN$. Poiché $1 \in MN$, il lemma 1 si applica a $u \pm v$, 1 e uv. Questi elementi sono interi su R, dunque A è un sottoanello e chiaramente contiene R.

Sia ora u intero su A. Esiste $M = Au_1 + \cdots + Au_n$ con $1 \in M$ e $uM \subseteq M$. Possiamo supporre $u_1 = 1$. Scriviamo $uu_i = \sum_j a_{ij}u_j$, $a_{ij} \in A$. Per ogni a_{ij} esiste un R-sottomodulo finitamente generato N_{ij} con $a_{ij}N_{ij} \subseteq N_{ij}$ e $1 \in N_{ij}$. Moltiplicando questi sottomoduli otteniamo $N = Rv_1 + \cdots + Rv_m$, con $v_1 = 1$, tale che $a_{ij}N \subseteq N$ per ogni i, j. Poniamo $P = \sum_{i,j} Ru_i v_j$. Si ha $1 = u_1 v_1 \in P$ e $u(u_i v_k) = \sum_j a_{ij}u_j v_k = \sum_j u_j(a_{ij}v_k)$. Poiché $a_{ij}v_k \in N$, quest'ultima è una combinazione R-lineare degli $u_j v_l$. Pertanto $uP \subseteq P$ e il lemma 1 mostra che u è intero su R. □

Nel caso in cui $R = F$ sia un sottocampo, il risultato afferma che gli elementi di E algebrici su F costituiscono un sottoanello. Inoltre per u algebrico si ha $F(u) = F[u]$; quindi anche u^{-1} è algebrico se $u \neq 0$. L'insieme degli elementi di E algebrici su F è dunque un sottocampo A di E e ogni elemento di E algebrico su A appartiene già ad A.

Poniamo ora $E = \mathbb{C}$ e $R = \mathbb{Q}$ oppure $\mathbb{Z}$. Gli elementi interi su $\mathbb{Q}$ sono i numeri algebrici e quelli interi su $\mathbb{Z}$ sono gli interi algebrici. Vale il criterio seguente.

Lemma 2. *Un numero complesso u è un intero algebrico se e solo se è algebrico e il suo polinomio minimo su $\mathbb{Q}$ appartiene a $\mathbb{Z}[x]$.*

Dimostrazione. La sufficienza è immediata. Se u è un intero algebrico, esiste $f(x) \in \mathbb{Z}[x]$ monico con $f(u) = 0$. Il polinomio minimo $\mu(x)$ di u divide $f(x)$. Poiché $\mathbb{Z}$ è fattoriale, ne segue che $\mu(x) \in \mathbb{Z}[x]$ (corollario al teorema 2.25). □

Teorema 4.24. *Un numero razionale è un intero algebrico se e solo se è intero. Per ogni numero algebrico u esiste un intero non nullo b tale che bu sia un intero algebrico.*

Dimostrazione. Ogni $a \in \mathbb{Z}$ è intero su $\mathbb{Z}$. D'altra parte, per $a \in \mathbb{Q}$ il polinomio minimo è $x - a$; perciò, se a è un intero algebrico, $a \in \mathbb{Z}$. Sia ora $u \in \mathbb{C}$ algebrico su $\mathbb{Q}$ e sia $f(x) = x^n + \alpha_1 x^{n-1} + \cdots + \alpha_n \in \mathbb{Q}[x]$ tale che $f(u) = 0$. Per $b \in \mathbb{Z}$, $b \neq 0$, il numero bu è radice di $b^n f(b^{-1}x) = x^n + b\alpha_1 x^{n-1} + \cdots + b^n \alpha_n$. Prendendo per b il prodotto dei denominatori degli α_i si ottiene un polinomio monico in $\mathbb{Z}[x]$ avente bu per radice. □

Ci servirà il teorema fondamentale dell'algebra: ogni polinomio di grado positivo in $\mathbb{C}[x]$ ha una radice in $\mathbb{C}$. Questo risultato, che proveremo nel §5.1, implica che ogni polinomio monico di grado positivo a coefficienti complessi si fattorizza come $\prod_i (x - r_i)$ in $\mathbb{C}[x]$. In altre parole, $\mathbb{C}$ contiene un campo di spezzamento per ogni polinomio monico diverso da 1 con coefficienti complessi. Quindi, per ogni insieme finito S di numeri algebrici, $\mathbb{Q}(S)$ si può immergere in un'estensione di Galois $K/\mathbb{Q}$ contenuta in $\mathbb{C}$.

Dimostrazione del teorema 4.22. Supponiamo, contro la tesi, che esistano numeri algebrici distinti $u_1, \ldots, u_n$ e numeri algebrici $v_1, \ldots, v_n$, non tutti nulli, tali che

$$v_1 e^{u_1} + v_2 e^{u_2} + \cdots + v_n e^{u_n} = 0. \tag{62}$$

Ne dedurremo dapprima una relazione dello stesso tipo con coefficienti razionali, e poi una relazione della forma

$$v_0 + v_1 \sum_{j=1}^{r} e^{\eta_j(u_1)} + \cdots + v_n \sum_{j=1}^{r} e^{\eta_j(u_n)} = 0, \tag{63}$$

con i v_i interi, $v_0 \neq 0$, e gli η_j elementi del gruppo di Galois di un'estensione di Galois $K/\mathbb{Q}$ contenuta in $\mathbb{C}$ e contenente gli u_i. Infine mostreremo analiticamente che (63) è impossibile.

Per chiarire il passaggio formale da (62) a (63) introduciamo l'algebra del gruppo additivo dei numeri algebrici sul campo dei numeri algebrici. È un caso particolare delle algebre di gruppo definite nell'esercizio 8, p. 127 dell'originale. Per distinguere il campo A dei numeri algebrici dal suo gruppo additivo, indichiamo quest'ultimo con A' e i suoi elementi con u', dove $u \mapsto u'$ è un isomorfismo di $(A, +, 0)$ su A'. Scriviamo moltiplicativamente l'operazione in A'. Allora $a \mapsto a'$ è biunivoca, $a'b' = (a+b)'$, $0'$ è l'unità e $(-a)'$ è l'inverso di a'. L'algebra di gruppo $A[A']$ è l'insieme delle somme finite $\sum_i v_i u_i'$, $v_i \in A$, $u_i' \in A'$, con l'addizione consueta e il prodotto determinato dalla distributività e dalla regola $(v_1 u_1')(v_2 u_2') = v_1 v_2 (u_1 + u_2)'$. Se $u_1, \ldots, u_n$ sono distinti, gli u_i' sono linearmente indipendenti su A: da $\sum_i v_i u_i' = 0$ segue $v_i = 0$ per ogni i. Poiché in $\mathbb{C}$ si ha $e^{u_1} e^{u_2} = e^{u_1 + u_2}$, la proprietà universale dell'algebra di gruppo (esercizio 8 citato) dà un omomorfismo $\varepsilon : A[A'] \to \mathbb{C}$ che manda $\sum_i v_i u_i'$ in $\sum_i v_i e^{u_i}$. Il teorema 4.22 equivale all'iniettività di ε.

L'algebra $A[A']$ è commutativa. Mostriamo che è un dominio. Introduciamo in $\mathbb{C}$ un ordinamento compatibile con l'addizione, il cosiddetto ordinamento lessicografico. Per $x = a + bi$, $y = c + di$, con a, b, c, d reali, poniamo $x > y$ se $a > c$ oppure se $a = c$ e $b > d$. Vale la tricotomia: per ogni x, y vale una sola fra $x > y$, $x = y$, $y > x$. Inoltre da $x > y$ e $z > t$ segue $x + z > y + t$. Siano $\sum_{i=1}^{n} v_i u_i'$ e $\sum_{j=1}^{m} z_j t_j'$ elementi non nulli di $A[A']$. Possiamo supporre $v_1, z_1 \neq 0$, $u_1 > u_2 > \cdots > u_n$ e $t_1 > t_2 > \cdots > t_m$. Il prodotto è $v_1 z_1 (u_1 + t_1)'$ più una somma di termini wq' con $q < u_1 + t_1$. È quindi non nullo e $A[A']$ è un dominio.

Sia $\sum_i v_i u_i' \in \ker \varepsilon$. Scegliamo un sottocampo $K \subseteq \mathbb{C}$ di Galois su $\mathbb{Q}$ che contenga gli u_i, v_i. Le somme $\sum_i x_i y_i'$ con $x_i, y_i \in K$ costituiscono un sottoanello $K[K']$ di $A[A']$. Ogni

$\eta \in G = \mathrm{Gal}(K/\mathbb{Q})$ vi definisce due automorfismi,

$$\sigma(\eta) : \sum_i x_i y_i' \longmapsto \sum_i \eta(x_i) y_i', \qquad \tau(\eta) : \sum_i x_i y_i' \longmapsto \sum_i x_i (\eta(y_i))'.$$

È immediato che sono automorfismi. Supponiamo che $\sum_i v_i u_i' \neq 0$. Per $G = \{\eta_1, \ldots, \eta_m\}$ ogni $\sigma(\eta_j)(\sum_i v_i u_i')$ è non nullo e quindi

$$U = \prod_{j=1}^{m} \sigma(\eta_j) \left(\sum_i v_i u_i' \right) = \prod_{j=1}^{m} \left(\sum_i \eta_j(v_i) u_i' \right) \neq 0.$$

Poiché uno dei fattori è $\sum_i v_i u_i'$, $U \in \ker \varepsilon$. La commutatività implica $\sigma(\eta)U = U$ per ogni $\eta \in G$. Scrivendo $U = \sum_i z_i t_i'$ con t_i' distinti si ha $\sum_i \eta(z_i) t_i' = \sum_i z_i t_i'$; dunque $\eta(z_i) = z_i$ per ogni i, η e $z_i \in \mathbb{Q} = \mathrm{Inv}\, G$. Se il nucleo contiene un elemento non nullo, ne contiene pertanto uno della forma $\sum_i v_i u_i'$ con v_i razionali. Applichiamo a questo gli automorfismi $\tau(\eta_j)$ e formiamo

$$V = \prod_{j=1}^{m} \tau(\eta_j) \left(\sum_i v_i u_i' \right) = \prod_{j=1}^{m} \left(\sum_i v_i (\eta_j(u_i))' \right).$$

Anche V è non nullo, appartiene al nucleo e soddisfa $\tau(\eta)V = V$ per ogni $\eta \in G$. Scrivendo $V = \sum_i z_i t_i'$ con $z_i \in \mathbb{Q}$ si ha $\sum_i z_i (\eta(t_i))' = \sum_i z_i t_i'$. Facciamo la media di queste espressioni:

$$V = \frac{1}{m} \sum_{j=1}^{m} \sum_i z_i (\eta_j(t_i))' = \frac{1}{m} \sum_i z_i \sum_{j=1}^{m} (\eta_j(t_i))' = \sum_i \frac{z_i}{m} T'(t_i),$$

dove, per ogni $t \in K$, $T'(t) = \sum_{j=1}^{m} (\eta_j(t))'$.

Abbiamo mostrato che, se $\ker \varepsilon \neq 0$, vi è un elemento non nullo del nucleo della forma $\sum_i v_i T'(t_i)$ con $v_i \in \mathbb{Q}$. Riunendo i termini simili possiamo supporre $T'(t_i) \neq T'(t_j)$ per $i \neq j$, da cui $t_j \neq \eta(t_i)$ per ogni $\eta \in G$. Per $s, t \in K$ calcoliamo

$$\begin{aligned} T'(s)T'(t) &= \left(\sum_j (\eta_j(s))' \right) \left(\sum_k (\eta_k(t))' \right) = \sum_{j,k} (\eta_j(s))' (\eta_k(t))' \\ &= \sum_{j,k} (\eta_j(s) + \eta_k(t))' = \sum_{j,k} (\eta_j(s + \eta_j^{-1}\eta_k(t)))' \\ &= \sum_k \sum_j (\eta_j(s + \eta_k(t)))' = \sum_k T'(s + \eta_k(t)). \end{aligned}$$

Se $\sum_i v_i T'(t_i) \in \ker \varepsilon$ con $v_i \in \mathbb{Q}$, $v_1 \neq 0$ e $t_j \neq \eta(t_i)$ per ogni $i \neq j$ e ogni η, la moltiplicazione per $T'(-t_1)$ dà quindi un elemento del nucleo della forma

$$v_0 + v_1 T'(u_1) + \cdots + v_n T'(u_n), \tag{64}$$

con $v_i \in \mathbb{Q}$, $v_0 \neq 0$, $u_i \neq 0$. Moltiplicando per un intero opportuno possiamo supporre interi tutti i v_i. Applicando ε a (64) si ottiene (63).

Fin qui il ragionamento è stato puramente algebrico. Passiamo alla parte analitica: dimostriamo che una relazione (63) con v_i interi, $v_0 \neq 0$ e u_i algebrici non nulli porta a contraddizione. Sia $f(x) = \sum_{k=0}^{t} a_k x^k \in \mathbb{Z}[x]$ con $a_0 \neq 0$ un polinomio di cui siano radici tutti gli u_i e quindi anche tutti gli $\eta_j(u_i)$. Dato un primo p poniamo

$$h(x) = x^{p-1} f(x)^p = \sum_{j=p-1}^{s} b_j x^j, \qquad s = tp + p - 1.$$

Si ha $b_j \in \mathbb{Z}$, $b_{p-1} = a_0^p$ e, per $p-1 \le j \le s$,

$$\begin{aligned} j!b_j e^x = &\left[j!b_j + \frac{j!}{1!}b_j x + \cdots + \frac{j!}{(j-p)!}b_j x^{j-p} \right] \\ &+ \left[\frac{j!}{(j-p+1)!}b_j x^{j-p+1} + \cdots + j b_j x^{j-1} \right] \\ &+ b_j x^j \left[1 + \frac{x}{j+1} + \frac{x^2}{(j+1)(j+2)} + \cdots \right]. \end{aligned} \tag{65}$$

La prima parentesi vale zero se $j = p-1$. Se $j \ge p$, si ha $j!/(j-p)! = p!\binom{j}{p}$, che è un multiplo intero di $p!$. A maggior ragione lo è $j!/k!$ per $0 \le k < j-p$. La prima parentesi in (65) è dunque $p!$ per un polinomio a coefficienti interi. Poniamo

$$N_p = \frac{1}{(p-1)!} \sum_{j=p-1}^{s} j!b_j. \tag{66}$$

Sommando (65) per $j = p-1, \ldots, s$, otteniamo

$$\begin{aligned} (p-1)!N_p e^x &= \sum_{j=p-1}^{s} j!b_j e^x \\ &= p!g_p(x) + \sum_{j=p-1}^{s} \left[\frac{j!}{(j-p+1)!}b_j x^{j-p+1} + \cdots + j b_j x^{j-1} \right] \\ &\quad + \sum_{j=p-1}^{s} b_j x^j \left[1 + \frac{x}{j+1} + \frac{x^2}{(j+1)(j+2)} + \cdots \right], \end{aligned} \tag{67}$$

con $g_p(x) \in \mathbb{Z}[x]$. Osserviamo che

$$\begin{aligned} h'(x) &= \sum_{j=p-1}^{s} j b_j x^{j-1}, \qquad h''(x) = \sum_{j=p-1}^{s} \frac{j!}{(j-2)!}b_j x^{j-2}, \\ &\vdots \\ h^{(p-1)}(x) &= \sum_{j=p-1}^{s} \frac{j!}{(j-p+1)!}b_j x^{j-p+1}. \end{aligned}$$

e tutte queste derivate sono divisibili per $f(x)$, poiché $h = x^{p-1}f^p$. La prima somma nella (67), ossia $h' + h'' + \cdots + h^{(p-1)}$, è divisibile per f e si annulla per $x = u_i$.

Stimiamo ora $|R(u_i)|$, dove $R(x)$ è l'ultima somma in (67). Scegliamo il primo $p > 2|u_i|$ per ogni i. Poiché anche $j+1 \ge p$, la serie è maggiorata da una serie geometrica e

$$\left| 1 + \frac{u_i}{j+1} + \frac{u_i^2}{(j+1)(j+2)} + \cdots \right| < 2.$$

Dunque

$$|R(u_i)| < 2 \sum_{j=p-1}^{s} |b_j||u_i|^j \le 2|u_i|^{p-1} \left(\sum_{k=0}^{t} |a_k||u_i|^k \right)^p \le 2M^p,$$

dove M è il massimo dei $2n$ numeri $\sum_{k=0}^{t} |a_k||u_i|^k$ e $\sum_{k=0}^{t} |a_k||u_i|^{k+1}$, $1 \le i \le n$. Pertanto, se $p > 2|u_i|$,

$$|N_p e^{u_i} - p g_p(u_i)| < \frac{2M^p}{(p-1)!}. \tag{68}$$

Se inoltre $p > |a_0|$, l'intero N_p dato da (66) non è divisibile per p, poiché $N_p \equiv b_{p-1} = a_0^p \equiv a_0 \pmod{p}$. Abbiamo così ottenuto il seguente lemma.

Lemma 3. *Siano $u_1, \dots, u_n$ numeri algebrici non nulli e sia $f(x) = \sum_{k=0}^{t} a_k x^k \in \mathbb{Z}[x]$, $a_0 \neq 0$, tale che $f(u_i) = 0$ per ogni i. Sia M il massimo dei $2n$ numeri $\sum_{k=0}^{t} |a_k||u_i|^k$ e $\sum_{k=0}^{t} |a_k||u_i|^{k+1}$ e sia p un primo maggiore di $\max(|a_0|, 2|u_1|, \dots, 2|u_n|)$. Allora esistono un intero N_p non divisibile per p e un polinomio $g_p(x) \in \mathbb{Z}[x]$ di grado minore di tp per cui valgono le disuguaglianze (68).*[14]

Torniamo alla relazione (63), con i v_i interi, $v_0 \neq 0$, e gli u_i numeri algebrici non nulli, radici di un polinomio $f \in \mathbb{Z}[x]$ come nel lemma. Anche gli $\eta_j(u_i)$, $\eta_j \in G = \mathrm{Gal}(K/\mathbb{Q})$, sono radici di f. Per il lemma 3, per tutti i primi p sufficientemente grandi esistono un intero N_p non divisibile per p e un polinomio intero g_p di grado minore di pt, $t = \deg f$, tali che

$$\left| N_p e^{\eta_j(u_i)} - p g_p(\eta_j(u_i)) \right| < 2M^p/(p-1)! \qquad (1 \leq i \leq n,\ 1 \leq j \leq r = |G|).$$

Sia k un intero positivo tale che ku_i^l sia un intero algebrico per ogni u_i e ogni $0 \leq l \leq t$. Un tale k esiste per il teorema 4.24. Allora $k^p g_p(u_i)$ è un intero algebrico e quindi lo è ogni $k^p g_p(\eta_j(u_i))$. Anche $\sum_{j=1}^{r} k^p g_p(\eta_j(u_i))$ è un intero algebrico. Essendo fisso sotto G, è razionale e quindi appartiene a $\mathbb{Z}$ per il criterio appena provato.

Moltiplicando (63) per $N_p k^p$ abbiamo

$$N_p k^p v_0 + \sum_{i=1}^{n} N_p k^p v_i \sum_{j=1}^{r} e^{\eta_j(u_i)} = 0.$$

Pertanto, detto L un maggiorante positivo di $|v_1|, \dots, |v_n|$, si ha

$$\begin{aligned} \left| N_p k^p v_0 + \sum_{i,j} p k^p v_i g_p(\eta_j(u_i)) \right| &= \left| \sum_{i=1}^{n} k^p v_i \left(\sum_{j=1}^{r} p g_p(\eta_j(u_i)) - \sum_{j=1}^{r} N_p e^{\eta_j(u_i)} \right) \right| \\ &\leq k^p \sum_{i,j} |v_i| \left| p g_p(\eta_j(u_i)) - N_p e^{\eta_j(u_i)} \right| \leq \frac{2k^p r n M^p L}{(p-1)!}. \end{aligned}$$

I numeri $pk^p v_i \sum_{j=1}^{r} g_p(\eta_j(u_i))$ sono interi divisibili per p, mentre N_p non è divisibile per p. Per p sufficientemente grande inoltre $p \nmid k$ e $p \nmid v_0$, quindi $p \nmid k^p v_0$. Il membro sinistro è perciò un intero non nullo, mentre il membro destro è positivo e minore di 1 per p sufficientemente grande. La contraddizione mostra che (63) è impossibile. Sono dunque provati il teorema 4.22 e quello di Lindemann–Weierstrass. □

Esercizi.

1. Mostrare che $\sin u$ è trascendente per ogni u algebrico non nullo. (*Suggerimento.* Usare $\sin u = (e^{iu} - e^{-iu})/(2i)$ e la trascendenza di e^{iu}.)

2. Mostrare che $\csc u, \cos u, \sec u, \tan u, \cot u$ sono trascendenti per ogni u algebrico non nullo.

3. Sia m un intero privo di fattori quadrati e sia $F = \mathbb{Q}(\sqrt{m})$ il sottocampo di $\mathbb{C}$ generato da $\sqrt{m}$. Mostrare che F è l'insieme dei numeri della forma $a + b\sqrt{m}$ con $a, b \in \mathbb{Q}$. Sia I l'insieme degli interi algebrici di F. Mostrare che I è un sottoanello di $\mathbb{C}$ e che, nel caso quadratico, consiste degli elementi $a+b\sqrt{m}$ con a, b razionali tali che $2a \in \mathbb{Z}$ e $a^2 - mb^2 \in \mathbb{Z}$.

4. Usare le notazioni dell'esercizio 3. Mostrare che se $m \equiv 2$ o $3 \pmod 4$, I consiste dei numeri $a + b\sqrt{m}$ con $a, b \in \mathbb{Z}$.

5. Con le stesse notazioni, mostrare che se $m \equiv 1 \pmod 4$, I consiste dei numeri $a+b\sqrt{m}$ con a, b entrambi interi oppure entrambi metà di interi dispari. Equivalentemente, mostrare che I è l'insieme dei numeri $a + b(1 + \sqrt{m})/2$ con $a, b \in \mathbb{Z}$.

[14] Sono debitore a E. Schenkman di varie comunicazioni sulla dimostrazione del teorema 4.22 e, in particolare, della forma di questo lemma.

4.13 Campi finiti

Applichiamo ora i risultati della teoria di Galois per ricavare le proprietà principali dei campi finiti. Osserviamo anzitutto che, se F è un campo finito, $|F| = p^n$ per qualche primo p. Sappiamo infatti che il suo campo primo si può identificare con il campo $\mathbb{Z}/(p)$ delle classi di resti modulo un primo p. Consideriamo F come spazio vettoriale su $\mathbb{Z}/(p)$ nel modo consueto. Il grado $[F : \mathbb{Z}/(p)]$ è chiaramente finito. Se vale n, esiste una base $(u_1, \dots, u_n)$ di F su $\mathbb{Z}/(p)$ e ogni elemento di F si scrive in un unico modo come combinazione lineare $a_1u_1 + a_2u_2 + \cdots + a_nu_n$, $a_j \in \mathbb{Z}/(p)$. Ne segue $|F| = p^n$. Lo stesso ragionamento mostra che, se $E \supseteq F$, $[E : F] = n$ e $|F| = q < \infty$, allora $|E| = q^n$. Possiamo ora ricavare rapidamente i risultati fondamentali sui campi finiti. In primo luogo abbiamo il seguente

Teorema 4.25. *Il numero di elementi di un campo finito è una potenza di un primo. Inoltre per ogni potenza di primo $q = p^n$ esiste, a meno di isomorfismi, uno e un solo campo F con $|F| = q$.*

Dimostrazione. Abbiamo già provato la prima affermazione. Per la seconda prendiamo il campo $P = \mathbb{Z}/(p)$ di p elementi e sia F un campo di spezzamento di $x^q - x$ su P. Affermiamo che $|F| = q$ e che F coincide con l'insieme R delle radici di $x^q - x$ in F. Osserviamo anzitutto che $(x^q - x)' = -1$, quindi $x^q - x$ ha q radici distinte in F. Mostriamo che $R = \{u_1, \dots, u_q\}$ è un sottocampo di F. Per la comoda formula binomiale $(a + b)^p = a^p + b^p$ in caratteristica p, per ogni i, j si ha $(u_i \pm u_j)^q = u_i^q \pm u_j^q = u_i \pm u_j$. Dunque $u_i \pm u_j \in R$. Inoltre $1 \in R$, $(u_iu_j)^q = u_i^q u_j^q = u_iu_j$ e, se $u_i \neq 0$, $(u_i^{-1})^q = (u_i^q)^{-1} = u_i^{-1}$. Queste proprietà mostrano che R è un sottocampo. Di conseguenza contiene il campo primo P e $R = P(R) = F$.

Siano ora F, F' due campi con $|F| = q = |F'|$. Sono entrambi estensioni di $P = \mathbb{Z}/(p)$. L'insieme F^* degli elementi non nulli di F è un gruppo moltiplicativo di ordine $q - 1$. Per $0 \neq u \in F$ si ha dunque $u^{q-1} = 1$ e $u^q = u$. Poiché l'ultima relazione vale anche per $u = 0$, ogni elemento di F è radice di $x^q - x$. Quest'equazione non ha più di q radici distinte in alcun campo, quindi F è un campo di spezzamento di $x^q - x$ su P. Lo stesso vale per F'. Per il teorema di isomorfismo dei campi di spezzamento (teorema 4.4), F e F' sono isomorfi. □

Consideriamo ora la teoria relativa dei campi finiti, ossia studiamo un campo finito rispetto a un suo sottocampo. Sia $|F| = q = p^m$ e sia E un'estensione di F con $[E : F] = n$. Come si è visto, $|E| = q^n$. Abbiamo anche visto (§ 4.4) che $a \mapsto a^p$ è un automorfismo di E, dunque anche $\eta : a \mapsto a^q$ lo è. Inoltre $|F| = q$, quindi $b^q = b$ per $b \in F$. Ne segue $\eta \in \mathrm{Gal}(E/F)$. Abbiamo così il seguente

Teorema 4.26. *Sia F un campo finito con $q = p^m$ elementi e sia E un'estensione di F tale che $[E : F] = n$. Allora E è ciclica su F, con gruppo di Galois $\langle\eta\rangle$, dove $\eta : a \mapsto a^q$.*

Dimostrazione. Mostriamo anzitutto che l'ordine di η è n. Poiché $|E| = q^n$, si ha $a^{q^n} = a$ per ogni $a \in E$. Pertanto $\eta^n = 1$. Se fosse $\eta^{n'} = 1$ per $0 < n' < n$, varrebbe $a^{q^{n'}} = a$ per ogni $a \in E$, in contrasto con il fatto che il polinomio $x^{q^{n'}} - x$ ha al più $q^{n'}$ radici in E. Quindi $o(\eta) = n$ e $|\langle\eta\rangle| = n$. Sia $F' = \mathrm{Inv}\langle\eta\rangle$. Per il teorema fondamentale della teoria di Galois, $[E : F'] = n$ e $\mathrm{Gal}(E/F') = \langle\eta\rangle$. D'altra parte, poiché $\eta \in \mathrm{Gal}(E/F)$, si ha $F \subseteq F' = \mathrm{Inv}\langle\eta\rangle$. Da $n = [E : F] = [E : F'][F' : F] = n[F' : F]$ segue $F' = F$. Quindi E è galoisiana su F e $\mathrm{Gal}(E/F) = \langle\eta\rangle$. □

Se K è un campo intermedio di E/F, allora $m = [K : F]$ divide $n = [E : F]$. Viceversa, sia m un divisore qualsiasi di n. Il gruppo ciclico $\mathrm{Gal}(E/F)$ ha uno e un solo sottogruppo di ordine n/m. Per il teorema fondamentale della teoria di Galois, esiste dunque uno e un solo campo intermedio K di E/F tale che $[K : F] = m$. Inoltre $|K| = q^m$. È ora chiaro il seguente

Corollario 1. *Siano F, E come nel teorema 4.26. Se K è un campo intermedio di E/F,*

$|K| = q^m$ *con* $m \mid n$. *Viceversa, se* $m \mid n$, *esiste uno e un solo campo intermedio* K *di* E/F *con* $|K| = q^m$.

Applicando questo risultato si ottiene una formula dovuta a Gauss per il numero dei polinomi monici irriducibili di grado n a coefficienti in F.

Corollario 2. *Sia* F *un campo finito con* $|F| = q$, *e sia* $N(n, q)$ *il numero dei polinomi monici irriducibili in* $F[x]$ *di grado* n. *Allora*

$$N(n,q) = \frac{1}{n} \sum_{d|n} \mu\left(\frac{n}{d}\right) q^d, \tag{69}$$

dove μ *è la funzione di Möbius (definita nell'esercizio 17, p. 151 dell'originale).*

Dimostrazione. Basta mostrare che

$$x^{q^n} - x = \prod g(x), \tag{70}$$

dove il prodotto si estende a tutti i polinomi monici irriducibili il cui grado divide n. Poiché $x^{q^n} - x$ non ha radici multiple, e quindi neppure fattori multipli in $F[x]$, basta verificare che un polinomio monico irriducibile $g(x)$ è un fattore di $x^{q^n} - x$ se e solo se il suo grado m divide n. Sia $g(x)$ un fattore monico irriducibile di $x^{q^n} - x$ in $F[x]$, con $\deg g = m$, e sia E un'estensione di F con $[E : F] = n$. Allora E è un campo di spezzamento di $x^{q^n} - x$ su F e contiene quindi una radice r di $g(x)$. Il polinomio minimo di r su F è $g(x)$, dunque $F(r)$ è un campo intermedio di E/F con $[F(r) : F] = m$. Ne segue $m \mid n$. Viceversa, sia $g(x)$ un polinomio monico irriducibile di $F[x]$ di grado $m \mid n$. Allora $K' = F[x]/(g(x))$ è un'estensione di F con $|K'| = q^m$. Poiché $m \mid n$, K' è isomorfo a un campo intermedio K di E/F. Dunque E contiene un elemento r con polinomio minimo $g(x)$ su F. Poiché $r^{q^m} = r$ e $m \mid n$, anche $r^{q^n} = r$, quindi $g(x) \mid (x^{q^n} - x)$. Ciò prova la fattorizzazione (70). Confrontando i gradi dei due membri, troviamo

$$q^n = \sum_{d|n} N(d,q)d. \tag{71}$$

La formula d'inversione di Möbius (p. 151 dell'originale) dà ora la formula di Gauss (69). □

Nel teorema 2.18 (p. 132 dell'originale) abbiamo dimostrato che ogni sottogruppo finito del gruppo moltiplicativo di un campo è ciclico. In particolare, per un campo finito E, il gruppo moltiplicativo E^* è ciclico. Se F è un sottocampo e z un generatore di E^*, allora evidentemente $E = F(z)$. Abbiamo quindi il seguente

Teorema 4.27. *Se* E *e* F *sono come nel teorema 4.26,* E *ammette un elemento primitivo su* F.

4.14 Basi speciali per le estensioni di grado finito

Sia E un'estensione di grado finito del campo F. Consideriamo anzitutto il problema dell'esistenza di un elemento primitivo per E. A Steinitz è dovuta una caratterizzazione particolarmente elegante delle estensioni che ammettono un elemento primitivo.

Teorema 4.28. *Sia* E *un'estensione di grado finito di* F. *Allora* E/F *ammette un elemento primitivo se e solo se esistono soltanto un numero finito di campi intermedi fra* F *ed* E.

Dimostrazione. Supponiamo anzitutto $E = F(u)$ e sia K un sottocampo contenente F. Siano $f(x)$ il polinomio minimo di u su F e $g(x)$ il suo polinomio minimo su K. Allora $g(x) \mid f(x)$. Sia K' il campo intermedio di E/F generato dai coefficienti di $g(x)$. Si ha $K' \subseteq K$ ed è chiaro che anche il polinomio minimo di u su K' è $g(x)$. Poiché $E = K(u) = K'(u)$, segue

$[E : K] = \deg g(x) = [E : K']$, da cui $K = K'$. I campi intermedi fra F ed E sono dunque precisamente quelli generati su F dai coefficienti di fattori monici di $f(x)$ in $E[x]$. Essendovi soltanto un numero finito di tali fattori, anche i campi intermedi sono in numero finito. Viceversa, supponiamo che E/F abbia solo un numero finito di campi intermedi. Per F finito si è visto nel paragrafo precedente che E ammette un elemento primitivo su F; possiamo così supporre F infinito. L'esistenza di un elemento primitivo seguirà per induzione sul numero dei generatori di un sistema finito di generatori se dimostriamo che, per due elementi qualsiasi u, v di E, il campo $F(u, v)$ ammette un elemento primitivo. Consideriamo i campi $F(u+av)$, $a \in F$. Sono in numero finito, mentre gli a sono infiniti. Esistono dunque $a \neq b$ tali che $F(u+av) = F(u+bv)$. Allora $v = (a-b)^{-1}(u+av-u-bv) \in F(u+av)$ e $u = (u+av) - av \in F(u+av)$. Quindi $z = u + av$ è un elemento primitivo di $F(u, v)$. □

Se E è separabile di grado finito su F, la chiusura normale K di E/F è galoisiana su F (p. 254 dell'originale). Inoltre vi è solo un numero finito di campi intermedi di K/F, poiché corrispondono ai sottogruppi di $\mathrm{Gal}(K/F)$. A maggior ragione E/F ha solo un numero finito di campi intermedi. Il criterio di Steinitz dà pertanto il seguente

Corollario. *Ogni estensione separabile di grado finito ammette un elemento primitivo.*

Sia ora K un'estensione galoisiana di F con gruppo di Galois $G = \{\eta_i \mid 1 \leq i \leq n\}$, $n = [K : F]$. Se $z \in K$ e $\{z_1, z_2, \ldots, z_r\}$ è l'orbita Gz, il polinomio minimo di z su F è $\prod_{i=1}^{r}(x - z_i)$. Dunque z è primitivo se e solo se l'orbita di z contiene n elementi, ossia se $\eta_1(z), \ldots, \eta_n(z)$ sono distinti. Una condizione più forte è che i coniugati $\eta_i(z)$ siano linearmente indipendenti su F. In tal caso $(\eta_1(z), \eta_2(z), \ldots, \eta_n(z))$ è una base di K/F, detta *base normale*. Nel resto di questo paragrafo ne dimostreremo l'esistenza per ogni estensione galoisiana. La dimostrazione si basa su importanti proprietà di indipendenza degli automorfismi dei campi.

Cominciamo con un risultato classico sull'indipendenza lineare dei caratteri. Siano H un monoide e F un campo. Per *carattere* χ di H in F (o *F-carattere* di H) si intende un omomorfismo di H nel gruppo moltiplicativo F^* degli elementi non nulli di F. Dunque $\chi : H \to F^*$ soddisfa $\chi(1) = 1$ e $\chi(h_1h_2) = \chi(h_1)\chi(h_2)$ per ogni $h_i \in H$. Dimostreremo una proprietà importante dei caratteri.

Teorema di indipendenza di Dedekind. *I caratteri distinti di un monoide a valori in un campo sono linearmente indipendenti. Ossia, se $\chi_1, \ldots, \chi_n$ sono caratteri distinti di H nel campo F, gli unici elementi $a_1, \ldots, a_n$ di F per i quali*

$$a_1\chi_1(h) + a_2\chi_2(h) + \cdots + a_n\chi_n(h) = 0 \tag{72}$$

per ogni $h \in H$ sono $a_1 = \cdots = a_n = 0$.

Dimostrazione. Procediamo per induzione su n. Per $n = 1$ il risultato è chiaro, poiché $a\chi(h) = 0$ per $a \neq 0$ darebbe $\chi(h) = 0$ per ogni h, il che è impossibile, essendo $\chi(1) = 1$. Sia $n > 1$ e supponiamo il teorema vero per $n - 1$ caratteri. Supponiamo che valga (72) per ogni $h \in H$, con $a_i \in F$. Per ipotesi induttiva possiamo supporre tutti gli a_i non nulli. Poiché $\chi_1 \neq \chi_2$, esiste $a \in H$ con $\chi_1(a) \neq \chi_2(a)$. Sostituendo ah ad h in (72) e usando la proprietà dei caratteri, si ottiene

$$a_1\chi_1(a)\chi_1(h) + a_2\chi_2(a)\chi_2(h) + \cdots + a_n\chi_n(a)\chi_n(h) = 0.$$

Moltiplicando invece la (72) per $\chi_1(a)$ si ha

$$a_1\chi_1(a)\chi_1(h) + a_2\chi_1(a)\chi_2(h) + \cdots + a_n\chi_1(a)\chi_n(h) = 0.$$

Sottraendo, si trova $a_2'\chi_2(h) + \cdots + a_n'\chi_n(h) = 0$ con $a_i' = a_i(\chi_i(a) - \chi_1(a))$, $2 \leq i \leq n$. Poiché $a_2' = a_2(\chi_2(a) - \chi_1(a)) \neq 0$, ciò contraddice il teorema per $n - 1$ caratteri e conclude la dimostrazione. □

Una delle principali applicazioni riguarda i monomorfismi di un campo in un altro.

Corollario. *Siano F_1, F_2 campi e $\eta_1, \dots, \eta_n$ monomorfismi distinti di F_1 in F_2. Essi sono linearmente indipendenti su F_2.*

Dimostrazione. Le restrizioni degli η_i agli elementi non nulli di F_1 sono caratteri del gruppo moltiplicativo $H = F_1^*$ in F_2. Il risultato segue dunque dal teorema di Dedekind. □

Supponiamo ora che E/F sia separabile di grado finito $n = [E : F]$ e che K/F ne sia la chiusura normale. Abbiamo il seguente criterio, espresso mediante un determinante, per riconoscere una base di E/F.

Teorema 4.29. *Sia E/F separabile di grado finito e sia K/F la sua chiusura normale. Il numero dei monomorfismi di E/F in K/F è $n = [E : F]$. Se questi sono $\eta_1 = 1, \eta_2, \dots, \eta_n$, una successione $(u_1, u_2, \dots, u_n)$ di n elementi di E è una base di E/F se e solo se*

$$\begin{vmatrix} \eta_1(u_1) & \eta_1(u_2) & \cdots & \eta_1(u_n) \\ \eta_2(u_1) & \eta_2(u_2) & \cdots & \eta_2(u_n) \\ \vdots & \vdots & & \vdots \\ \eta_n(u_1) & \eta_n(u_2) & \cdots & \eta_n(u_n) \end{vmatrix} \neq 0. \tag{73}$$

Dimostrazione. Sia $G = \mathrm{Gal}(K/F)$ e sia H il sottogruppo di G che fissa E. Allora $n = [E : F] = [G : H]$ e si può scrivere $G = \zeta_1 H \cup \cdots \cup \zeta_n H$ con classi distinte e $\zeta_1 = 1$. Poniamo $\eta_i = \zeta_i|_E$. Si tratta di un monomorfismo di E/F in K/F e per $i \neq j$ si ha $\eta_i \neq \eta_j$. Infatti $\eta_i = \eta_j$ implicherebbe $\zeta_i^{-1}(\zeta_j(u)) = u$ per ogni $u \in E$, dunque $\zeta_i^{-1}\zeta_j \in H$ e $\zeta_i H = \zeta_j H$, contro l'ipotesi. Sia ora η un qualsiasi monomorfismo di E/F in K/F. Il campo K è campo di spezzamento su F di un polinomio $f(x) \in F[x]$, dunque lo è anche su E e su $\eta(E)$. Per il teorema 4.4 l'isomorfismo η di E su $\eta(E)$ si può estendere a un automorfismo ζ di K/F. Allora $\zeta \in \mathrm{Gal}(K/F)$ e quindi $\zeta = \zeta_i\lambda$ per qualche $\lambda \in H$. Pertanto $\eta = \zeta|_E = \zeta_i|_E = \eta_i$. I monomorfismi di E/F in K/F sono dunque precisamente $\eta_1 = 1, \eta_2, \dots, \eta_n$.

Se gli elementi $u_1, \dots, u_n$ di E sono linearmente dipendenti su F, esistono $a_i \in F$ non tutti nulli tali che $\sum_i a_i u_i = 0$. Applicando η_j, otteniamo

$$\sum_i a_i \eta_j(u_i) = 0, \qquad 1 \leq j \leq n.$$

Il sistema lineare omogeneo $\sum_i \eta_j(u_i)x_i = 0$, $1 \leq j \leq n$, ha dunque la soluzione non nulla $(x_1, \dots, x_n) = (a_1, \dots, a_n)$. Ne segue $\det(\eta_j(u_i)) = 0$. Viceversa, supponiamo $\det(\eta_j(u_i)) = 0$. Considerando la matrice trasposta, otteniamo una soluzione non nulla $(a_1, \dots, a_n)$, $a_j \in K$, del sistema $\sum_j \eta_j(u_i)x_j = 0$, $1 \leq i \leq n$. Quindi $\sum_j a_j\eta_j(u_i) = 0$. Allora $(u_1, \dots, u_n)$ non è una base: altrimenti ogni $u \in E$ si scriverebbe $\sum_i c_i u_i$ con $c_i \in F$, e si avrebbe

$$\sum_j a_j \eta_j(u) = \sum_{i,j} a_j c_i \eta_j(u_i) = \sum_i c_i \Big(\sum_j a_j \eta_j(u_i) \Big) = 0.$$

Ciò contraddice l'indipendenza lineare dei monomorfismi $\eta_1, \dots, \eta_n$ di E in K (corollario al teorema di Dedekind). □

Dimostriamo ora l'esistenza di una base normale per ogni estensione galoisiana K/F. Distinguiamo due casi: (I) K/F è ciclica; (II) F è infinito. Se F è finito, anche K è finito e, come abbiamo visto nel § 4.13, K è ciclico su F. I due casi esauriscono quindi le possibilità.

Sia anzitutto K/F ciclica, ossia $G = \mathrm{Gal}(K/F) = \langle \eta \rangle$, con η di ordine $n = [K : F]$. Allora $G = \{1, \eta, \dots, \eta^{n-1}\}$. Osserviamo che η è una trasformazione lineare di K su F: si ha $\eta(u + v) = \eta(u) + \eta(v)$, e per $a \in F$ si ha $\eta(au) = a\eta(u)$. Possiamo applicare a η la teoria di una

singola trasformazione lineare di uno spazio vettoriale di dimensione finita (§ 3.10, pp. 195–197 dell'originale). Ricordiamo che il modulo su $F[x]$ ($F[\lambda]$ nella notazione del capitolo 3) determinato da una trasformazione lineare è somma diretta di moduli ciclici con annullatori $(d_i(x))$, dove $d_1(x), \dots, d_r(x)$ sono i fattori invarianti. L'ultimo è il polinomio minimo, mentre il prodotto di tutti i fattori invarianti è il polinomio caratteristico della trasformazione. Il modulo è quindi ciclico se e solo se i polinomi caratteristico e minimo coincidono o, equivalentemente, se il grado del polinomio minimo è la dimensione dello spazio. Affermiamo che ciò avviene per la trasformazione η di K su F. Anzitutto $\eta^n = 1$, quindi η soddisfa $x^n - 1 = 0$. D'altra parte, se $f(x) = x^m + a_1 x^{m-1} + \cdots + a_m$, con $a_i \in F$ e $m < n$, allora $f(\eta) \neq 0$, poiché gli automorfismi $1, \eta, \dots, \eta^m$ sono distinti e quindi linearmente indipendenti su K, e a maggior ragione su F. Il fatto che K sia un $F[x]$-modulo ciclico (mediante la trasformazione lineare η) significa che K/F ha una base della forma $(u, \eta(u), \eta^2(u), \dots, \eta^{n-1}(u))$. Questa è una base normale di K/F.

Supponiamo ora F infinito. Seguendo Artin, dimostreremo l'esistenza di una base normale servendoci della nozione di indipendenza algebrica dei monomorfismi di campi. Se E, K sono campi e $\eta_1, \dots, \eta_n$ sono monomorfismi di E in K, li diciamo *algebricamente indipendenti su K* se l'unico polinomio $f(x_1, \dots, x_n) \in K[x_1, \dots, x_n]$, con le x_i indeterminate, tale che $f(\eta_1(u), \dots, \eta_n(u)) = 0$ per ogni $u \in E$ è $f = 0$.

Teorema 4.30. *Sia F un campo infinito, E un'estensione separabile di grado finito di F e K una chiusura normale di E/F. Siano $\eta_1, \dots, \eta_n$ gli $n = [E : F]$ monomorfismi distinti di E/F in K/F (teorema 4.29). Allora gli η_i sono algebricamente indipendenti su K.*

Dimostrazione. Sia $f(x_1, \dots, x_n) \in K[x_1, \dots, x_n]$ tale che $f(\eta_1(u), \dots, \eta_n(u)) = 0$ per ogni $u \in E$. Sia $(u_1, \dots, u_n)$ una base di E/F. Per ogni scelta di $a_i \in F$ si ha

$$0 = f\Big(\eta_1\Big(\sum_i a_i u_i\Big), \dots, \eta_n\Big(\sum_i a_i u_i\Big)\Big) = f\Big(\sum_i a_i \eta_1(u_i), \dots, \sum_i a_i \eta_n(u_i)\Big).$$

Se poniamo $g(x_1, \dots, x_n) = f(\sum_i \eta_1(u_i)x_i, \dots, \sum_i \eta_n(u_i)x_i)$, allora $g(a_1, \dots, a_n) = 0$ per tutti gli $a_i \in F$. Sia $(v_1, \dots, v_m)$ una base di K/F. Possiamo scrivere $g = \sum_j g_j v_j$ con $g_j \in F[x_1, \dots, x_n]$. La condizione $g(a_1, \dots, a_n) = 0$ dà $g_j(a_1, \dots, a_n) = 0$ per ogni j. Poiché vale per tutti gli $a_i \in F$, dal teorema 2.19 (p. 136 dell'originale) segue $g_j = 0$ per ogni j, e quindi $g = 0$. Per il teorema 4.29 si ha $\det(\eta_j(u_i)) \neq 0$. La matrice $A = (\eta_j(u_i))$ (riga j, colonna i) ha perciò un'inversa $A^{-1} = (v_{ij}) \in M_n(K)$. Scrivendo x come vettore colonna, la relazione $g(x) = f(Ax)$ dà $f(x) = g(A^{-1}x)$. Da $g = 0$ segue $f = 0$, come richiesto. □

Possiamo ora completare la dimostrazione del seguente teorema.

Teorema della base normale. *Ogni estensione galoisiana K/F di grado finito ammette una base normale.*

Dimostrazione. Avendo dimostrato il risultato quando $G = \mathrm{Gal}(K/F)$ è ciclico, e quindi per F finito, possiamo supporre F infinito. Per il risultato appena provato gli automorfismi $\eta_1, \dots, \eta_n$ di K/F sono algebricamente indipendenti su K. Abbiamo anche visto che, per $u \in K$, la successione $(\eta_1(u), \dots, \eta_n(u))$ è una base se e solo se

$$\det\big((\eta_i\eta_j)(u)\big) \neq 0.$$

Scriviamo $\eta_i\eta_j = \eta_{i(j)}$. Allora $j \mapsto i(j)$ è una permutazione di $1, \dots, n$. Nell'anello $K[x_1, \dots, x_n]$ consideriamo la matrice X il cui elemento (i, j) è $x_{i(j)}$. Affermiamo che $\det X \neq 0$. Specializziamo $x_1 = 1$ e $x_i = 0$ per $i > 1$. Poiché $j \mapsto i(j)$ è una permutazione di $1, \dots, n$ e, al variare di i, i prodotti $\eta_i\eta_j$ sono a loro volta distinti, x_1 compare una e una sola volta in ogni riga e in ogni colonna di X. Dopo la specializzazione il determinante vale quindi ± 1. Dunque $d(x_1, \dots, x_n) = \det X \neq 0$. Per l'indipendenza algebrica degli η_i su K esiste $u \in K$ con $\det((\eta_i\eta_j)(u)) \neq 0$. Allora $(\eta_1(u), \dots, \eta_n(u))$ è una base normale. □

Esercizi

1. Trovare un elemento primitivo di $\mathbb{Q}(\sqrt{2},\sqrt{3})$ su $\mathbb{Q}$.
2. Trovare un elemento primitivo di un campo di spezzamento di x^5-2 su $\mathbb{Q}$.
3. Siano x,y indeterminate, $E=(\mathbb{Z}/(p))(x,y)$ e $F=(\mathbb{Z}/(p))(x^p,y^p)$. Mostrare che $[E:F]=p^2$ e che E non ammette un elemento primitivo su F. Esibire un'infinità di campi intermedi distinti di E/F.
4. Determinare una base normale per il campo dell'esercizio 1.
5. Dimostrare che, se $E=F(u,v)$ con u,v algebrici su F e u separabile su F, allora E ammette un elemento primitivo su F.
6. Siano E,K estensioni dello stesso campo F e sia $\mathrm{Hom}_F(E,K)$ l'insieme delle applicazioni lineari dallo spazio vettoriale E/F in K/F. Osservare che $\mathrm{Hom}_F(E,K)$ diviene uno spazio vettoriale su K definendo come al solito la somma e ponendo, per $k\in K$, $l\in\mathrm{Hom}_F(E,K)$, $(kl)(x)=k\,l(x)$, $x\in E$. Sia $[E:F]<\infty$ e sia $(\xi_1,\dots,\xi_n)$ una base di E/F. Sia l_i l'elemento di $\mathrm{Hom}_F(E,K)$ tale che $l_i(\xi_j)=0$ per $j\neq i$ e $l_i(\xi_i)=1$. Mostrare che $(l_1,\dots,l_n)$ è una base di $\mathrm{Hom}_F(E,K)$ su K. Ne segue che la dimensione di $\mathrm{Hom}_F(E,K)$ su K è $[E:F]$.
7. Con le notazioni e le ipotesi dell'esercizio 6, usare tale esercizio e il teorema di indipendenza di Dedekind per mostrare che vi sono al più $[E:F]$ monomorfismi di E/F in K/F.

4.15 Tracce e norme

Sia E/F galoisiana, con $G=\mathrm{Gal}(E/F)=\{\eta_1=1,\eta_2,\dots,\eta_n\}$. Per $u\in E$ definiamo

$$T_{E/F}(u)=\sum_{i=1}^{n}\eta_i(u),\qquad N_{E/F}(u)=\prod_{i=1}^{n}\eta_i(u), \tag{74}$$

dette rispettivamente *traccia* e *norma* di u in E/F. Entrambe sono evidentemente fisse sotto il gruppo di Galois e quindi appartengono a F. Abbiamo così le applicazioni traccia e norma

$$T_{E/F}:u\mapsto T_{E/F}(u),\qquad N_{E/F}:u\mapsto N_{E/F}(u)$$

di E in F. Per $u,v\in E$ e $a\in F$ si ha

$$\begin{aligned}
T_{E/F}(u+v)&=\sum_i\eta_i(u+v)=\sum_i\eta_i(u)+\sum_i\eta_i(v)=T_{E/F}(u)+T_{E/F}(v),\\
T_{E/F}(au)&=\sum_i\eta_i(au)=a\sum_i\eta_i(u)=aT_{E/F}(u),\\
N_{E/F}(uv)&=\prod_i\eta_i(uv)=\prod_i\eta_i(u)\prod_i\eta_i(v)=N_{E/F}(u)N_{E/F}(v),\\
N_{E/F}(au)&=\prod_i\eta_i(au)=a^n\prod_i\eta_i(u)=a^nN_{E/F}(u).
\end{aligned}$$

Le prime due identità mostrano che $T=T_{E/F}$ è F-lineare, cioè è una forma lineare sullo spazio vettoriale E su F. Per $N=N_{E/F}$ abbiamo osservato che è moltiplicativa e omogenea di grado n. Evidentemente $T(0)=0$, $T(-u)=-T(u)$, $N(0)=0$, $N(1)=1$ e, per $u\neq0$, $N(u)\neq0$ e $N(u^{-1})=N(u)^{-1}$. La restrizione di N al gruppo moltiplicativo E^* degli elementi non nulli di E è quindi un omomorfismo di E^* in F^*.

Consideriamo per esempio l'estensione quadratica $E = \mathbb{Q}(\sqrt{m})$, dove m è un intero privo di fattori quadrati. Ogni $u \in E$ è della forma $a + b\sqrt{m}$, con $a, b \in \mathbb{Q}$, e il gruppo di Galois consiste dell'identità e dell'applicazione $a + b\sqrt{m} \mapsto a - b\sqrt{m}$. Dunque

$$T(a + b\sqrt{m}) = 2a, \qquad N(a + b\sqrt{m}) = a^2 - b^2 m.$$

L'esempio forse più familiare si ottiene considerando $\mathbb{C}$ come estensione quadratica di $\mathbb{R}$ mediante $\sqrt{-1}$. Per $u = a + b\sqrt{-1}$, $a, b \in \mathbb{R}$, $T(u) = 2a$ è il doppio della parte reale di u, mentre $N(u) = a^2 + b^2 = |u|^2$.

Essendo T e N omomorfismi, è naturale chiedersi quali siano le immagini $T(E)$ e $N(E^*)$ e i nuclei di T e di N (quest'ultima come omomorfismo di E^* in F^*). La prima immagine si determina facilmente. Si ha $T(E) \subseteq F$ e, poiché $T(E)$ è un sottospazio dello spazio unidimensionale F/F, o $T(E) = 0$ oppure $T(E) = F$. Il primo caso è escluso: significherebbe $\sum_i \eta_i(u) = 0$ per ogni $u \in E$, contro l'indipendenza lineare degli automorfismi η_i. Di solito è meno agevole ottenere informazioni su $N(E^*)$. Per esempio, se $E = \mathbb{Q}(\sqrt{m})$, il problema generale è determinare i numeri razionali c per cui $x^2 - my^2 = c$ ha una soluzione razionale $(x, y) = (a, b)$. È un problema aritmetico non banale. Per $m = -1$ l'aritmetica dell'anello degli interi di Gauss fornisce una risposta, indicata nell'esercizio 4 alla fine di questo paragrafo.

Ricaviamo ora due teoremi generali sui nuclei della traccia e della norma. Il primo è universalmente noto come «Satz 90 di Hilbert», poiché Hilbert lo pubblicò nel 1897 nella sua classica relazione sulla teoria algebrica dei numeri, dove figurava come novantesimo teorema.[15] È il risultato seguente.

Teorema 4.31. *Sia E un'estensione ciclica del campo F e sia η un generatore del gruppo (ciclico) di Galois di E/F. Per $u \in E$ vale $N_{E/F}(u) = 1$ se e solo se esiste $v \in E^*$ tale che $u = v(\eta(v))^{-1}$ (talvolta si scrive $v^{1-\eta}$).*

In una direzione il risultato è immediato: se $u = v(\eta(v))^{-1}$, allora

$$N(u) = N(v)N(\eta(v)^{-1}) = N(v)N(v)^{-1} = 1.$$

Per il viceversa dimostreremo un risultato più generale sulle estensioni galoisiane, dovuto ad A. Speiser (per le matrici).

Teorema 4.32. *Sia E un'estensione galoisiana di grado finito di F con gruppo di Galois G. Sia $\eta \mapsto u_\eta$ un'applicazione di G nel gruppo moltiplicativo E^* tale che*

$$u_{\zeta\eta} = \zeta(u_\eta)u_\zeta \tag{75}$$

per ogni $\eta, \zeta \in G$. Allora esiste $v \neq 0$ in E tale che

$$u_\eta = v(\eta(v))^{-1}. \tag{76}$$

Dimostrazione. Poiché gli u_η sono non nulli e gli automorfismi $\eta \in G$ sono linearmente indipendenti su E, esiste $w \in E$ tale che

$$v = \sum_{\eta \in G} u_\eta \eta(w) \neq 0.$$

Allora, per $\zeta \in G$, si ha

$$\begin{aligned}\zeta(v) &= \sum_\eta \zeta(u_\eta)(\zeta\eta)(w) = \sum_\eta u_{\zeta\eta}u_\zeta^{-1}(\zeta\eta)(w) \\ &= \Big(\sum_\eta u_{\zeta\eta}(\zeta\eta)(w)\Big)u_\zeta^{-1} = \Big(\sum_\eta u_\eta\eta(w)\Big)u_\zeta^{-1} = vu_\zeta^{-1}.\end{aligned}$$

[15] D. Hilbert, *Theorie der algebraischen Zahlkörper*, 1897, p. 149.

Quindi $u_\zeta = v\zeta(v)^{-1}$, come richiesto. □

Per completare la dimostrazione del Satz 90, supponiamo G ciclico con generatore η e sia $u \in E$ tale che $N(u) = 1$. Definiamo

$$u_{\eta^i} = u\eta(u)\eta^2(u)\cdots\eta^{i-1}(u), \qquad 1 \le i \le n. \tag{77}$$

Per $i + j \le n$ si ha $u_{\eta^j}\eta^j(u_{\eta^i}) = u\eta(u)\cdots\eta^{j-1}(u)\eta^j(u)\cdots\eta^{i+j-1}(u) = u_{\eta^{i+j}}$. La stessa relazione vale per $i + j > n$, poiché $u_1 = u_{\eta^n} = N(u) = 1$. Le (75) sono dunque soddisfatte per $G = \langle\eta\rangle$. Esiste perciò v tale che $u = u_\eta = v(\eta(v))^{-1}$.

I due risultati che abbiamo dimostrato hanno analoghi additivi. Cominciamo con la forma additiva del teorema 4.32.

Teorema 4.33. *Siano E, F, G come nel teorema 4.32 e sia $\eta \mapsto d_\eta$ un'applicazione di G in E tale che*

$$d_{\zeta\eta} = d_\zeta + \zeta(d_\eta) \tag{78}$$

per ogni $\eta, \zeta \in G$. Allora esiste $c \in E$ tale che

$$d_\eta = c - \eta(c), \qquad \eta \in G. \tag{79}$$

Dimostrazione. Abbiamo visto che esiste $u \in E$ con $T(u) \neq 0$. Poniamo

$$c = T(u)^{-1}\sum_\eta d_\eta\eta(u).$$

Allora

$$\begin{aligned} c - \zeta(c) &= T(u)^{-1}\sum_\eta [d_\eta\eta(u) - \zeta(d_\eta)(\zeta\eta)(u)] \\ &= T(u)^{-1}\sum_\eta [d_\eta\eta(u) + d_\zeta(\zeta\eta)(u) - d_{\zeta\eta}(\zeta\eta)(u)] \\ &= T(u)^{-1}d_\zeta\sum_\eta(\zeta\eta)(u) = d_\zeta T(u)^{-1}\sum_\eta\eta(u) \\ &= d_\zeta T(u)^{-1}T(u) = d_\zeta, \end{aligned}$$

come richiesto. □

Sia ora $G = \langle\eta\rangle$ e sia $d \in E$ tale che $T(d) = 0$. Poniamo

$$d_{\eta^i} = d + \eta(d) + \cdots + \eta^{i-1}(d), \qquad 1 \le i \le n. \tag{80}$$

Come nel caso delle norme, si vede che vale la condizione (78) e quindi la (79). Il teorema 4.33 dà il seguente analogo additivo del Satz 90.

Teorema 4.34. *Sia E/F ciclica con gruppo di Galois $G = \langle\eta\rangle$. Sia $d \in E$ di traccia zero. Allora esiste $c \in E$ tale che $d = c - \eta(c)$.*

Applichiamo il Satz 90 e il suo analogo additivo per ottenere risultati sulla struttura delle estensioni cicliche. Diamo anzitutto una dimostrazione più semplice e un'estensione di un risultato precedente (lemma 3 del § 4.7, p. 253 dell'originale).

Teorema 4.35. *Supponiamo che F contenga n radici n-esime distinte di 1 e che E/F sia ciclica di grado n. Allora $E = F(u)$ con $u^n \in F$.*

Dimostrazione. Sia z una radice primitiva n-esima di 1. Si ha $N_{E/F}(z) = z^n = 1$. Esiste quindi $u \in E$ tale che $z = u(\eta(u))^{-1}$, dove η genera il gruppo di Galois. Allora $\eta(u) = z^{-1}u$ e

$\eta(u^n) = \eta(u)^n = (z^{-1}u)^n = u^n$. Pertanto $u^n \in F$. Inoltre da $\eta(u) = z^{-1}u$ segue $\eta^i(u) = z^{-i}u$, e quindi l'orbita di u sotto $\mathrm{Gal}(E/F)$ contiene n elementi distinti. Il polinomio minimo di u su F ha dunque grado n e $E = F(u)$. □

Otteniamo ora la struttura di un'estensione ciclica di grado p in caratteristica p.

Teorema 4.36. *Sia F un campo di caratteristica $p \neq 0$ e sia E/F ciclica di grado p. Allora $E = F(c)$ con $c^p - c \in F$.*

Dimostrazione. Si ha $T_{E/F}(1) = 1 + 1 + \cdots + 1$ (p termini) $= 0$. Per il teorema 4.34 esiste dunque $c \in E$ tale che $\eta(c) = c + 1$. Allora $\eta^i(c) = c + i$ e l'orbita di c sotto $\mathrm{Gal}(E/F)$ contiene p elementi. Dunque $E = F(c)$. Inoltre $\eta(c^p - c) = \eta(c)^p - \eta(c) = (c+1)^p - (c+1) = c^p - c$. Quindi $c^p - c \in F$. □

Esercizi

1. Mostrare che, se E è un campo finito e F un sottocampo, cosicché E/F è ciclica, l'omomorfismo norma $N_{E/F}$ da E^* a F^* è suriettivo.

2. (Albert.) Sia E un'estensione ciclica di grado n di F e sia η un generatore di $\mathrm{Gal}(E/F)$. Sia $r \mid n$, $n = rm$, e supponiamo $c \in F^*$ tale che $c^r = N_{E/F}(u)$ per qualche $u \in E$. Mostrare che esiste v nell'unico campo intermedio K di grado m su F tale che $c = N_{K/F}(v)$.

3. Siano E, F, n, u come nel teorema 4.35. Mostrare che $v \in E$ soddisfa un'equazione della forma $x^n = a$ con $a \in F$ se e solo se $v = bu^k$ con $b \in F$ e $1 \leq k \leq n$.

4. Mostrare che un numero razionale $a \neq 0$ è norma di un elemento di $\mathbb{Q}(\sqrt{-1})$ se e solo se i primi dispari che compaiono con molteplicità dispari nel numeratore o nel denominatore di a, scritto in forma ridotta b/c, $(b, c) = 1$, sono della forma $4n + 1$ (cfr. esercizio 10, p. 150 dell'originale).[16]

5. Supponiamo che F contenga p radici p-esime distinte di 1, con p primo, e che E/F sia ciclica di grado p^f. Sia z una radice primitiva p-esima di 1. Se E/F si può immergere in un'estensione ciclica K/F di grado p^{f+1}, mostrare che $z = N_{E/F}(u)$ per qualche $u \in E$.

6. Mostrare che, se $E = \mathbb{Q}(\sqrt{m})$ con $m \in \mathbb{Z}$, $m < 0$, allora E non si può immergere in un'estensione ciclica quartica di $\mathbb{Q}$.

7. (Albert.) Con le notazioni dell'esercizio 5, sia η un generatore di $\mathrm{Gal}(E/F)$. Supponiamo che esista $u \in E$ con $N_{E/F}(u) = z$. Mostrare che esiste $v \in E$ con $\eta(v)v^{-1} = u^p$. Mostrare che v non è una potenza p-esima in E e che, per $K = E(w)$ con $w^p = v$, l'estensione K/F è ciclica di grado p^{f+1}.

8. Osservare che gli esercizi 5 e 7 implicano il teorema seguente. Se F contiene p radici p-esime distinte di 1, con p primo, e E/F è ciclica di grado $p^f > 1$, allora E si può immergere in un'estensione ciclica di grado p^{f+1} su F se e solo se una radice primitiva p-esima z di 1 è una norma in E/F. Usare questo risultato per dimostrare che, se la caratteristica di F è diversa da 2 ed $E = F(\sqrt{c}) \neq F$, allora E/F si può immergere in un'estensione ciclica quartica di F se e solo se c è somma di due quadrati di elementi di F.

9. (Uchida.) Sia F_0 un campo di caratteristica p e $F = F_0(s, t)$, con s, t indeterminate. Mostrare che il gruppo di Galois di $x^p - sx - t$ su F è isomorfo al gruppo delle applicazioni $x \mapsto ax + b$, $a, b \in \mathbb{Z}/(p)$, $a \neq 0$.

[16] Nota alla traduzione. Nell'enunciato a stampa manca la condizione $a > 0$, necessaria poiché la norma è una somma di due quadrati.

10. (Uchida.) Con le notazioni dell'esercizio 9, mostrare che il gruppo di Galois di $x^{p+1} - sx - t$ su F è isomorfo al gruppo delle trasformazioni lineari fratte $x \mapsto (ax + b)/(cx + d)$, per $a, b, c, d \in \mathbb{Z}/(p)$, $ad - bc \neq 0$.

11. Sia $W = F(\zeta)$, dove ζ è una radice primitiva p-esima di 1, p primo. Mostrare che W/F è ciclica e $[W : F] = s$ divide $p - 1$. Mostrare che $\mathrm{Gal}(W/F) = \langle\tau\rangle$, dove $\tau(\zeta) = \zeta^t$ e s è l'ordine di $t + (p)$ nel gruppo moltiplicativo di $\mathbb{Z}/(p)$.

12. Con le notazioni dell'esercizio 11, sia $a \in W$ tale che a non sia una potenza p-esima in W ma $(\tau a)a^{-t} = b^p$, $b \in W$. Posto $K = W[x]/(x^p - a)$, mostrare che K/F è ciclica con $[K : F] = ps$ e che contiene un'unica estensione ciclica intermedia E/F con $[E : F] = p$.

13. Siano s, t, p come nell'esercizio 11. Poiché $(s, p) = 1 = (t, p)$, esistono interi s', t' tali che $ss' \equiv 1 \pmod{p}$ e $tt' \equiv 1 \pmod{p}$. Poniamo $t_k = s'(t')^k$, ossia $t_0 = s'$ e $t_k = t_{k-1}t'$ per $1 \leq k \leq s$. Mostrare che

$$\sum_{k=1}^{s} t^k t_k \equiv 1 \pmod{p}, \qquad t_s \equiv s' \pmod{p}.$$

14. Con le notazioni degli esercizi 11 e 13, sia $a \in W$ non nullo e si ponga $M(a) = \prod_{k=1}^{s} (\tau^k a)^{t_k}$. Mostrare che $\tau(M(a))M(a)^{-t}$ è una potenza p-esima in W. Se $M(a)$ non è una potenza p-esima, può quindi servire come elemento a dell'esercizio 12 per costruire un'estensione ciclica di grado p su F.

Nota. Gli esercizi 12–14 espongono risultati dovuti ad Albert, il quale mostrò anche che ogni estensione ciclica E/F con $[E : F] = p$ si può ottenere in questo modo.

4.16 Riduzione modulo p

Determinare il gruppo di Galois di un'equazione a coefficienti razionali è in generale un problema difficile. Se il grado non supera 4, i risultati delle pp. 256–261 dell'originale forniscono un metodo efficace. Anche per un grado arbitrario n abbiamo ottenuto informazioni pertinenti. Per esempio, abbiamo visto che il gruppo di Galois G_f di un polinomio f di grado n è un sottogruppo di A_n se e solo se il discriminante $d(f)$ è un quadrato (p. 257 dell'originale). Abbiamo pure mostrato che G_f è transitivo sull'insieme delle radici se e solo se f è irriducibile. Ricaveremo ora un risultato importante, detto *riduzione modulo p*: applicato per diversi primi p, esso dà informazioni su G_f che, insieme ai risultati ricordati, spesso permettono di determinare G_f.

È facile vedere che possiamo limitarci ai polinomi monici $f \in \mathbb{Z}[x]$ (esercizio 1 qui sotto). Sia p primo. Riducendo i coefficienti modulo p si ottiene l'omomorfismo canonico di $\mathbb{Z}[x]$ su $(\mathbb{Z}/(p))[x]$. Per $f(x) \in \mathbb{Z}[x]$ denotiamo con $f_p(x)$ il polinomio corrispondente in $(\mathbb{Z}/(p))[x]$. Se $f(x)$ è monico di grado n, anche $f_p(x)$ è monico di grado n. Abbiamo visto (pp. 258–259 dell'originale) che $d(f)$ è un polinomio a coefficienti interi nei coefficienti a_i di f. Quindi $d(f) \in \mathbb{Z}$ e $d(f_p) = d(f)_p$ si ottiene riducendo $d(f)$ modulo p. Pertanto, se $d(f_p) \neq 0$, anche $d(f) \neq 0$ e sia f sia f_p hanno radici distinte. In questo caso dimostreremo il seguente risultato.

Teorema 4.37 (Dedekind). *Sia $f(x) \in \mathbb{Z}[x]$ monico di grado n e sia p un primo tale che $f_p(x)$ abbia radici distinte (equivalentemente $d(f_p) \neq 0$). Supponiamo che $f_p(x)$ si scomponga in $(\mathbb{Z}/(p))[x]$ in fattori irriducibili di gradi $n_1, \ldots, n_r$, con $\sum_i n_i = n$. Allora il gruppo di Galois G_f contiene una permutazione delle radici di f la cui decomposizione in cicli, per un opportuno ordinamento delle radici, è*

$$(1\ 2\ \cdots\ n_1)(n_1 + 1\ \cdots\ n_1 + n_2)(n_1 + n_2 + 1\ \cdots\ n_1 + n_2 + n_3)\cdots .$$

Prima di dimostrare il teorema, illustriamone l'impiego determinando G_f per $f(x) = x^6 + 22x^5 + 21x^4 + 12x^3 - 37x^2 - 29x - 15$. Riducendo modulo 2 otteniamo $f_2(x) = x^6 + x^4 + x^2 + x + 1$. Verificando la divisibilità per i polinomi irriducibili modulo 2 di grado ≤ 3, si vede che $f_2(x)$ è irriducibile. Dunque G_f contiene un 6-ciclo ed è transitivo. Si ha poi $f_3(x) = x(x^5 + x^4 - x + 1)$ e si può mostrare che $x^5 + x^4 - x + 1$ è irriducibile modulo 3. Pertanto G_f contiene un 5-ciclo. Modulo 5 si ha $f_5(x) = x(x-1)(x+1)(x+2)(x^2+2)$ e x^2+2 è irriducibile modulo 5. Il gruppo G_f contiene quindi un 2-ciclo. Si può dimostrare (esercizio 3 qui sotto) che un sottogruppo transitivo di S_n che contenga un $(n-1)$-ciclo e una trasposizione coincide con S_n. Pertanto $G_f = S_6$.

La dimostrazione del teorema 4.37 che daremo è dovuta a John Tate e si basa sul seguente risultato.

Teorema 4.38. *Sia $f(x) \in \mathbb{Z}[x]$ monico di grado n, sia E un campo di spezzamento di $f(x)$ su $\mathbb{Q}$ e sia p un primo tale che $f_p(x)$ abbia radici distinte nel suo campo di spezzamento E_p su $\mathbb{Z}/(p)$. Sia D il sottoanello di E generato dalle radici di $f(x)$. Allora*

(a) *esistono omomorfismi ψ di D in E_p;*

(b) *ogni tale omomorfismo induce una biiezione dall'insieme R delle radici di $f(x)$ in E sull'insieme R_p delle radici di $f_p(x)$ in E_p;*

(c) *se ψ, ψ' sono due tali omomorfismi, allora $\psi' = \psi\sigma$ per qualche $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$.*

Dimostrazione. (a) Si ha $E = \mathbb{Q}(r_1, \ldots, r_n)$ e $f(x) = \prod_{i=1}^n (x - r_i)$ in $E[x]$. Le r_i sono distinte, poiché $d(f_p) \neq 0$ e quindi $d(f) \neq 0$. Inoltre $D = \mathbb{Z}[r_1, \ldots, r_n]$. Poniamo

$$D' = \sum_{0 \leq e_i \leq n-1} \mathbb{Z} r_1^{e_1} \cdots r_n^{e_n},$$

insieme delle combinazioni $\mathbb{Z}$-lineari degli elementi $r_1^{e_1} \cdots r_n^{e_n}$, $0 \leq e_i \leq n-1$. Poiché $f(r_i) = 0$, r_i^n è combinazione $\mathbb{Z}$-lineare di $1, r_i, \ldots, r_i^{n-1}$. Quindi $r_i D' \subseteq D'$ e, iterando, $r_1^{f_1} \cdots r_n^{f_n} D' \subseteq D'$ per ogni scelta di interi positivi f_i. Pertanto D' è un sottoanello di D che contiene le r_i, e dunque $D' = D$. Ne segue che D è un $\mathbb{Z}$-modulo finitamente generato. Poiché E ha caratteristica zero, il sottomodulo di torsione di D è nullo. Quindi D è un $\mathbb{Z}$-modulo libero con una base, sia essa $(u_1, \ldots, u_N)$: $D = \mathbb{Z}u_1 \oplus \cdots \oplus \mathbb{Z}u_N$ (p. 190 dell'originale). Gli u_i costituiscono una base di $E/\mathbb{Q}$, sicché $N = [E : \mathbb{Q}]$. Infatti l'indipendenza lineare su $\mathbb{Q}$ è chiara: da una relazione $\mathbb{Q}$-lineare non banale fra gli u_i, moltiplicando per un intero non nullo, se ne ricaverebbe una $\mathbb{Z}$-lineare non banale. Consideriamo $\mathbb{Q}D = \sum_i \mathbb{Q}u_i$. È un sottoanello di E contenente $\mathbb{Q}$. Per l'esercizio 7, p. 216 dell'originale, $\mathbb{Q}D$ è un sottocampo di E. Contenendo le r_i, coincide con E. Dunque gli u_i generano lo spazio vettoriale $E/\mathbb{Q}$ e ne formano una base.

Consideriamo $pD = \sum_i \mathbb{Z}(pu_i)$. È un ideale di D e $|D/pD| = p^N$. L'anello finito D/pD contiene un ideale massimale proprio, della forma M/pD dove M è un ideale massimale di D contenente pD. Allora D/M è un campo, immagine omomorfa di D/pD, poiché $(D/pD)/(M/pD) \cong D/M$. Come D/pD, anche D/M ha caratteristica p. Il suo campo primo è dunque $\mathbb{Z}/(p)$ e $|D/M| = p^m$, con $m \leq N$. L'omomorfismo canonico ν di D su D/M manda $\mathbb{Z}$ sul campo primo $\mathbb{Z}/(p)$. Poiché $D = \mathbb{Z}[r_1, \ldots, r_n]$ e $f(x) = \prod_{i=1}^n (x - r_i)$ in $D[x]$, si ha

$$D/M = (\mathbb{Z}/(p))[\bar{r}_1, \ldots, \bar{r}_n], \qquad \bar{r}_i = \nu(r_i) = r_i + M, \qquad \bar{f}(x) = \prod_{i=1}^n (x - \bar{r}_i).$$

Da $f(x) \in \mathbb{Z}[x]$ segue che i coefficienti di $\bar{f}(x)$ appartengono a $\mathbb{Z}/(p)$ e $\bar{f}(x) = f_p(x)$. Pertanto D/M è un campo di spezzamento di $f_p(x)$ su $\mathbb{Z}/(p)$. Poiché E_p è stato scelto con la stessa proprietà, esiste un isomorfismo di D/M su E_p. Componendolo con ν, otteniamo un omomorfismo ψ di D su E_p. È provato (a).

(b) Sia ψ un omomorfismo di D in E_p. La restrizione $\psi|_{\mathbb{Z}}$ è un omomorfismo di $\mathbb{Z}$ sul campo primo di E_p. Poiché manda 1 nell'unità di E_p, è l'omomorfismo canonico di $\mathbb{Z}$ su $\mathbb{Z}/(p)$. Quindi $f_p(x) = \psi(f(x)) = \prod_i (x - \psi(r_i))$, applicando ψ ai coefficienti. Gli $\psi(r_i)$ sono perciò le radici di $f_p(x)$ in E_p e $\psi|_R$ è una biiezione di R su R_p.

(c) Fissiamo un omomorfismo $\psi : D \to E_p$. Sia $\sigma \in G = \mathrm{Gal}(E/\mathbb{Q})$. Poiché permuta le r_i, σ manda D in sé. La restrizione $\sigma|_D$ è un omomorfismo di D in sé (anzi un automorfismo) e $\psi\sigma$ è un omomorfismo di D in E_p. Elementi distinti $\sigma, \sigma' \in G$ inducono permutazioni distinte delle r_i. Essendo $\psi|_R$ biiettiva su R_p, anche $\psi\sigma$ e $\psi\sigma'$ sono distinti. Otteniamo così $N = [E : \mathbb{Q}]$ omomorfismi distinti $\psi_j = \psi\sigma_j$, con $G = \{\sigma_1, \ldots, \sigma_N\}$. Affermiamo che non ve ne sono altri. Sia infatti ψ_{N+1} distinto da tutti gli ψ_j, $1 \le j \le N$. Per il teorema di indipendenza di Dedekind, applicato al monoide moltiplicativo H del dominio D e al campo $F = E_p$, gli ψ_i, incluso ψ_{N+1}, sono linearmente indipendenti su E_p. Consideriamo d'altra parte il sistema

$$\sum_{i=1}^{N+1} x_i \psi_i(u_j) = 0, \qquad 1 \le j \le N.$$

Poiché vi sono più incognite che equazioni, il sistema lineare omogeneo a coefficienti $\psi_i(u_j) \in E_p$ ha una soluzione non banale $(a_1, \ldots, a_{N+1})$, $a_i \in E_p$. Per $y \in D$ si ha $y = \sum_{j=1}^{N} n_j u_j$, $n_j \in \mathbb{Z}$. Quindi $\psi_i(y) = \sum_j \bar{n}_j \psi_i(u_j)$, con $\bar{n}_j = n_j + (p)$, e $\sum_i a_i \psi_i(y) = \sum_{i,j} \bar{n}_j a_i \psi_i(u_j) = 0$. Ciò contraddice l'indipendenza degli ψ_i e conclude la prova di (c). □

Possiamo ora servirci di questo risultato per la

Dimostrazione del teorema 4.37. Poiché E_p è un campo di p^m elementi, l'applicazione $\pi : a \mapsto a^p$ è un automorfismo di E_p. Per ogni omomorfismo ψ di D in E_p anche $\pi\psi$ è dunque un tale omomorfismo. Esiste perciò un unico $\sigma(\psi) \in G$ con $\pi\psi = \psi\sigma(\psi)$. L'automorfismo $\sigma = \sigma(\psi)$ si chiama *automorfismo di p-Frobenius* di $E/\mathbb{Q}$ corrispondente a ψ. Restringendo ψ e σ a R e usando la biiettività di ψ da R su R_p, otteniamo $\sigma = \psi^{-1}\pi\psi$. Dunque ψ^{-1} manda le orbite di $\langle\pi\rangle$ in R_p nelle orbite di $\langle\sigma\rangle$ in R. Le orbite di $\langle\pi\rangle$ in R_p sono gli insiemi delle radici dei fattori irriducibili di $f_p(x)$ in $(\mathbb{Z}/(p))[x]$. Se questi hanno gradi $n_1, \ldots, n_r$, le orbite di $\langle\sigma\rangle$ in R hanno cardinalità $n_1, \ldots, n_r$, quindi σ, come permutazione di R, ha la decomposizione in cicli $(1\ 2\ \cdots\ n_1)(n_1 + 1\ \cdots\ n_1 + n_2)\cdots$ per un opportuno ordinamento delle radici. □

Esercizi

1. Sia $f(x) \in \mathbb{Q}[x]$ monico; si scriva $f(x) = x^n - a_1 x^{n-1} + a_2 x^{n-2} - \cdots + (-1)^n a_n$, con $a_i = b_i d^{-1}$ e $b_i, d \in \mathbb{Z}$. Mostrare che $d^n f(d^{-1}x) \in \mathbb{Z}[x]$ è monico e ha lo stesso campo di spezzamento di $f(x)$ su $\mathbb{Q}$.

2. Sia $f(x) \in \mathbb{Z}[x]$ monico con radici distinte. Mostrare che il teorema 4.37 si applica a tutti i primi salvo un numero finito.

3. Mostrare che se G è un sottogruppo transitivo di S_n che contiene un $(n-1)$-ciclo e una trasposizione, allora $G = S_n$.

4. Determinare G_f per $f(x) = x^6 - 12x^4 + 15x^3 - 6x^2 + 15x + 12$ su $\mathbb{Q}$.

5. (Tate.) Mostrare che per ogni primo p e ogni intero positivo n esiste un polinomio monico irriducibile di grado n in $(\mathbb{Z}/(p))[x]$ (usare la (69) o la sua dimostrazione). Fissato n, si prendano $g(x)$ monico irriducibile di grado n in $(\mathbb{Z}/(2))[x]$, $h(x)$ monico irriducibile di grado $n-1$ in $(\mathbb{Z}/(3))[x]$ e $k(x)$ monico irriducibile quadratico modulo un primo $p > n-2$. Usare il teorema cinese del resto (esercizio 10, p. 110 dell'originale) per mostrare che esiste $f(x) \in$

$\mathbb{Z}[x]$ monico tale che $f_2(x) = g(x)$, $f_3(x) = xh(x)$ e $f_p(x) = x(x+1)\cdots(x+n-3)k(x)$. Mostrare che $G_f = S_n$.

6. Mostrare che ogni sottogruppo transitivo di A_5 è isomorfo a uno dei seguenti tre gruppi: (a) il gruppo ciclico Z_5, (b) il gruppo diedrale D_5, (c) A_5.

7. (Jensen e Yui.) Sia $f(x) = x^5 - 5x + 12$. Il polinomio è irriducibile in $\mathbb{Q}[x]$ e $d(f) = (2^6 5^3)^2$. Se $r_1, \ldots, r_5$ sono le radici di f, sia $P(x) = \prod_{1 \le i < j \le 5}(x - (r_i + r_j))$. Allora $P(x)$ è prodotto di due distinti polinomi monici irriducibili in $\mathbb{Q}[x]$:

$$P(x) = (x^5 - 5x^3 - 10x^2 + 30x - 36)(x^5 + 5x^3 + 10x^2 + 10x + 4).$$

Usare queste informazioni, l'esercizio 6 e f_3 per mostrare che $G_f \cong D_5$.

8. (Jensen e Yui.) Sia $f(x) = x^5 + 20x + 16$. Si ha $d(f) = (2^8 5^3)^2$ e, definendo $P(x)$ come nell'esercizio 7, il polinomio

$$P(x) = x^{10} - 60x^6 - 176x^5 - 1600x^2 + 1280x - 256$$

è irriducibile in $\mathbb{Q}[x]$. Usare queste informazioni per mostrare che $G_f \cong A_5$.

Capitolo 5
Equazioni e disuguaglianze polinomiali reali

L'obiettivo principale del capitolo è la teoria delle equazioni e delle disuguaglianze polinomiali in più incognite nel campo $\mathbb{R}$ dei numeri reali. Le proprietà fondamentali di $\mathbb{R}$ da cui prende le mosse l'analisi sono racchiuse nell'affermazione che $\mathbb{R}$ è un campo ordinato completo: vi è cioè una relazione $>$ che soddisfa gli assiomi di campo ordinato (dati nel § 5.1) e vale l'assioma di completezza, secondo cui ogni sottoinsieme non vuoto di $\mathbb{R}$ dotato di un maggiorante ha un estremo superiore. Poiché ci occuperemo soltanto di funzioni polinomiali, è naturale che qui non sia necessaria tutta la forza di queste proprietà. Vedremo che basta assumere di avere un campo ordinato R tale che (1) gli elementi positivi di R ammettano radici quadrate in R e (2) ogni equazione di grado dispari in un'incognita e a coefficienti in R abbia una radice in R. Un campo ordinato che soddisfi queste condizioni si dirà *reale chiuso*. È chiaro che $\mathbb{R}$ le soddisfa. Anche il sottocampo degli elementi di $\mathbb{R}$ algebrici su $\mathbb{Q}$ è un campo reale chiuso, eppure questo campo ordinato non gode della classica proprietà di completezza.

Mostreremo che se R è reale chiuso, $R(\sqrt{-1})$ è algebricamente chiuso. Prendendo $R = \mathbb{R}$ si ha $\mathbb{C} = \mathbb{R}(\sqrt{-1})$: ne seguirà come corollario il «teorema fondamentale dell'algebra», secondo il quale il campo dei numeri complessi è algebricamente chiuso.

Ci interesserà soprattutto sviluppare algoritmi per decidere se un dato sistema di equazioni, condizioni di diversità ($\neq$) e disuguaglianze ($>$) polinomiali a coefficienti in R ammetta soluzioni in R. Il primo risultato definitivo di questo genere è un teorema classico dimostrato da J. C. F. Sturm nel 1836. Il risultato più generale in questa direzione è un'ampia estensione del teorema di Sturm, dimostrata da A. Tarski intorno al 1930. Ne daremo una dimostrazione alternativa dovuta ad A. Seidenberg. Prima di passare dal teorema di Sturm a quello di Tarski studieremo l'eliminazione delle variabili nei sistemi di equazioni e condizioni di diversità a coefficienti in un campo qualsiasi.

5.1 Campi ordinati. Campi reali chiusi

Definiamo un campo ordinato mediante il suo insieme P degli elementi positivi.

Definizione 5.1. *Un campo ordinato (F, P) è un campo F con un sottoinsieme P di F (l'insieme degli elementi positivi) tale che (1) $0 \notin P$; (2) per $a \in F$ si ha $a \in P$, oppure $a = 0$, oppure $-a \in P$; (3) se $a, b \in P$, allora $a + b, ab \in P$. Un campo F è ordinabile se è possibile specificare in F un sottoinsieme P che abbia queste proprietà.*

Poiché ogni campo ha più di un elemento, in un campo ordinato (F, P) l'insieme P non è vuoto. Se $N = \{-a \mid a \in P\}$, la (2) dice che $F = P \cup \{0\} \cup N$. Dalla (1) risulta inoltre che $P \cap \{0\} = N \cap \{0\} = \varnothing$. Si ha anche $P \cap N = \varnothing$: se $a \in P \cap N$, anche $-a \in P \cap N$, e allora $0 = a + (-a) \in P$, contro la (1). La decomposizione di F è dunque in sottoinsiemi disgiunti. L'insieme N è chiuso rispetto all'addizione, poiché per $a, b \in P$ si ha $-a + (-b) = -(a + b) \in N$. Se invece $a, b \in N$, allora $ab = (-a)(-b) \in P$.

In (F, P) introduciamo la relazione d'ordine $a > b$ ponendo $a > b$ quando $a - b \in P$. Per due elementi $a, b \in F$ vale la tricotomia: sussiste una e una sola delle tre alternative $a > b$, $a = b$, $b > a$. Se $a > b$, allora $a + c > b + c$ per ogni c, e $ap > bp$ per ogni p positivo. Viceversa, si potrebbe partire da una relazione $>$ su un campo che soddisfi la tricotomia, la transitività e le due proprietà per cui da $a > b$ seguono $a + c > b + c$ e, per $p > 0$, $ap > bp$. Ponendo $P = \{p \mid p > 0\}$ si ottiene un campo ordinato nel senso della definizione precedente, e la relazione $>$ associata coincide con quella data.

Come di consueto scriveremo $a < b$ per $b > a$. Si stabiliscono facilmente le proprietà elementari delle disuguaglianze familiari nel campo reale. Ne elenchiamo alcune. Se $a > 0$, allora $a^{-1} > 0$; se $a > b > 0$, allora $b^{-1} > a^{-1} > 0$. Se $a > b$, si ha $-a < -b$, e se inoltre $c > d$, si ha $a + c > b + d$. Poniamo $|a| = a$ per $a \geq 0$ e $|a| = -a$ per $a < 0$; si dimostra che $|a + b| \leq |a| + |b|$ e che $|ab| = |a||b|$.

Se F' è un sottocampo di (F, P), allora (F', P'), con $P' = F' \cap P$, è un campo ordinato. Si parla di *ordine indotto* su F'. Dati due campi ordinati (F, P) e (F', P'), un isomorfismo η di F su F' è un *isomorfismo d'ordine* se $\eta(P) \subset P'$. Poiché $\eta(0) = 0$ e $\eta(N) \subset N'$, segue che $\eta(P) = P'$.

In un campo ordinato (F, P), da $a \neq 0$ segue $a^2 > 0$. Se $a_1, \ldots, a_r \neq 0$, allora $\sum a_i^2 \neq 0$. In particolare $1 + 1 + \cdots + 1 = 1^2 + \cdots + 1^2 \neq 0$: ogni campo ordinato ha quindi caratteristica zero. Non si può neppure avere $-1 = \sum a_i^2$, poiché ne seguirebbe $1^2 + \sum a_i^2 = 0$. In particolare $-1 \neq a^2$ per ogni $a \in F$: F non contiene alcuna radice quadrata di -1. Ne risulta che il campo complesso $\mathbb{C}$ non è ordinabile.

Nel campo $\mathbb{R}$ dei numeri reali è facile dedurre dall'assioma di completezza le due proprietà seguenti.

(i) Ogni elemento positivo ha una radice quadrata in $\mathbb{R}$.

(ii) Ogni equazione polinomiale $f(x) = 0$, dove $f(x) \in \mathbb{R}[x]$ ha grado dispari, ha una radice in $\mathbb{R}$.

Entrambe seguono dal teorema dei valori intermedi: se f è continua, $a < b$ e $f(a)f(b) < 0$, esiste un c con $a < c < b$ e $f(c) = 0$. Diremo dunque *reale chiuso*[1] un campo ordinato (R, P) che abbia le proprietà (i) e (ii), con R al posto di $\mathbb{R}$. Vale il seguente teorema.

Teorema 5.1. *Un campo reale chiuso ammette un unico ordinamento che lo renda un campo ordinato. Ogni suo automorfismo è un isomorfismo d'ordine. Se R è reale chiuso, è reale chiuso anche il suo sottocampo degli elementi algebrici su $\mathbb{Q}$ ($\subset R$).*

Dimostrazione. Sia (R, P) reale chiuso e sia (R, P') una qualsiasi struttura di campo ordinato su R. Se $a \in P$, allora $a = b^2$ con $b \neq 0$. Dunque $a \in P'$. Ne segue $P \subset P'$, e quindi $P = P'$. La seconda affermazione si ottiene allo stesso modo. Sia ora R_0 il sottocampo di R formato dagli elementi algebrici su $\mathbb{Q}$ (cfr. § 4.12, p. 280 dell'originale). Se $a \in R_0$ e $a > 0$, esiste $b \in R$ con $b^2 = a$. Quindi b è algebrico su R_0 e perciò appartiene a R_0. La condizione (i) vale in R_0. Analogamente si verifica la (ii), e R_0 è reale chiuso. □

In particolare, il campo dei numeri algebrici reali, cioè il sottocampo dei numeri reali algebrici su $\mathbb{Q}$, è reale chiuso. Questo sottocampo, naturalmente, non è completo. Gli assiomi qui adottati sono dunque più deboli dell'assioma di completezza.

Dimostriamo ora, per i campi reali chiusi, l'analogo del «teorema fondamentale dell'algebra».

Teorema 5.2. *Se R è reale chiuso, $R(\sqrt{-1})$ è algebricamente chiuso.*

Dimostrazione. La dimostrazione che daremo è dovuta ad Artin e segue piuttosto da vicino una dimostrazione di Gauss del risultato classico. Osserviamo anzitutto che $\sqrt{-1} \notin R$ e che su $C = R(\sqrt{-1})$ abbiamo l'automorfismo $r = a + b\sqrt{-1} \mapsto \bar{r} = a - b\sqrt{-1}$, con $a, b \in R$. Se $f(x) \in C[x]$, allora $f(x)\bar{f}(x) \in R[x]$; se questo prodotto ha una radice in C, anche f ha una radice in C. Per dimostrare che C è algebricamente chiuso basta dunque provare che ogni polinomio monico di grado positivo a coefficienti in R ha una radice in C. Per i polinomi di grado dispari ciò segue da (ii). Mostriamo ora che ogni elemento di C ha una radice quadrata

[1] Questa nozione equivale a un'altra che ha un ruolo centrale nella teoria dei campi formalmente reali dovuta ad Artin e Schreier. Se ne dà un'esposizione nel capitolo 11 del secondo volume.

in questo campo. Per gli elementi $a \geq 0$ di R ciò segue da (i); se $a \in R$, $a < 0$ e $b^2 = -a$, allora $(b\sqrt{-1})^2 = a$. Sia ora $r = a + b\sqrt{-1}$ con $a, b \in R$ e $b \neq 0$. Poniamo $i = \sqrt{-1}$ e siano $x, y \in R$. La condizione $(x + iy)^2 = r$ equivale a

$$x^2 - y^2 = a, \qquad 2xy = b. \tag{1}$$

Poiché $b \neq 0$, moltiplicando per un opportuno elemento di R (che ha una radice quadrata in C) possiamo supporre $b = 2$. La seconda equazione diventa $xy = 1$, ossia $y = x^{-1}$. La prima diventa allora $x^2 - x^{-2} = a$, cioè $z - z^{-1} = a$ per $z = x^2$. L'equazione $z^2 - az - 1 = 0$ ammette in R la soluzione $(a + \sqrt{a^2 + 4})/2$, poiché $a^2 + 4 > 0$.[2] Inoltre $a + \sqrt{a^2 + 4} > 0$, poiché la disuguaglianza contraria porterebbe a $4 \leq 0$. Esiste quindi $x \neq 0$ in R con $x^2 = \frac{1}{2}(a + \sqrt{a^2 + 4})$. Allora $x^4 - ax^2 = 1$ e $x^2 - x^{-2} = a$. Dunque x e $y = x^{-1}$ soddisfano (1) con $b = 2$.

Abbiamo così provato che ogni elemento di C ha una radice quadrata in C. Ne segue che non esistono estensioni E/C con $[E : C] = 2$. Serviamoci di questo fatto per provare che ogni polinomio monico non costante a coefficienti in R ha una radice in C. Sia $f(x)$ un tale polinomio e sia E un campo di spezzamento su R di $f(x)(x^2 + 1)$ che supponiamo contenga C. Poiché la caratteristica è zero, E è un'estensione di Galois di R. Sia $G = \mathrm{Gal}(E/R)$, con $|G| = 2^e m$ e m dispari. Per il teorema di Sylow, G ha un sottogruppo H di ordine 2e. Se D è il sottocampo corrispondente dell'estensione E/R, si ha $[E : D] = 2^e$ e $[D : R] = m$. Poiché R non ha estensioni proprie di grado dispari, deve essere $m = 1$, sicché $D = R$ e $[E : R] = 2^e$. Il gruppo di Galois è $G = H$, di ordine 2e. Un tale gruppo è risolubile. Se $e > 1$, dalla teoria di Galois (cfr. § 4.11, p. 271 dell'originale) segue facilmente che E contiene un sottocampo F contenente C con $[F : C] = 2$, in contraddizione con quanto abbiamo appena provato. Pertanto $e = 1$ ed $E = C$. Il campo C contiene una radice di $f(x)$ ed è dunque algebricamente chiuso. □

Dal teorema segue subito che i polinomi monici irriducibili di $R[x]$ hanno grado uno o due. Dalla formula risolutiva delle equazioni di secondo grado risulta inoltre che $x^2 + ax + b$ è irriducibile in $R[x]$ se e solo se $a^2 < 4b$.

Il fatto che $R(\sqrt{-1})$ sia algebricamente chiuso permette di stabilire, per le funzioni polinomiali su R, varie proprietà fondamentali delle funzioni continue e derivabili di una variabile reale. Ci occorrerà, in particolare, il teorema dei valori intermedi per i polinomi.

Teorema 5.3. *Sia R un campo reale chiuso, sia $f(x) \in R[x]$ e siano $a, b \in R$ tali che $f(a)f(b) < 0$. Esiste allora un elemento c fra a e b tale che $f(c) = 0$.*

Dimostrazione. Possiamo supporre $f(x)$ monico. In $R[x]$ esso si fattorizza nella forma

$$f(x) = (x - r_1) \cdots (x - r_m) g_1(x) \cdots g_s(x), \qquad g_i(x) = x^2 + c_i x + d_i, \quad c_i^2 < 4d_i.$$

Abbiamo

$$g_i(x) = \left(x + \frac{c_i}{2}\right)^2 + \frac{4d_i - c_i^2}{4} = \left(x + \frac{c_i}{2}\right)^2 + e_i^2, \qquad e_i = \tfrac{1}{2}\sqrt{4d_i - c_i^2}.$$

Dunque $g_i(u) > 0$ per ogni $u \in R$. Se $a, b < r_i$ per ogni i, i fattori $(a - r_i)(b - r_i)$ sono positivi e anche $f(a)f(b) > 0$. Lo stesso vale se $a, b > r_i$ per ogni i. Più in generale, affinché il prodotto sia negativo occorre che almeno uno dei fattori $(a - r_i)(b - r_i)$ sia negativo. Dunque uno degli r_i si trova fra a e b, e $f(r_i) = 0$. □

Esercizi

1. (Veblen.) Sia F un campo che soddisfa due assiomi: (i) -1 non è un quadrato in F; (ii) la somma di due elementi non quadrati è un elemento non quadrato. Dimostrare che esiste uno e un solo ordinamento che renda F un campo ordinato.

[2] Come di consueto per $\mathbb{R}$, il radicale di un elemento positivo indica la radice quadrata positiva.

2. Dimostrare che $\mathbb{Q}(\sqrt{2})$ ammette esattamente due ordinamenti che lo rendono un campo ordinato.

3. Sia F un campo ordinato e sia x un'indeterminata su F. Dimostrare che $F(x)$ è ordinato se si pone
$$\frac{a_0x^n + a_1x^{n-1} + \cdots + a_n}{b_0x^m + b_1x^{m-1} + \cdots + b_m} > 0 \quad \Longleftrightarrow \quad a_0b_0 > 0.$$

4. Sia F un campo ordinato e sia $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$ a coefficienti in F. Posto $M = \max(1, |a_1| + \cdots + |a_n|)$, provare che da $|u| > M$ segue $|f(u)| > 0$. Dedurne che ogni radice di $f(x)$ in F appartiene all'intervallo $-M \leq x \leq M$.

   Nei tre esercizi seguenti R è un campo reale chiuso.

5. Dimostrare il teorema di Rolle per i polinomi: se $f(a) = 0 = f(b)$ e $a < b$, esiste c, $a < c < b$, con $f'(c) = 0$.

6. Dimostrare il teorema del valor medio per i polinomi: se $a < b$, esiste c, $a < c < b$, tale che $f(b) - f(a) = (b - a)f'(c)$.

7. Dimostrare che $f(x)$ ammette massimo su ogni intervallo chiuso e limitato $a \leq x \leq b$.

5.2 Il teorema di Sturm

In questo paragrafo dedurremo un risultato classico, il teorema di Sturm, che permette di determinare il numero esatto delle radici di un'equazione polinomiale $f(x) = 0$ in un campo reale chiuso. Seguiremo piuttosto da vicino l'esposizione di Weber nel *Lehrbuch der Algebra* (1898), vol. 1, pp. 301–313.

Sia R un campo reale chiuso e sia $f(x)$ un polinomio di grado positivo a coefficienti in R. Seguendo Weber, diremo che una successione di polinomi

$$f_0(x) = f(x), f_1(x), \ldots, f_s(x) \tag{2}$$

è una *successione di Sturm* per $f(x)$ sull'intervallo chiuso $[a, b]$ (ossia $a \leq x \leq b$) se gli $f_j(x) \in R[x]$ soddisfano le condizioni seguenti:

(i) $f_s(x)$ non ha radici in $[a, b]$.

(ii) $f_0(a)f_0(b) \neq 0$.

(iii) Se $c \in [a, b]$ è radice di $f_j(x)$, con $0 < j < s$, allora $f_{j-1}(c)f_{j+1}(c) < 0$.

(iv) Se $f(c) = 0$ per $c \in [a, b]$, esistono intervalli aperti (c_1, c) e (c, c_2) tali che $f_0(u)f_1(u) < 0$ per ogni u nel primo intervallo e $f_0(u)f_1(u) > 0$ per ogni u nel secondo.

Stabiliremo l'esistenza di tali successioni per ogni polinomio a radici distinte. Vediamo prima in che modo una tale successione permetta di determinare il numero delle radici di $f(x)$ nell'intervallo aperto (a, b). Consideriamo il numero delle variazioni di segno nelle due successioni di elementi di R
$$f_0(a), f_1(a), \ldots, f_s(a), \qquad f_0(b), f_1(b), \ldots, f_s(b).$$
Se $c = (c_1, c_2, \ldots, c_m)$ è una successione finita di elementi non nulli di R, il suo *numero di variazioni di segno* è il numero degli indici i, $1 \leq i \leq m - 1$, per cui $c_ic_{i+1} < 0$. Per una

successione arbitraria c si prende il numero delle variazioni di segno della sottosuccessione c' ottenuta sopprimendo gli zeri in c. Per esempio,

$$(1, 0, 0, 2, -1, 0, 3, 4, -2)$$

presenta tre variazioni di segno. Possiamo ora enunciare il teorema.

Teorema 5.4. *Sia $f(x)$ un polinomio di grado positivo a coefficienti in un campo reale chiuso R e sia $f_0(x) = f(x), f_1(x), \ldots, f_s(x)$ una successione di Sturm per $f(x)$ su $[a, b]$. Il numero delle radici distinte di $f(x)$ in (a, b) è $V_a - V_b$, dove V_c indica, in generale, il numero delle variazioni di segno nella successione $(f_0(c), f_1(c), \ldots, f_s(c))$.*

Dimostrazione. Le radici dei polinomi $f_j(x)$ della successione di Sturm suddividono $[a, b]$ in sottointervalli. Si ha dunque una successione $a = a_0 < a_1 < \cdots < a_m = b$ tale che nessuno degli $f_j(x)$ ha radici in (a_i, a_{i+1}). Sia dapprima $c \in (a_0, a_1)$: nessun f_j ha radici in (a_0, c). Per il teorema dei valori intermedi (teorema 5.3), $f_j(a_0)f_j(c) \geq 0$ per $0 \leq j \leq s$. Se tutti gli $f_j(a_0)$ sono diversi da zero, i prodotti sono positivi, e quindi $V_{a_0} = V_c$. Supponiamo ora che $f_k(a_0) = 0$ per qualche indice k. Per le proprietà della successione $f_0(a) \neq 0$ e $f_s(a) \neq 0$, perciò $0 < k < s$. La (iii) dà $f_{k-1}(a_0)f_{k+1}(a_0) < 0$. Poiché f_{k-1} e f_{k+1} non hanno radici in (a_0, c), si ha $f_{k-1}(a_0)f_{k-1}(c) > 0$ e $f_{k+1}(a_0)f_{k+1}(c) > 0$. Ne segue $f_{k-1}(c)f_{k+1}(c) < 0$. Dunque le terne $f_{k-1}(a_0), 0, f_{k+1}(a_0)$ e $f_{k-1}(c), f_k(c), f_{k+1}(c)$ contribuiscono ciascuna con una variazione di segno, rispettivamente a V_{a_0} e a V_c. Considerando tutti gli indici k si ottiene $V_{a_0} = V_c$. In modo analogo, se $d \in (a_{m-1}, a_m)$, si ha $V_d = V_{a_m}$.

Siano ora $c \in (a_{i-1}, a_i)$ e $d \in (a_i, a_{i+1})$, con $1 \leq i < m$. Lo stesso ragionamento dà $V_c = V_d$ purché $f(a_i) \neq 0$. Supponiamo invece $f(a_i) = 0$. Per (iv) si ha $f_0(c)f_1(c) < 0$ e $f_0(d)f_1(d) > 0$. La coppia $f_0(c), f_1(c)$ presenta dunque una variazione di segno, mentre $f_0(d), f_1(d)$ non ne presenta alcuna. Il ragionamento precedente mostra che, per $j > 1$, le terne $f_{j-1}(c), f_j(c), f_{j+1}(c)$ e $f_{j-1}(d), f_j(d), f_{j+1}(d)$ danno lo stesso contributo in corrispondenza di uno zero intermedio. Pertanto $V_c - V_d = 1$ se $f(a_i) = 0$. Scegliamo ora $a_i' \in (a_{i-1}, a_i)$. Allora

$$V_a - V_b = (V_a - V_{a_1'}) + \sum_{i=1}^{m-1} (V_{a_i'} - V_{a_{i+1}'}) + (V_{a_m'} - V_b).$$

Ciascuna parentesi del secondo membro vale zero o uno, e il numero di volte in cui vale uno coincide con il numero degli a_i radici di $f(x)$. Dunque $V_a - V_b$ conta le radici di $f(x)$ in (a, b). □

Sia ora $f(x)$ un qualsiasi polinomio di grado positivo in $R[x]$. Definiamo la sua *successione standard* mediante

$$\begin{aligned}
f_0(x) &= f(x), \quad f_1(x) = f'(x) && \text{(derivata formale)},\\
f_0(x) &= q_1(x)f_1(x) - f_2(x), && \deg f_2 < \deg f_1,\\
&\vdots\\
f_{i-1}(x) &= q_i(x)f_i(x) - f_{i+1}(x), && \deg f_{i+1} < \deg f_i,\\
&\vdots\\
f_{s-1}(x) &= q_s(x)f_s(x) && \text{(ossia } f_{s+1}(x) = 0).
\end{aligned} \tag{3}$$

Si ottengono dunque gli f_i modificando l'algoritmo euclideo per il massimo comune divisore di f e f': a ogni passo si prende l'opposto del resto della divisione. Per esempio, per $f(x) = x^3 + x + 1$ si ha $f_0(x) = f(x)$, $f_1(x) = 3x^2 + 1$ e

$$f_0(x) = \tfrac{1}{3}x f_1(x) - (-\tfrac{2}{3}x - 1).$$

Quindi $f_2(x) = -\frac{2}{3}x - 1$. Poiché $f_1(x) = (-\frac{9}{2}x + \frac{27}{4})f_2(x) - (-\frac{31}{4})$ si ha $f_3(x) = -\frac{31}{4}$. La successione standard è dunque

$$x^3 + x + 1, \quad 3x^2 + 1, \quad -\tfrac{2}{3}x - 1, \quad -\tfrac{31}{4}.$$

In generale, dalla (3) risulta che $f_s(x)$ divide ciascun $f_i(x)$ ed è un massimo comune divisore di f e f'. Poniamo $g_i(x) = f_i(x)/f_s(x)$ e consideriamo la successione

$$g_0(x), g_1(x), \ldots, g_s(x). \tag{4}$$

Mostriamo che essa è una successione di Sturm per $g_0(x)$ su ogni intervallo $[a, b]$ con $g_0(a) \neq 0$ e $g_0(b) \neq 0$. La condizione (ii) è immediata, e vale la (i) poiché $g_s(x) = 1$. Dividendo le uguaglianze (3) per $f_s(x)$ si ottiene

$$g_{j-1}(x) = q_j(x)g_j(x) - g_{j+1}(x). \tag{5}$$

Se $g_j(c) = 0$, la (5) dà $g_{j-1}(c)g_{j+1}(c) \leq 0$; inoltre $g_{j-1}(c) = 0$ se e solo se $g_{j+1}(c) = 0$. In quest'ultimo caso si avrebbe $0 = g_{j-1}(c) = g_j(c) = g_{j+1}(c) = \cdots$, contro $g_s = 1$. Dunque il prodotto è strettamente negativo, e vale la (iii).

Supponiamo ora $g_0(c) = 0$ con $c \in [a, b]$. Si ha $f(x) = (x - c)^e h(x)$, dove $e > 0$ e $h(c) \neq 0$, e

$$f'(x) = (x - c)^e h'(x) + e(x - c)^{e-1} h(x).$$

Inoltre $f_s(x) = (x - c)^{e-1} k(x)$, con $k(c) \neq 0$. Segue che $h(x) = k(x)l(x)$ con $l(c) \neq 0$, e $h'(x) = k(x)m(x)$. Da queste relazioni si ricava

$$\begin{aligned} g_0(x) &= (x - c)l(x), \qquad l(c) \neq 0, \\ g_1(x) &= (x - c)m(x) + e\, l(x). \end{aligned} \tag{6}$$

Pertanto $g_1(c) = e\, l(c) \neq 0$. Scegliamo un intervallo $[c_1, c_2]$ che contenga c al suo interno e in cui $g_1(x)l(x)$ non si annulli. Per il teorema dei valori intermedi e per l'uguaglianza $g_1(c) = e\, l(c)$, il prodotto è positivo in tutto l'intervallo. Quindi $g_0(x)g_1(x) = (x-c)g_1(x)l(x)$ ha il segno di $x-c$: è negativo per $c_1 \leq x < c$ e positivo per $c < x \leq c_2$. Vale così (iv), e la (4) è una successione di Sturm per $g_0(x)$.

Se f non ha radici multiple, f e f' sono relativamente primi e la successione degli f_i si ottiene da quella dei g_i moltiplicando per una costante non nulla di R. La successione degli f_i è allora una successione di Sturm per $f = f_0$. Se invece f ha radici multiple, la successione standard non è una successione di Sturm su un intervallo contenente una radice multipla. Essa può tuttavia servire ancora a determinare il numero delle radici distinte di f in (a, b). È quanto afferma il seguente risultato.

Teorema di Sturm. *Sia $f(x)$ un polinomio di grado positivo a coefficienti in un campo reale chiuso R e sia $f_0 = f, f_1 = f', \ldots, f_s$ la sua successione standard (3). Sia $[a, b]$ un intervallo con $f(a) \neq 0$, $f(b) \neq 0$. Il numero delle radici distinte di $f(x)$ in (a, b) è $V_a - V_b$, dove V_c è il numero delle variazioni di segno nella successione $(f_0(c), f_1(c), \ldots, f_s(c))$.*

Dimostrazione. Poniamo $g_i = f_i/f_s$ come sopra. A prescindere dalle molteplicità, f e g_0 hanno le stesse radici in $[a, b]$ (esercizio 1, p. 233 dell'originale). Poiché i g_i formano una successione di Sturm per g_0, il numero delle radici è $V_a(g) - V_b(g)$, dove $V_c(g)$ conta le variazioni di segno nella successione dei $g_i(c)$. Poiché $f_i(c) = g_i(c)f_s(c)$ e $f_s(a), f_s(b) \neq 0$, si ha $V_a(g) = V_a$ e $V_b(g) = V_b$. Quindi $V_a - V_b$ è il numero richiesto. □

Abbiamo indicato nell'esercizio 4 del § 5.1 che le radici in R del polinomio $x^n + a_1x^{n-1} + \cdots + a_n$ appartengono a $[-M, M]$, dove $M = \max(1, |a_1| + \cdots + |a_n|)$. Se poniamo $\mu = 1 + |a_1| + \cdots + |a_n|$,

le radici appartengono a $(-\mu, \mu)$; pertanto il loro numero è $V_{-\mu} - V_{\mu}$ per la successione standard. Si ottiene così un procedimento costruttivo per determinare il numero delle radici in R. Talvolta è preferibile servirsi di una maggiorazione η polinomiale negli a_j. Osservando che $1 + a_i^2 > |a_i|$, possiamo prendere

$$\eta = 1 + \sum_{i=1}^{n}(1 + a_i^2) = (n+1) + \sum_{i=1}^{n} a_i^2. \tag{7}$$

Le radici in R si trovano allora in $(-\eta, \eta)$.

Esercizi

1. Applicare il teorema di Sturm per dimostrare che $x^3 - 7x - 7$ ha due radici reali in $(-2, -1)$.

2. Determinare con questo teorema il numero delle radici reali di $x^4 + 12x^2 + 5x - 9$.

3. Sia $f(x) = x^3 + px + q$, $p \neq 0$. Mostrare che $f_0 = f$, $f_1 = 3x^2 + p$, $f_2 = -2px - 3q$, $f_3 = -4p^3 - 27q^2$ costituiscono una successione di Sturm per f su ogni $[a, b]$ con $f(a)f(b) \neq 0$ (nel caso $f_3 \neq 0$). Osservare che $f_3 = d$ è il discriminante di f (p. 259 dell'originale). Dal teorema di Sturm dedurre che f ha una sola radice reale se $d < 0$, e tre radici reali distinte se $d > 0$.

4. Siano $f(x) = x^4 + qx^2 + rx + s$, $L = 8qs - 2q^3 - 9r^2$, e sia d il discriminante di f. Provare che, se $d < 0$, il numero delle radici reali è due; se $d > 0$, $q < 0$, $L > 0$, le radici sono quattro e reali distinte; se $d > 0$ e vale almeno una delle condizioni $q \geq 0$, $L \leq 0$, non vi sono radici reali.

5. Si definiscano i polinomi di Legendre $P_0, P_1, \ldots$ mediante la ricorrenza

$$nP_n(x) - (2n-1)xP_{n-1}(x) + (n-1)P_{n-2}(x) = 0, \qquad P_0(x) = 1, \quad P_1(x) = x.$$

   Provare che $P_m, P_{m-1}, \ldots, P_0$ è una successione di Sturm per P_m su $[-1, 1]$ e che P_m ha m radici reali distinte in $(-1, 1)$.

6. Sia $f(x) \in R[x]$, con R reale chiuso e $\deg f = n$. Sia W_c il numero di variazioni di segno della successione $f(c), f'(c), \ldots, f^{(n)}(c)$. Dimostrare il *teorema di Budan*: se $a < b$ e $f(a)f(b) \neq 0$, la differenza fra $W_a - W_b$ e il numero delle radici di f in (a, b), contate con le molteplicità, è un intero pari non negativo.

7. Dedurre dall'esercizio 6 la *regola dei segni di Cartesio*. Sia $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_lx^{n-l}$, con $a_0a_l \neq 0$ e $a_i \in R$. Se P è il numero delle variazioni di segno di $a_0, a_1, \ldots, a_l$, mostrare che la differenza fra P e il numero delle radici positive di $f(x)$, contate con le molteplicità, è un intero pari non negativo.

5.3 Formalizzazione dell'algoritmo euclideo e teorema di Sturm

Nell'ultima parte del capitolo svilupperemo un metodo per stabilire se un sistema finito qualsiasi di equazioni, condizioni di diversità ($F \neq 0$) e disuguaglianze ($F > 0$) polinomiali in più incognite ammetta soluzioni in un campo reale chiuso R. Il risultato principale (teorema di Tarski) sarà che, dato un tale sistema, si possono determinare in un numero finito di passi un numero finito di sistemi di equazioni, condizioni di diversità e disuguaglianze polinomiali nei coefficienti, in modo che il sistema dato abbia soluzione in R se e solo se i coefficienti soddisfano tutte le relazioni

di almeno uno dei sistemi ottenuti. Per illustrare il tipo di risultato consideriamo una quartica «ridotta» $x^4 + qx^2 + rx + s = 0$, con $q, r, s \in R$. Si può mostrare che essa ha una radice in R se e solo se, indicando con

$$d = 4\left(4s + \frac{q^2}{3}\right)^3 - 27\left(\frac{8}{3}qs - r^2 - \frac{2}{27}q^3\right)^2 \quad \text{e} \quad L = 8qs - 2q^3 - 9r^2$$

il discriminante e l'espressione ausiliaria, si verifica una delle alternative

I. $d < 0$;
II. $d > 0,\ q < 0,\ L > 0$;
III. $d = 0,\ r \neq 0$;
IV. $d = 0,\ r = 0,\ q \leq 0$.

Ciò segue facilmente dall'esercizio 4 del § 5.2 e dal fatto che $d = 0$ se e solo se l'equazione ha radici multiple.

Mostreremo che una simile versione del teorema di Sturm vale per qualsiasi equazione i cui coefficienti dipendano da parametri che assumono valori in un campo reale chiuso. Per questo occorre una versione parametrica dell'algoritmo euclideo per il massimo comune divisore, che ora dedurremo. Partiamo da un anello dei coefficienti della forma $A = K[t_1, \ldots, t_r]$, dove le t_i sono indeterminate e K è $\mathbb{Z}$ oppure uno dei campi $\mathbb{Z}/(p)$, con p primo. Siano $F(t_1, \ldots, t_r; x), G(t_1, \ldots, t_r; x) \in A[x]$, ossia

$$F(t_i; x) = u_n x^n + u_{n-1}x^{n-1} + \cdots + u_0, \qquad G(t_i; x) = v_m x^m + v_{m-1}x^{m-1} + \cdots + v_0$$

con $u_i, v_j \in A$. Supponiamo $G \neq 0$ e consideriamone un «troncamento» $G_k(t_i; x) = v_k x^k + v_{k-1}x^{k-1} + \cdots + v_0$, con $v_k \neq 0$, ottenuto sopprimendo i termini $v_j x^j$ di indice $j > k$. Il suo grado in x è $\deg_x G_k = k$, dove k assume alcuni dei valori fra 0 e m. Applicando l'algoritmo di divisione possiamo scrivere

$$v_k(t_i)^{e_k} F(t_i; x) = Q_k(t_i; x)G_k(t_i; x) - R_k(t_i; x), \tag{8}$$

dove $\deg R_k < k$ e si può prendere come intero non negativo e_k il maggiore fra 0 e $n-k+1$ (p. 129 dell'originale). Abbiamo indicato con $-R_k$ il resto usuale: il segno scelto è più comodo per il teorema di Sturm ed è altrettanto adatto alle altre applicazioni. Per il teorema di Sturm occorre inoltre che e_k sia pari. Fissiamo dunque $e_k = 0$ quando $n < k$ e, negli altri casi, prendiamo il più piccolo intero pari maggiore o uguale a $n - k + 1$.[3] Con questa scelta ben determinata si ottengono Q_k e R_k che soddisfano (8). Considerazioni di grado ne assicurano l'unicità.

Sia ora R un campo contenente K, dunque con sottoanello primo K ($= \mathbb{Z}$ oppure $\mathbb{Z}/(p)$), e sia $(c_1, \ldots, c_r) \in R^{(r)} = R \times \cdots \times R$ (r volte). Manteniamo le notazioni precedenti. O $v_j(c_1, \ldots, c_r) = 0$ per tutti gli indici $0 \leq j \leq m$, nel qual caso $G(c_i; x) = 0$, oppure esiste k con $v_k(c_1, \ldots, c_r) \neq 0$ e $v_j(c_1, \ldots, c_r) = 0$ per $j > k$. Allora $G(c_i; x) = G_k(c_i; x)$ ha grado k in x e coefficienti in R. Per la (8)

$$v_k(c_i)^{e_k} F(c_i; x) = Q_k(c_i; x)G(c_i; x) - R_k(c_i; x).$$

Il fattore $v_k(c_i)^{e_k}$ è non nullo e $\deg R_k(c_i; x) < \deg G(c_i; x)$. Per ottenere il quoziente e il resto della divisione di $F(c_i; x)$ per $G(c_i; x) \neq 0$ basta dunque moltiplicare rispettivamente $Q_k(c_i; x)$ e $-R_k(c_i; x)$ per $v_k(c_i)^{-e_k}$. Tale fattore è positivo se R è reale chiuso.

[3] *Nota alla traduzione.* A p. 317 l'originale stampa $n \leq k$ nel primo caso. Per $n = k$ la divisione su A può richiedere una potenza del coefficiente direttore, come mostra già la divisione di x per $t_1 x + 1$.

Introduciamo i seguenti sistemi di equazioni e condizioni di diversità, definiti da polinomi in A:

$$\begin{aligned} \Gamma_{-\infty} &= \{v_0 = 0,\ v_1 = 0, \dots, v_m = 0\}, \\ \Gamma_k &= \{v_j = 0\ (j > k),\ v_k \neq 0\} \qquad \text{se } v_k(t_1, \dots, t_r) \neq 0. \end{aligned} \tag{9}$$

La famiglia $\gamma = \{\Gamma_{-\infty}, \Gamma_k\}$, con $0 \leq k \leq m$ e $v_k \neq 0$, è un *ricoprimento* di A nel senso seguente: per ogni campo R contenente K, se $\Gamma_j(R)$ indica l'insieme delle r-uple di $R^{(r)}$ che soddisfano le condizioni di Γ_j, allora $R^{(r)} = \bigcup_j \Gamma_j(R)$. Qui $\Gamma_j(R)$ è precisamente l'insieme delle r-uple per cui $\deg_x G(c_i; x) = j$ (con grado $-\infty$ se e solo se il polinomio è nullo). In generale, i termini Γ_j di un ricoprimento sono sistemi finiti di equazioni e di condizioni di diversità i cui primi membri appartengono ad A. Più condizioni $l_1 \neq 0, \dots, l_h \neq 0$ equivalgono all'unica condizione $l \neq 0$, con $l = l_1 \cdots l_h$, perché R non ha divisori dello zero non nulli. Per uniformità, se Γ_j non contiene equazioni aggiungiamo $0 = 0$, e se non contiene condizioni di diversità aggiungiamo $1 \neq 0$. Possiamo così supporre che Γ_j sia costituito da un insieme finito non vuoto di equazioni e da una sola condizione di diversità. Per R reale chiuso, inoltre, le equazioni $d_1 = 0, \dots, d_h = 0$ equivalgono all'unica equazione $\sum d_i^2 = 0$. Se si considerano soltanto campi reali chiusi contenenti K (necessariamente $K = \mathbb{Z}$), ogni Γ_j si può dunque ridurre a un'equazione e a una condizione di diversità.

Riassumiamo. Dati $F, G \in A[x]$, con $G = v_m x^m + \cdots + v_0$, sia γ il ricoprimento definito dai v_j mediante (9). A ogni $k \neq -\infty$ che vi compare corrispondono polinomi $Q_k, R_k \in A[x]$ tali che, per ogni campo R contenente K e ogni $(c_1, \dots, c_r) \in \Gamma_k(R)$, si ha $G(c_i; x) \neq 0$, e il quoziente e il resto della divisione in $R[x]$ di $F(c_i; x)$ per $G(c_i; x)$ differiscono da $Q_k(c_i; x)$ e $-R_k(c_i; x)$ per un fattore non nullo, positivo se R è reale chiuso.

Sia Γ un sistema finito di equazioni e di un'unica condizione di diversità, definito da elementi di A, e sia $\delta = \{\Delta_1, \dots, \Delta_s\}$ un ricoprimento di A. Per $1 \leq j \leq s$ costruiamo $\Gamma^{(j)}$ prendendo tutte le equazioni di Γ e Δ_j e, come condizione di diversità, il prodotto delle due condizioni corrispondenti. Per ogni R si ha allora $\Gamma^{(j)}(R) = \Gamma(R) \cap \Delta_j(R)$. Poiché $\bigcup_{j=1}^s \Delta_j(R) = R^{(r)}$, segue $\Gamma(R) = \bigcup_{j=1}^s \Gamma^{(j)}(R)$. Pertanto, se $\gamma = \{\Gamma_1 = \Gamma, \Gamma_2, \dots, \Gamma_q\}$ ricopre A, lo ricopre anche $\gamma' = \{\Gamma^{(1)}, \dots, \Gamma^{(s)}, \Gamma_2, \dots, \Gamma_q\}$. Chiameremo *raffinamenti* di γ i ricoprimenti ottenuti in questo modo o ripetendo il procedimento un numero finito di volte.

Ricordiamo che, dati $f(x), g(x) \in R[x]$ con $g(x) \neq 0$ e R campo, l'algoritmo euclideo per un massimo comune divisore di f e g costruisce per divisioni successive la successione di polinomi

$$f_0 = f, \quad f_1 = g, \quad f_2, \dots, f_s, \quad f_{s+1} = 0, \tag{10}$$

con $\deg f_{i+1} < \deg f_i$ per $i \geq 1$ e $f_{i-1} = q_i f_i - f_{i+1}$ per opportuni q_i (cfr. esercizio 11, p. 150 dell'originale, e la (3)). Ne segue che $f_s \neq 0$ è un massimo comune divisore di f e g in $R[x]$. Chiameremo la (10) *successione euclidea della coppia* (f, g) quando $g \neq 0$. Conviene estendere la definizione al caso $(f, 0)$, prendendo come successione euclidea $(f, 0, 0)$.

Lemma. *Siano $F, G \in A[x]$, $G \neq 0$. In un numero finito di passi possiamo costruire un ricoprimento $\delta = \{\Delta_1, \dots, \Delta_h\}$ che raffina il ricoprimento (9) definito dai coefficienti di G, e successioni di polinomi $F_{j0} = F, F_{j1}, \dots, F_{js_j} \in A[x]$, $1 \leq j \leq h$, tali che, per ogni campo R contenente K e ogni $(c_1, \dots, c_r) \in \Delta_j(R)$, i termini della successione*

$$F_{j0}(c_i; x), F_{j1}(c_i; x), \dots, F_{js_j}(c_i; x), F_{j,s_j+1}(c_i; x) = 0$$

differiscono da quelli della successione euclidea di $(F(c_i; x), G(c_i; x))$ per fattori non nulli in R. Tali fattori sono positivi se R è reale chiuso.

Dimostrazione. Come sopra, per ogni $k \neq -\infty$ della (9) determiniamo $Q_k, R_k \in A[x]$ dividendo F per $G_k = v_k x^k + \cdots + v_0$ mediante l'algoritmo descritto. Se $R_k = 0$, i polinomi $F_{k0} = F$,

$F_{k1} = G_k$, $F_{k2} = 0$ soddisfano la condizione. Supponiamo dunque $R_k \neq 0$. La somma dei gradi di G_k e R_k è minore della somma dei gradi di F e G nei passi in cui occorre proseguire dopo la prima divisione. Per induzione sulla somma dei gradi applichiamo il risultato alla coppia formata dal divisore e dall'opposto del resto. Si ottengono così un ricoprimento $\delta_k = \{\Delta_{k1}, \dots, \Delta_{kh_k}\}$ e successioni di polinomi che iniziano con G_k e proseguono l'algoritmo. Raffiniamo γ sostituendo ogni Γ_k, $k \neq -\infty$, con i sistemi $\Gamma_k^{(1)}, \dots, \Gamma_k^{(h_k)}$: le equazioni di $\Gamma_k^{(l)}$ sono quelle di Γ_k e Δ_{kl}, e la sua condizione di diversità è il prodotto delle due condizioni corrispondenti. Sia δ il ricoprimento così ottenuto per tutti i k. A $\Gamma_k^{(l)}$ associamo la successione che si ottiene anteponendo F a quella già costruita per il passo successivo; al termine $\Gamma_{-\infty}$ associamo $(F, 0, 0)$. Si verifica subito che δ, con queste successioni, soddisfa le condizioni richieste. □

A titolo di esempio consideriamo la cubica ridotta $F(p, q; x) = x^3 + px + q$, in cui p, q sono indeterminate, e la derivata $G(p, q; x) = F'(p, q; x) = 3x^2 + p$. Prendiamo $K = \mathbb{Z}$. Si ha $v_2 = 3$, $v_1 = 0$, $v_0 = p$, dunque $\Gamma_{-\infty} = \{3 = 0,\ 0 = 0,\ p = 0;\ 1 \neq 0\}$; le prime tre sono le equazioni, l'ultima è la condizione di diversità banale. Inoltre $\Gamma_2 = \{0 = 0;\ 3 \neq 0\}$ e $\Gamma_0 = \{3 = 0;\ p \neq 0\}$. Per ogni R considerato, $\Gamma_{-\infty}(R) = \Gamma_0(R) = \varnothing$, perciò si può prendere $\gamma = \{\Gamma = \Gamma_2\}$. Il primo passo dell'algoritmo dà, con la convenzione dei resti cambiati di segno, $-(6px + 9q)$; ripetiamo il procedimento con $3x^2 + p$ e $-(6px + 9q)$. Lasciamo al lettore la verifica che si ottiene $\delta = \{\Delta_1, \Delta_2, \Delta_3, \Delta_4\}$, dove

$$\begin{aligned}
\Delta_1 &= \{0 = 0;\ p(27q^2 + 4p^3) \neq 0\},\\
\Delta_2 &= \{27q^2 + 4p^3 = 0;\ p \neq 0\},\\
\Delta_3 &= \{p = 0;\ q \neq 0\},\\
\Delta_4 &= \{p = 0,\ q = 0;\ 1 \neq 0\}.
\end{aligned}$$

Le successioni corrispondenti sono

$$\begin{aligned}
&\text{I.} \quad x^3 + px + q,\ 3x^2 + p,\ -(6px + 9q),\ -9(27q^2 + 4p^3),\ 0;\\
&\text{II.} \quad x^3 + px + q,\ 3x^2 + p,\ -(6px + 9q),\ 0;\\
&\text{III.} \quad x^3 + px + q,\ 3x^2 + p,\ -9q,\ 0;\\
&\text{IV.} \quad x^3 + px + q,\ 3x^2 + p,\ 0.
\end{aligned}$$

Possiamo ora dare la versione parametrica del teorema di Sturm.

Teorema 5.5. *Sia $F(t_i; x) = u_n x^n + u_{n-1}x^{n-1} + \cdots + u_0 \in A[x]$, dove $u_j = u_j(t_1, \dots, t_r) \in A = \mathbb{Z}[t_1, \dots, t_r]$ e t_i, x sono indeterminate. In un numero finito di passi si può determinare una famiglia finita $\{\Gamma_1, \dots, \Gamma_s\}$, in cui ciascun Γ_k è un insieme finito di relazioni polinomiali $C = 0$, $C > 0$, $C \neq 0$, con $C \in A$, tale che per qualsiasi campo reale chiuso R e qualsiasi scelta di $c_1, \dots, c_r \in R$ l'equazione $F(c_i; x) = 0$ ha una radice in R se e solo se per $t_i = c_i$ valgono tutte le relazioni di almeno uno dei Γ_k.*

Dimostrazione. Poniamo $G(t_i; x) = F'(t_i; x) = nu_n x^{n-1} + (n-1)u_{n-1}x^{n-2} + \cdots + u_1$. Possiamo supporre $G \neq 0$: altrimenti $F(t_i; x) = u_0(t_1, \dots, t_r)$ e l'asserto è banale. Applicando il lemma otteniamo con un procedimento finito il ricoprimento $\delta = \{\Delta_1, \dots, \Delta_h\}$ e le corrispondenti successioni $F_{j0} = F, F_{j1}, \dots, F_{js_j} \in A[x]$, per $1 \leq j \leq h$. Se R è reale chiuso e $(c_1, \dots, c_r) \in \Delta_j(R)$, i polinomi specializzati differiscono per fattori positivi dai termini della successione standard di Sturm di $F(c_i; x)$. Per semplicità fissiamo uno dei Δ_j, scriviamolo Δ e indichiamo la successione corrispondente con $F_0 = F, F_1, \dots, F_s$. Poiché δ raffina il ricoprimento associato a G, su $\Delta(R)$ si ha o $u_n(c_i) = \cdots = u_1(c_i) = 0$, oppure un m, $1 \leq m \leq n$, con $u_m(c_i) \neq 0$ e $u_j(c_i) = 0$ per $j > m$. Nel primo caso l'equazione ha una radice se e solo se $u_0(c_i) = 0$. Nel secondo, tutte le radici reali di $F(c_i; x)$ sono in $(-\eta, \eta)$, dove

$$\eta = (m + 1) + \sum_{j=0}^{m-1} u_j(c_i)^2 u_m(c_i)^{-2}. \tag{11}$$

Poiché le due successioni differiscono per fattori positivi, il teorema di Sturm dice che una radice esiste se e solo se il numero di variazioni nella successione $F_0(c_i;-\eta),\dots,F_s(c_i;-\eta)$ supera quello nella successione $F_0(c_i;\eta),\dots,F_s(c_i;\eta)$.[4] Per esprimere tale condizione mediante polinomi nei c_i, associamo a ogni F_k i polinomi

$$g_k(t_1,\dots,t_r)=u_m^{2n_k}F_k\left(t_i;\ m+1+\sum_{j=0}^{m-1}u_j^2u_m^{-2}\right),$$
$$h_k(t_1,\dots,t_r)=u_m^{2n_k}F_k\left(t_i;\ -(m+1)-\sum_{j=0}^{m-1}u_j^2u_m^{-2}\right),$$

dove $u_j=u_j(t_1,\dots,t_r)$ e n_k è il grado di F_k in x. Abbiamo $g_k,h_k\in A$ e, dopo specializzazione, essi differiscono rispettivamente da $F_k(c_i;\eta)$ e $F_k(c_i;-\eta)$ per fattori positivi. Per $(c_1,\dots,c_r)\in\Delta(R)$ esiste pertanto una radice se e solo se la successione degli $h_k(c_i)$ ha più variazioni di segno della successione dei $g_k(c_i)$. Consideriamo tutti i sistemi possibili di relazioni

$$\begin{array}{lll} g_0\gtrless 0, & g_k\begin{matrix}\geqslant\\ \leqslant\end{matrix}0, & g_s\gtrless 0,\\ h_0\gtrless 0, & h_k\begin{matrix}\geqslant\\ \leqslant\end{matrix}0, & h_s\gtrless 0,\end{array}\qquad 1\leq k\leq s-1. \tag{12}$$

Accoppiamo le relazioni sui g con quelle sugli h in tutti i modi per i quali il numero di variazioni dei segni prescritti agli h supera quello relativo ai g. Per $(c_1,\dots,c_r)\in\Delta(R)$, l'equazione ha soluzione se e solo se i c_i soddisfano uno dei sistemi così accoppiati. Aggiungendo a ciascuno le relazioni di Δ otteniamo uno dei sistemi Γ richiesti. Procedendo per tutte le coppie e tutti i Δ, si ottiene una famiglia finita di sistemi Γ con le proprietà enunciate. □

Lo stesso metodo permette di ottenere un risultato analogo per l'esistenza di una radice in un intervallo dato $(-c,c)$, con $c>0$. Si può inoltre sostituire $A=\mathbb{Z}[t_1,\dots,t_r]$ con $F[t_1,\dots,t_r]$, dove F è un sottocampo di un campo reale chiuso.

5.4 Procedimenti di eliminazione. Risultanti

Prima di estendere il teorema 5.5 ai sistemi di equazioni, condizioni di diversità e disuguaglianze in più incognite, conviene affrontare un problema più semplice: come stabilire se un sistema di equazioni polinomiali e condizioni di diversità ammetta soluzione in qualche estensione di un campo dato. Il teorema fondamentale che vogliamo dimostrare è il seguente.

Teorema 5.6. *Sia $K=\mathbb{Z}$ oppure $\mathbb{Z}/(p)$, con p primo, e siano $A=K[t_1,\dots,t_r]$, $B=A[x_1,\dots,x_n]$, dove le t e le x sono indeterminate. Dato $\Gamma=\{F_1,\dots,F_m;G\}\subset B$, si può determinare in un numero finito di passi una famiglia finita $\{\Gamma_1,\dots,\Gamma_s\}$, dove $\Gamma_j=\{f_{j1},\dots,f_{jm_j};g_j\}\subset A$, con la seguente proprietà. Per ogni campo F contenente K e ogni $(c_1,\dots,c_r)\in F^{(r)}$, il sistema di equazioni e della condizione di diversità*

$$\Gamma(c_i):\quad F_1(c_i;x_1,\dots,x_n)=\cdots=F_m(c_i;x_1,\dots,x_n)=0,\quad G(c_i;x_1,\dots,x_n)\neq 0 \tag{13}$$

ammette soluzione rispetto alle x in qualche estensione E/F se e solo se i c_i soddisfano uno dei sistemi

$$\Gamma_j(c_i):\quad f_{j1}(c_i)=\cdots=f_{jm_j}(c_i)=0,\quad g_j(c_i)\neq 0,\qquad 1\leq j\leq s. \tag{14}$$

Inoltre, se è soddisfatto uno di questi sistemi, esiste una soluzione di (13) in un'estensione algebrica E/F.

[4] *Nota alla traduzione.* A p. 321 dell'originale i due estremi del confronto sono scambiati. Qui si applica la formula $V_{-\eta}-V_\eta$ dimostrata nel paragrafo precedente.

Premettiamo alla dimostrazione un lemma dovuto a Tarski.

Lemma. *Siano $f(x), g(x)$ polinomi non nulli in $F[x]$, con F campo, e sia $h = \deg f$. Se $f \mid g^h$, non esiste in alcuna estensione E/F un elemento a che soddisfi $f(a) = 0$, $g(a) \neq 0$. Se invece $f \nmid g^h$, un tale elemento a esiste in un'estensione algebrica E/F.*

Dimostrazione. La prima affermazione è chiara. Supponiamo $f \nmid g^h$. Esiste un fattore irriducibile $p(x)$ di f che non divide g. Altrimenti, detti $p_1, \dots, p_m$ i fattori irriducibili distinti di f, avremmo $p_1 \cdots p_m \mid g$, quindi $(p_1 \cdots p_m)^h \mid g^h$. Ma $f \mid (p_1 \cdots p_m)^h$ e ne seguirebbe $f \mid g^h$, assurdo. Scegliamo quindi un tale p e poniamo $E = F[x]/(p(x))$. È un'estensione algebrica di F, e l'elemento $a = x + (p(x))$ soddisfa $f(a) = 0$, $g(a) \neq 0$. □

Dimostrazione del teorema 5.6. Consideriamo dapprima $n = 1$ e procediamo per induzione sulla somma dei gradi in x degli F_i e di G, ponendo qui $\deg 0 = 0$. Se la somma è nulla non c'è nulla da dimostrare. Quando almeno uno dei polinomi ha grado positivo, si applicano le riduzioni seguenti.

Caso I. $\deg F_i > 0$ per $i = 1, 2$. Siano $b_0 x^h$ e $d_0 x^k$ i termini direttori di F_1 e F_2 rispettivamente: $F_1 = b_0 x^h + b_1 x^{h-1} + \cdots$, $F_2 = d_0 x^k + d_1 x^{k-1} + \cdots$, con $b_0, d_0 \neq 0$. Supponiamo $h \geq k$. Definiamo

$$\Gamma' = \{d_0, F_1, F_2' = F_2 - d_0 x^k, F_3, \dots, F_m; G\},$$
$$\Gamma'' = \{F_1'' = d_0 F_1 - b_0 x^{h-k} F_2, F_2, \dots, F_m; d_0 G\}.$$

Se $(c_1, \dots, c_r) \in F^{(r)}$ soddisfa $d_0(c_i) = 0$, allora $\Gamma(c_i)$ è risolubile in un'estensione E/F se e solo se lo è $\Gamma'(c_i)$ in E. Se invece $d_0(c_i) \neq 0$, l'equivalenza vale con $\Gamma''(c_i)$. La somma dei gradi in x diminuisce sia in Γ' sia in Γ''. Il risultato nel caso I segue dunque dall'ipotesi induttiva.

Caso II. $\deg F_1 > 0$, $\deg G > 0$ e $\deg F_i = 0$ per $i > 1$. Siano $b_0 x^h$ e $d_0 x^k$ i termini direttori di F_1 e G. Mediante divisione, moltiplicando G^h per una potenza sufficiente b_0^N, otteniamo $Q, R \in B$ tali che

$$b_0^N G^h = Q F_1 + R, \qquad R = r_0 x^{h-1} + r_1 x^{h-2} + \cdots + r_{h-1}, \quad r_i \in A.$$

Definiamo

$$\Gamma' = \{b_0, F_1' = F_1 - b_0 x^h, F_2, \dots, F_m; G\}, \qquad \Gamma_i = \{F_2, \dots, F_m; b_0 r_i\} \subset A, \quad 0 \leq i < h.$$

Se $b_0(c_i) = 0$, il sistema $\Gamma(c_i)$ ha una soluzione in E/F se e solo se $\Gamma'(c_i)$ ne ha una in E. Supponiamo ora $b_0(c_i) \neq 0$. Allora $F_1(c_i; x) \nmid G(c_i; x)^h$ se e solo se $R(c_i; x) \neq 0$, ossia se e solo se almeno uno degli $r_j(c_i)$ è diverso da zero. Per il lemma, $F_1(c_i; x) = 0$, $G(c_i; x) \neq 0$ ammettono una soluzione in un'estensione se e solo se $R(c_i; x) \neq 0$; in tal caso una soluzione esiste in un'estensione algebrica di F. Se dunque $b_0(c_i) \neq 0$, il sistema $\Gamma(c_i)$ è risolubile in un'estensione se e solo se vale uno dei sistemi $\Gamma_j(c_i)$, e allora è risolubile in un'estensione algebrica. Poiché la somma dei gradi in Γ' e in ogni Γ_j è inferiore a quella in Γ, si conclude per induzione.

Caso III. $\deg F_1 > 0$, $\deg F_i = \deg G = 0$ per $i > 1$. Sia $b_0 x^h$ il termine direttore di F_1. Definiamo

$$\Gamma' = \{b_0, F_1' = F_1 - b_0 x^h, F_2, \dots, F_m; G\}, \qquad \Gamma'' = \{F_2, \dots, F_m; G b_0\}.$$

Se $b_0(c_i) = 0$, il sistema $\Gamma(c_i)$ ha una soluzione in E/F se e solo se $\Gamma'(c_i)$ ne ha una nella stessa estensione. Se $b_0(c_i) \neq 0$, il polinomio $F_1(c_i; x)$ ha una radice in qualche estensione algebrica E/F, e in tale estensione $\Gamma(c_i)$ ha soluzione se sono soddisfatte le condizioni $\Gamma''(c_i)$. Il grado di F_1' è minore di quello di F_1, e il caso segue per induzione.

Caso IV. $\deg F_i = 0$ per $1 \leq i \leq m$ e $\deg G > 0$. Sia $d_0 x^k$ il termine direttore di G e definiamo

$$\Gamma' = \{d_0, F_1, \dots, F_m; G' = G - d_0 x^k\}, \qquad \Gamma'' = \{F_1, \dots, F_m; d_0\}.$$

Se $d_0(c_i) = 0$, la risolubilità di $\Gamma(c_i)$ in E/F equivale a quella di $\Gamma'(c_i)$ in E. Se $d_0(c_i) \neq 0$, la condizione $G(c_i; x) \neq 0$ è soddisfatta da qualche x in ogni campo con più di k elementi che contenga i coefficienti. Esiste dunque una soluzione in un'estensione algebrica E/F quando sono soddisfatte le condizioni $\Gamma''(c_i)$. Poiché il grado di G' è minore di quello di G, il risultato segue anche qui.

Questi casi esauriscono le possibilità in cui la somma dei gradi in x degli F_i e di G sia positiva, e il teorema è provato per $n = 1$. Procediamo per induzione su n, supponendo $n > 1$. Trattiamo $x_1, \dots, x_{n-1}$ come ulteriori parametri t. Il risultato appena provato fornisce in un numero finito di passi sistemi $\Lambda_k = \{F_{k1}, \dots, F_{kh_k}; G_k\}$, $1 \leq k \leq u$, con $F_{kj}, G_k \in K[t_1, \dots, t_r, x_1, \dots, x_{n-1}]$, che hanno le due proprietà seguenti. (1) Se F è un campo contenente K e $(c_1, \dots, c_{r+n-1}) \in F^{(r+n-1)}$ soddisfa uno dei sistemi Λ_k nel senso di (14), allora $\Gamma(c_1, \dots, c_{r+n-1})$ è risolubile in x_n in un'estensione algebrica E/F. (2) Se tale sistema è risolubile in x_n in qualche estensione E/F, vale uno dei sistemi $\Lambda_k(c_1, \dots, c_{r+n-1})$.

Per induzione sul numero delle x, a ogni Λ_k associamo una famiglia $\{\Gamma_{kl} \mid 1 \leq l \leq u_k\}$ di sistemi finiti di polinomi in A che soddisfa la conclusione del teorema per Λ_k nelle $n-1$ incognite. Affermiamo che la famiglia di tutti i Γ_{kl}, con $1 \leq k \leq u$ e $1 \leq l \leq u_k$, ha la proprietà richiesta per Γ. Supponiamo dapprima che $(c_1, \dots, c_r)$ soddisfi Γ_{kl} per qualche k, l. Allora esiste un'estensione algebrica E/F nella quale $\Lambda_k(c_1, \dots, c_r)$ ha soluzione in $x_1, \dots, x_{n-1}$. Detti $(c_{r+1}, \dots, c_{r+n-1})$ i valori di una soluzione, per (1) il sistema $\Gamma(c_1, \dots, c_r)$ ammette soluzione in un'estensione algebrica E' di E, anch'essa algebrica su F. Viceversa, sia $(c_{r+1}, \dots, c_{r+n-1}, c)$ una soluzione di $\Gamma(c_1, \dots, c_r)$ in qualche E/F. Allora $\Gamma(c_1, \dots, c_{r+n-1})$ è risolubile in x_n in E. Per (2) vale uno dei sistemi $\Lambda_k(c_1, \dots, c_{r+n-1})$, e questo implica che per qualche l è soddisfatto $\Gamma_{kl}(c_1, \dots, c_r)$. La dimostrazione è completa. □

Un secondo metodo di eliminazione, più classico, si fonda sui risultanti. Ne daremo ora il risultato principale e ne indicheremo alcune estensioni negli esercizi. Cerchiamo un criterio affinché due polinomi abbiano un fattore comune di grado positivo. Consideriamo $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$ e $g(x) = b_m x^m + b_{m-1} x^{m-1} + \cdots + b_0$ in $F[x]$, dove F è un campo. Supponiamo $m, n > 0$, ammettendo però $a_n = 0$ o $b_m = 0$. Vale il risultato seguente.

Teorema 5.7. *Siano f, g come sopra e si ponga*

$$\text{(15)} \qquad \mathrm{Res}(f,g) = \begin{vmatrix} a_n & a_{n-1} & \cdots & a_0 & 0 & \cdots & & 0 \\ 0 & a_n & a_{n-1} & \cdots & a_0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & & a_0 \\ b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & & 0 \\ 0 & b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & b_m & b_{m-1} & \cdots & & b_0 \end{vmatrix}.$$

Il primo blocco consta di m righe con i coefficienti di f traslati ogni volta di un posto a destra, il secondo di n righe analoghe per g. Allora $\mathrm{Res}(f,g) = 0$ *se e solo se $a_n = b_m = 0$ oppure $f(x)$ e $g(x)$ hanno un fattore comune di grado positivo in $F[x]$.*

Dimostrazione. Se $a_n = b_m = 0$, la prima colonna del determinante è nulla e quindi $\mathrm{Res}(f,g) = 0$. Supponiamo ora che f, g abbiano un fattore comune h di grado positivo e che almeno uno tra a_n, b_m sia diverso da zero. Si ha $f = f_1 h$, $g = g_1 h$, e $f_1 \neq 0$ se $a_n \neq 0$, $g_1 \neq 0$ se $b_m \neq 0$. Per simmetria possiamo supporre $a_n \neq 0$, $f_1 \neq 0$. Se $\deg h = r$, allora $\deg f_1 = n - r$. Se $g = 0$, anche $g_1 = 0$; altrimenti la relazione $f g_1 = g f_1$ dà $\deg g_1 \leq m - r$. Possiamo dunque scrivere $f_1 = -c_{n-1} x^{n-1} - \cdots - c_0$, $g_1 = d_{m-1} x^{m-1} + \cdots + d_0$, con qualche $c_i \neq 0$, e si ha la relazione

$$\text{(16)} \quad (a_n x^n + \cdots + a_0)(d_{m-1} x^{m-1} + \cdots + d_0) + (b_m x^m + \cdots + b_0)(c_{n-1} x^{n-1} + \cdots + c_0) = 0.$$

Uguagliando i coefficienti di $x^{m+n-1}, x^{m+n-2}, \ldots, 1$ otteniamo

$$\begin{aligned} a_n d_{m-1} + b_m c_{n-1} &= 0, \\ a_n d_{m-2} + a_{n-1} d_{m-1} + b_m c_{n-2} + b_{m-1} c_{n-1} &= 0, \\ &\vdots \\ a_0 d_0 + b_0 c_0 &= 0. \end{aligned} \tag{17}$$

Consideriamo il sistema lineare con le incognite nell'ordine $d_{m-1}, \ldots, d_0, c_{n-1}, \ldots, c_0$. Poiché la soluzione non è nulla, il determinante dei coefficienti è nullo. Esso è $\mathrm{Res}(f, g)$, dunque $\mathrm{Res}(f, g) = 0$. Viceversa, se il risultante è nullo, ripercorrendo (17) e (16) a ritroso troviamo f_1, g_1 non entrambi nulli, di gradi rispettivamente al più $n-1$ e $m-1$, tali che $fg_1 = gf_1$. Supponiamo $f_1 \neq 0$. Se $g_1 = 0$, allora $g = 0$ e $b_m = 0$: o $a_n = 0$, oppure f è un fattore comune di grado positivo. Se $g_1 \neq 0$ e $g = 0$, la stessa argomentazione dà l'alternativa richiesta. Supponiamo infine $g_1, g \neq 0$. Dalle relazioni precedenti segue $f \neq 0$. O $a_n = b_m = 0$, oppure possiamo supporre per simmetria $a_n \neq 0$, cosicché $\deg f = n$. Poiché $\deg f_1 \leq n-1$, dall'identità $fg_1 = gf_1$ e dalla fattorizzazione dei quattro polinomi in fattori irriducibili si deduce che f e g hanno un fattore comune di grado positivo. □

Chiameremo $\mathrm{Res}(f, g)$ il *risultante* di f e g (rispetto a x). Se almeno uno dei due coefficienti direttori è non nullo, l'annullarsi del risultante è una condizione polinomiale a coefficienti interi negli a_i, b_j, equivalente all'esistenza di un fattore comune di grado positivo.

Esercizi

1. Se $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ e $f'(x) = nx^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1$, dimostrare che $\mathrm{Res}(f, f') = (-1)^{n(n-1)/2} d$, dove d è il discriminante di f (cfr. § 4.8, p. 258).

2. Servirsi del teorema sui risultanti per dimostrare il teorema 5.6 nel caso di due equazioni $F_1(t_1, \ldots, t_r; x) = F_2(t_1, \ldots, t_r; x) = 0$ e $G = 1$.

3. Siano $f_1, \ldots, f_m \in F[x]$, con $f_i(x) = a_{n_i,i}x^{n_i} + a_{n_i-1,i}x^{n_i-1} + \cdots + a_{0i}$. Si scelga un intero $n \geq n_i$ per ogni i e si ponga $g_i(x) = x^{n-n_i} f_i(x)$, $g_{m+i}(x) = (x-1)^{n-n_i} f_i(x)$, per $1 \leq i \leq m$. Mostrare che gli f_i hanno un fattore comune di grado positivo se e solo se lo hanno i g_j, $1 \leq j \leq 2m$. Aggiungere $4m$ indeterminate $u_1, \ldots, u_{2m}, v_1, \ldots, v_{2m}$ e porre $E = F(u_1, \ldots, u_{2m}, v_1, \ldots, v_{2m})$. Siano $u(x) = \sum_{j=1}^{2m} u_j g_j(x)$ e $v(x) = \sum_{j=1}^{2m} v_j g_j(x)$. Formare $\mathrm{Res}(u(x), v(x)) \in F[u_1, \ldots, u_{2m}, v_1, \ldots, v_{2m}]$. Dimostrare che tale risultante è nullo se e solo se tutti gli $a_{n_i,i}$ sono nulli oppure gli $f_1, \ldots, f_m$ hanno un fattore comune di grado positivo.

4. Mostrare che il sistema (13) è risolubile se e solo se lo è il seguente sistema in $x_1, \ldots, x_{n+1}$:

$$F_1(c_i; x_1, \ldots, x_n) = \cdots = F_m(c_i; x_1, \ldots, x_n) = x_{n+1}G(c_i; x_1, \ldots, x_n) - 1 = 0.$$

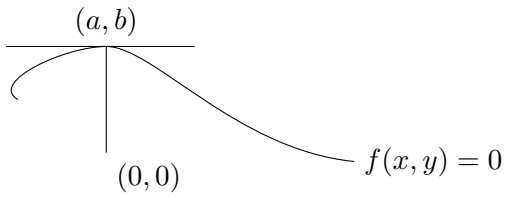
   Osservare che in questo modo si eliminano le condizioni di diversità.

5.5 Un metodo di decisione per una curva algebrica

In questo paragrafo daremo un metodo, dovuto ad A. Seidenberg, per decidere se una data equazione $f(x, y) = 0$, con $f(x, y) \in R[x, y]$, abbia una soluzione in $R^{(2)}$. In altri termini, la curva algebrica $f(x, y) = 0$ ha punti reali? Per esempio, $x^2 + y^2 = 0$ ne ha, mentre $x^2 + y^2 = -1$ non ne ha. Il metodo di Seidenberg si fonda su un'osservazione semplice. Se $f(x, y) = 0$ ha un

punto reale, essa ha anche un punto reale (a, b) di minima distanza dall'origine. Si può allora provare che (a, b) è anche una soluzione di $g(x, y) = x(\partial f/\partial y) - y(\partial f/\partial x) = 0$. Ne segue che a è una radice del polinomio $h(x)$, risultante rispetto a y di $f(x, y)$ e $g(x, y)$. Dunque l'esistenza di una soluzione di $f(x, y) = 0$ in $R^{(2)}$ implica l'esistenza di una radice di $h(x) = 0$ in R, cosa che si può decidere col metodo di Sturm. Vedremo che il ragionamento si può invertire purché si sostituisca l'origine con un punto opportuno e si prendano al posto di x, y due altri generatori x', y', ossia purché si effettui un opportuno cambiamento affine di coordinate nello spazio vettoriale $R^{(2)}$. Otterremo così un algoritmo per verificare la risolubilità di $f(x, y) = 0$ in $R^{(2)}$.

I primi due passi indicati si compiono facilmente se $R = \mathbb{R}$, il campo dei numeri reali. Se (x_0, y_0) è una soluzione, consideriamo l'insieme dei punti (x, y) di $\mathbb{R}^{(2)}$ tali che $f(x, y) = 0$ e $x^2 + y^2 \leq x_0^2 + y_0^2$. È un sottoinsieme chiuso e limitato di $\mathbb{R}^{(2)}$, dunque compatto, e contiene pertanto un punto (a, b) di minima distanza dall'origine. Dal calcolo differenziale sappiamo che o $(\partial f/\partial x)_{(a,b)} = 0 = (\partial f/\partial y)_{(a,b)}$, oppure la retta che congiunge l'origine con (a, b) è normale alla curva data.



In quest'ultimo caso (a, b) è proporzionale alla normale alla curva, $((\partial f/\partial x)_{(a,b)}, (\partial f/\partial y)_{(a,b)})$. In entrambi i casi $a(\partial f/\partial y)_{(a,b)} - b(\partial f/\partial x)_{(a,b)} = 0$, e quindi (a, b) è radice di $g(x, y) = 0$.

Dimostriamo ora questi risultati, in due lemmi, per ogni campo reale chiuso.

Lemma 1. *Sia $f(x, y) \in R[x, y]$, con x, y indeterminate ed R un campo reale chiuso. Se $f(x, y) = 0$ ammette una soluzione in R, essa ammette una soluzione (a, b) di minima distanza dall'origine.*

Dimostrazione. Consideriamo nello spazio $R^{(2)}$ le intersezioni del luogo C di equazione $f(x, y) = 0$ con le circonferenze $x^2 + y^2 = c^2$, $c \geq 0$. Per ipotesi almeno una di queste intersezioni è non vuota. Dobbiamo provare che l'insieme S dei $c \geq 0$ per cui C incontra la circonferenza di raggio c ha minimo. Consideriamo $f(x, y)$ e $x^2 + y^2 - t^2$ in $R[x, y, t]$, dove x, y, t sono indeterminate, e formiamo il loro risultante rispetto a y, ossia trattandoli come polinomi in y. Questo risultante è un elemento $g(t, x) \in R[t, x]$. Affermiamo che S coincide con l'insieme dei $c \geq 0$ per cui $g(c, x)$ ha una radice nell'intervallo $[-c, c]$. Infatti, se $c \in S$ e (a, b) è un punto comune alla circonferenza e a C, $f(a, y)$ e $y^2 + a^2 - c^2$ hanno il fattore comune $y - b$. Pertanto $g(c, a) = 0$ e $g(c, x)$ ha la radice $a \in R$. Inoltre $-c \leq a \leq c$. Viceversa, supponiamo che $g(c, x)$ abbia, per un $c \geq 0$, una radice $a \in R$ con $-c \leq a \leq c$. Poiché il coefficiente direttivo in y di $y^2 + a^2 - c^2$ è 1, dal teorema 5.7 segue che questo polinomio e $f(a, y)$ hanno un fattore comune non costante in $R[y]$. Ma i fattori del primo sono $y \pm b$, con $b = (c^2 - a^2)^{1/2}$. Di conseguenza (a, b) oppure $(a, -b)$ è un punto di intersezione di C con $x^2 + y^2 = c^2$. Dunque $c \in S$.

Indichiamo con S' il sottoinsieme di S formato dai c per cui $g(c, c)g(c, -c) \neq 0$. Per tali c una radice in $[-c, c]$ appartiene all'intervallo aperto $(-c, c)$. Dalle osservazioni successive al teorema 5.5 segue che S' è unione finita di insiemi definiti da sistemi finiti di equazioni $p(c) = 0$, condizioni di non annullamento $q(c) \neq 0$ e disuguaglianze $r(c) > 0$, con $p, q, r \in R[t]$. Esaminando in R i luoghi $p(t) = 0$, $q(t) \neq 0$, $r(t) > 0$, si vede che l'insieme dei punti c che soddisfano queste condizioni e $c \geq 0$ è un'unione, eventualmente vuota, di un numero finito di intervalli. Questi

possono essere aperti, chiusi o semiaperti, ridursi a punti oppure estendersi fino a $+\infty$. Anche S' ha questa forma. Poiché l'insieme dei $c \geq 0$ per cui $g(c,c)g(c,-c) = 0$ è finito oppure comprende tutti i $c \geq 0$, anche S ha la medesima struttura. Basterà provare che il complementare di S nella semiretta degli elementi non negativi è aperto in tale semiretta. Seguirà che S è unione finita di intervalli chiusi e ha un minimo.

Siano dunque $d \geq 0$, $d \notin S$. Il sistema $g(d,x) = 0$, $-d \leq x \leq d$, non ha soluzione in R. Scriviamo

$$g(t,x) = g_0(x) + g_1(x)(t-d) + \cdots + g_m(x)(t-d)^m, \qquad g_i \in R[x].$$

Si ha $g_0(u) \neq 0$ per $-d \leq u \leq d$. Esiste pertanto $d' > d$ tale che $g_0(u) \neq 0$ per $-d' \leq u \leq d'$. Esistono allora $b, B > 0$ tali che $|g_0(u)| \geq b$ e $|g_i(u)| \leq B$ per $i \geq 1$ e ogni $u \in [-d', d']$ (esercizio 7, p. 311 dell'originale). Se $|c-d| < 1/2$, $|c-d| < b/(4B)$ e $u \in [-d', d']$, abbiamo

$$\begin{aligned} |g(c,u)| &\geq |g_0(u)| - |g_1(u)(c-d) + \cdots + g_m(u)(c-d)^m| \\ &\geq b - 2B|c-d| > b - b/2 = b/2. \end{aligned}$$

Ogni $c \geq 0$ con $|c-d| < 1/2$, $|c-d| < b/(4B)$ e $c \leq d'$ appartiene perciò al complementare di S. Ogni punto d di questo complementare vi possiede dunque un intorno relativo. Ciò dimostra il lemma. □

Come nel caso classico del campo dei numeri reali, diciamo *semplice* un punto (a,b) di $C\colon f(x,y) = 0$ per cui

$$((\partial f/\partial x)_{(a,b)}, (\partial f/\partial y)_{(a,b)}) \neq (0,0).$$

Il membro di sinistra è allora il vettore normale a C in (a,b) e la *retta tangente* in questo punto ha equazione

$$(\partial f/\partial x)_{(a,b)}(x-a) + (\partial f/\partial y)_{(a,b)}(y-b) = 0.$$

Sia ora (a,b) un punto di C di minima distanza dall'origine. Vogliamo mostrare che $b(\partial f/\partial x)_{(a,b)} - a(\partial f/\partial y)_{(a,b)} = 0$. Se $(a,b) = (0,0)$ o se il punto non è semplice, l'uguaglianza è evidente. Negli altri casi essa afferma che il vettore che congiunge $(0,0)$ a (a,b) e il vettore normale a C in (a,b) sono linearmente dipendenti. Equivalentemente, la curva C e la circonferenza di centro l'origine e raggio $(a^2+b^2)^{1/2}$ hanno la stessa tangente in (a,b). Se così non fosse, la tangente a C in (a,b) conterrebbe punti interni al cerchio, mentre C non ne conterrebbe, per la scelta di (a,b). Che ciò sia impossibile, e quindi che l'uguaglianza desiderata valga, risulta dal lemma seguente.

Lemma 2. *Sia p un punto comune, in $R^{(2)}$, a una circonferenza e a una curva $C\colon f(x,y) = 0$. Se p è semplice per C e la tangente a C in p contiene punti interni al cerchio, anche C contiene punti interni al cerchio.*

Dimostrazione. Scegliendo opportunamente gli assi possiamo prendere $p = (0,0)$ e come asse x la tangente a C in p. Allora $f(0,0) = 0$, $(\partial f/\partial x)_{(0,0)} = 0$ e possiamo supporre $(\partial f/\partial y)_{(0,0)} = 1$. Il centro del cerchio non appartiene all'asse y. Indichiamolo con (a,b), con $a \neq 0$. Indicando al pedice 0 la valutazione nell'origine si ha

$$f(x,y) = f(0,0) + f_{x,0}x + f_{y,0}y + \frac{1}{2!}(f_{xx,0}x^2 + 2f_{xy,0}xy + f_{yy,0}y^2) + \cdots .$$

Tenuto conto delle condizioni su f, possiamo scrivere $f(x,y) = y(1+h(x,y)) + g(x)$, dove $h(0,0) = 0$ e x^2 divide $g(x)$. Scegliamo $\delta > 0$ tale che $|h(x,y)| \leq 1/2$ se $|x|, |y| \leq \delta$. Allora $1/2 \leq 1 + h(x,y) \leq 3/2$; per $|x| \leq \delta$, $\delta(1+h(x,\delta))$ è compreso fra $\delta/2$ e $3\delta/2$, mentre $-\delta(1+h(x,-\delta))$ è compreso fra $-3\delta/2$ e $-\delta/2$. Poiché $g(0) = 0$, esiste $0 < \delta' < \delta$ tale che, per $|x| \leq \delta'$,

$f(x,\delta) = \delta(1+h(x,\delta)) + g(x) > 0$ e $f(x,-\delta) < 0$. Per ogni x_0 con $|x_0| \le \delta'$ esiste pertanto un $y_0 \in [-\delta,\delta]$ con $f(x_0,y_0) = 0$. Dunque $y_0 = -g(x_0)/(1+h(x_0,y_0))$ e

$$\begin{aligned}(a-x_0)^2 + (b-y_0)^2 &= (a-x_0)^2 + \left(b + \frac{g(x_0)}{1+h(x_0,y_0)}\right)^2 \\ &= a^2 + b^2 - 2ax_0 + x_0^2 + \frac{2bg(x_0)}{1+h(x_0,y_0)} + \frac{g(x_0)^2}{(1+h(x_0,y_0))^2}.\end{aligned}$$

Poiché x_0^2 divide $g(x_0)$ nel senso della divisibilità polinomiale appena osservata, prendendo x_0 sufficientemente piccolo e dello stesso segno di a (così che $ax_0 > 0$), si ottiene $(a-x_0)^2+(b-y_0)^2 < a^2+b^2$. Il punto (x_0,y_0) di C è dunque interno al cerchio. □

Abbiamo mostrato che se $C\colon f(x,y) = 0$ ha un punto in $R^{(2)}$, allora C e $D\colon yf_x - xf_y = 0$ hanno un punto comune in $R^{(2)}$. Sostituendo all'origine il punto (c,d), si vede allo stesso modo che, se C ha un punto in $R^{(2)}$, allora C e $D\colon (y-d)f_x - (x-c)f_y = 0$ hanno un punto comune in $R^{(2)}$.

Applichiamo ora il risultato per ottenere il metodo di decisione di Seidenberg. Con l'algoritmo euclideo determiniamo un massimo comune divisore $d(x)$ dei coefficienti delle potenze di y in $f(x,y)$, e scriviamo $f(x,y) = d(x)f_1(x,y)$, dove f_1 non è divisibile per alcun polinomio di grado positivo nella sola x. L'equazione $f(x,y) = 0$ è risolubile se e solo se è risolubile $d(x) = 0$ oppure $f_1(x,y) = 0$. Possiamo pertanto limitarci ai polinomi di $R[x,y] = (R[x])[y]$ primitivi rispetto a y (su $R[x]$), ossia privi di fattori di grado positivo nella sola x. Riduciamoci ora al caso dei polinomi senza fattori multipli. Calcoliamo col medesimo algoritmo un massimo comune divisore di $f(x,y)$ e $(\partial/\partial y)f(x,y)$ in $R(x)[y]$, dove $R(x)$ è il campo delle frazioni di $R[x]$. Possiamo scriverlo nella forma $u(x)v(x)^{-1}d(x,y)$, con $d \in R[x,y]$ primitivo rispetto a y. Per il lemma di Gauss (p. 152 dell'originale) d divide f in $R[x,y]$. Inoltre $g = f/d$ ha gli stessi fattori irriducibili di f e non ha fattori multipli (esercizio 1, p. 233 dell'originale). Supponiamo quindi f primitivo rispetto a y e privo di fattori multipli. Quest'ultima condizione implica che f e f_y non hanno in $R(x)[y]$ fattori comuni di grado positivo in y.

Sia t un'altra indeterminata e formiamo il risultante $h(t,x)$ di $f(x,y)$ e $g(t,x,y) = yf_x - (x-t)f_y$ considerati come polinomi in y. Per la definizione (15) del teorema 5.7, $h \in R[t,x]$. Affermiamo che $h(t,x) \neq 0$. In caso contrario $h(c,x) = 0$ per ogni $c \in R$, e allora $f(x,y)$ e $g(c,x,y) = yf_x - (x-c)f_y$ avrebbero in $R(x)[y]$ un fattore comune di grado positivo in y. Ciò segue dal teorema sui risultanti (p. 325), poiché il coefficiente della massima potenza di y in f è un elemento non nullo di $R[x]$. Un fattore comune non banale in $R(x)[y]$ dà un fattore comune non banale in $R[x,y]$. Ma, a meno di fattori invertibili, f ha un numero finito di fattori irriducibili in $R[x,y]$. Esistono dunque $c_1 \neq c_2$ tali che $g(c_1,x,y)$, $g(c_2,x,y)$ e $f(x,y)$ abbiano un fattore comune $d(x,y) \in R[x,y]$ di grado positivo. Allora $f(x,y)$ e $f_y = (c_1-c_2)^{-1}(g(c_1,x,y) - g(c_2,x,y))$ avrebbero un fattore comune non banale, contro l'ipotesi. Pertanto $h(t,x) \neq 0$.

Scegliamo $c \in R$ tale che $h(x) = h(c,x) \neq 0$ e poniamo $g(x,y) = g(c,x,y) = yf_x - (x-c)f_y$. Essendo $h(x)$ il risultante di f e g, questi non hanno fattori comuni di grado positivo in y; essendo f primitivo, non ne hanno neppure nella sola x. I loro fattori comuni in $R[x,y]$ sono dunque invertibili. Ne segue anche che il loro risultante $k(y)$ in $R[y][x]$ è un polinomio non nullo in y.

Abbiamo visto che, se $f(x,y) = 0$ ha una soluzione in $R^{(2)}$, allora $f(x,y) = 0$ e $g(x,y) = 0$ hanno una soluzione comune $(a,b) \in R^{(2)}$. I polinomi $f(a,y)$ e $g(a,y)$ hanno quindi il fattore comune $y-b$, e il loro risultante $h(a)$ è zero. La risolubilità di $f(x,y) = 0$ in $R^{(2)}$ implica dunque quella di $h(x) = 0$ in R. Che dire del viceversa? Mostreremo che vale, purché si scelgano opportunamente i generatori x, y di $R[x,y]$. Si ricordi che al loro posto si possono usare $x' = a_{11}x + a_{12}y$, $y' = a_{21}x + a_{22}y$ con $a_{ij} \in R$ e $\det(a_{ij}) \neq 0$.

Per ricondurre la questione della risolubilità di $f(x,y)=0$ in $R^{(2)}$ a quella di $h(x)=0$ in R, passiamo al campo algebricamente chiuso $A=R(\sqrt{-1})$. Sia V l'intersezione, in $A^{(2)}$, dei luoghi $f(x,y)=0$ e $g(x,y)=0$. Se $(a,b)\in V$, allora $h(a)=0$ per il risultante rispetto a y. Analogamente $k(b)=0$ per quello rispetto a x. Poiché $h(x)\neq 0$ e $k(y)\neq 0$, ognuna di queste equazioni ha solo un numero finito di radici in A. Pertanto V è finito. Se C ha un punto in $R^{(2)}$, ne ha uno anche V e h ha una radice in R. Viceversa, supponiamo che $a\in R$ sia radice di h. Se a non annulla il coefficiente $l(x)$ della massima potenza di y in $f(x,y)$, da $h(a)=0$ segue che esiste $b\in A$ con $(a,b)\in V$. Se $b\in R$, il punto è un punto reale di V, dunque di C. Altrimenti anche $(a,\bar{b})\in V$, dove $\bar{b}$ è il coniugato di b rispetto all'automorfismo non identico di A/R. Poiché $\bar{b}\neq b$, si avrebbero due punti distinti di V con la stessa ascissa. Pertanto, se il coefficiente direttivo non si annulla alle ascisse considerate e due punti distinti di V non hanno mai la stessa ascissa, la risolubilità di $h(x)=0$ in R implica quella di $f(x,y)=0$ in $R^{(2)}$.

Realizziamo entrambe le condizioni scegliendo opportunamente le coordinate. Per $m\in R$, $m\neq 0$, poniamo $x'=m^{-1}x-y$, $y'=y$. Allora $x=m(x'+y')$, $y=y'$, per cui $R[x,y]=R[x',y']$ e $f(x,y)=f(m(x'+y'),y')$. Indichiamo con f_n la parte omogenea di grado massimo $n>0$ in f. Il coefficiente di $(y')^n$ in $f(m(x'+y'),y')$ è $f_n(m,1)$. Siccome $f_n(x,1)\neq 0$, possiamo scegliere $m\in R$ che non ne sia radice. Essendo n il grado totale di f, la potenza massima di y' è allora $(y')^n$ e il suo coefficiente è la costante non nulla $f_n(m,1)$. Questo provvede alla prima condizione. Per la seconda, calcoliamo col procedimento euclideo un massimo comune divisore d di h e h'. Dividendo h per d otteniamo un polinomio h_1 con radici semplici $r_1,\dots,r_u$, le stesse di h. Analogamente calcoliamo $k_1(y)$ con radici semplici $s_1,\dots,s_v$, le stesse di k. Dividendo per i rispettivi coefficienti direttivi possiamo supporre h_1 e k_1 monici. Formiamo il polinomio

$$\prod_{\substack{i\neq i'\\ j\neq j'}}[(y_j-y_{j'})x-(x_i-x_{i'})], \tag{18}$$

dove le x e le y sono indeterminate, $1\leq i,i'\leq u$ e $1\leq j,j'\leq v$. Applicando due volte il teorema sui polinomi simmetrici (teorema 2.20, p. 139), mostreremo che (18) si esprime come polinomio in x i cui coefficienti sono polinomi a coefficienti interi nei polinomi simmetrici elementari delle x_i e delle y_j. Consideriamo dapprima il seguente polinomio nelle indeterminate x_i,t, a coefficienti interi:

$$\begin{aligned} l(x_1,\dots,x_n;t)&=\prod_{i\neq i'}[t-(x_i-x_{i'})]\\ &=t^m-l_1(x_1,\dots,x_n)t^{m-1}+l_2(x_1,\dots,x_n)t^{m-2}-\cdots, \end{aligned} \tag{19}$$

dove $m=n(n-1)$ e $l_i\in\mathbb{Z}[x_1,\dots,x_n]$. Il prodotto è invariante per ogni permutazione delle x_i, quindi gli l_i sono polinomi simmetrici nelle x_i a coefficienti interi. Possiamo allora scrivere ciascun l_j, in modo unico, come polinomio nei simmetrici elementari $p_1=\sum x_i$, $p_2=\sum_{i<j}x_ix_j,\dots,p_n=x_1\cdots x_n$. Così $l_j=m_j(p_1,\dots,p_n)\in\mathbb{Z}[p_1,\dots,p_n]$ e

$$l=t^m-m_1(p_1,\dots,p_n)t^{m-1}+m_2(p_1,\dots,p_n)t^{m-2}-\cdots.$$

Il polinomio (18) si scrive $\prod_{j\neq j'}l(x_1,\dots,x_u;(y_j-y_{j'})x)$. Sostituendo l'espressione di l e scrivendo $m=u(u-1)$ e $M=mv(v-1)$, otteniamo

$$\begin{aligned} &\prod_{\substack{i\neq i'\\ j\neq j'}}[(y_j-y_{j'})x-(x_i-x_{i'})]\\ &=\prod_{j\neq j'}[(y_j-y_{j'})^mx^m-m_1(p_1,\dots,p_u)(y_j-y_{j'})^{m-1}x^{m-1}\\ &\qquad\qquad+m_2(p_1,\dots,p_u)(y_j-y_{j'})^{m-2}x^{m-2}-\cdots]\\ &=z_0x^M-z_1x^{M-1}+z_2x^{M-2}-\cdots, \end{aligned}$$

dove $z_k \in \mathbb{Z}[p_1, \ldots, p_u][y_1, \ldots, y_v]$.[5] Anche le permutazioni delle y lasciano invariato il prodotto, per cui gli z_k sono simmetrici nelle y. Dunque

$$z_k \in \mathbb{Z}[p_1, \ldots, p_u][q_1, \ldots, q_v] = \mathbb{Z}[p_1, \ldots, p_u, q_1, \ldots, q_v],$$

dove $q_1 = \sum y_i$, $q_2 = \sum_{i<j} y_i y_j, \ldots$ sono i simmetrici elementari nelle y. Ciò prova che (18) si può scrivere come polinomio in x, nei p_i e nei q_j a coefficienti interi. Inoltre il calcolo è costruttivo, poiché era costruttiva la dimostrazione del teorema fondamentale sui polinomi simmetrici data nel § 2.13.

Sostituendo ai p_i e q_j nella formula di (18) i corrispondenti coefficienti di $h_1(x)$ e $k_1(y)$, con i segni alterni, otteniamo un polinomio $p(x) \in R[x]$ le cui radici sono gli elementi

$$(r_i - r_{i'})(s_j - s_{j'})^{-1} \in A, \qquad i \neq i', \quad j \neq j',$$

con gli stessi intervalli di indici di prima. Scegliamo m in modo che non sia radice né di $p(x)$ né di $f_n(x, 1)$. L'insieme dei punti (r_i, s_j) in $A^{(2)}$ contiene V, e due suoi punti distinti non hanno la stessa ascissa nel sistema (x', y'). Infatti le nuove coordinate di (x, y) sono $(m^{-1}x - y, y)$, e $m^{-1}r_i - s_j \neq m^{-1}r_{i'} - s_{j'}$ se $(i, j) \neq (i', j')$.

Ne segue che sostituendo $f(x, y)$ con $f(m(x + y), y)$ e $g(x, y)$ con $g(m(x + y), y)$ si soddisfano le condizioni per cui la nuova equazione $f(x, y) = 0$ è risolubile in $R^{(2)}$ se e solo se il risultante $h(x)$ dei due polinomi rispetto a y ha una radice in R. Il teorema di Sturm permette di decidere quest'ultima questione. Abbiamo così ottenuto una procedura per decidere la risolubilità dell'equazione originaria.

Nel prossimo paragrafo useremo un procedimento induttivo per polinomi in più parametri e più variabili. Il passo induttivo richiederà una piccola estensione della procedura appena descritta. Considereremo un'equazione $f(x, y) = 0$ soggetta alla condizione $g(x) \neq 0$. Come prima possiamo supporre f primitivo in y. Per evitare casi banali supponiamo inoltre $\deg_x f > 0$ e $\deg g > 0$. Sia $t(y)$ il risultante di $f(x, y)$ e $g(x)$ rispetto a x. Allora $t(y) \neq 0$, poiché f è primitivo rispetto a y. Scegliamo $c \in R$ tale che $t(c) \neq 0$, e al posto di f prendiamo $f_1(x, y) = f(x, y + c)$. Evidentemente il sistema $f(x, y) = 0$, $g(x) \neq 0$ è risolubile se e solo se lo è $f_1(x, y) = 0$, $g(x) \neq 0$. Il risultante rispetto a x di $f_1(x, y)$ e $g(x)$ è $t(y + c)$, che non si annulla per $y = 0$. Pertanto $g(x)$ e $f_1(x, 0)$ sono primi tra loro in $R[x]$. Poniamo $f_2(x, y) = f_1(x, g(x)y)$. Affermiamo che il sistema $f_1(x, y) = 0$, $g(x) \neq 0$ è risolubile in $R^{(2)}$ se e solo se lo è l'equazione $f_2(x, y) = 0$. Se (a, b) soddisfa il sistema, $f_2(a, g(a)^{-1}b) = f_1(a, b) = 0$. Viceversa, se $f_2(a, c) = f_1(a, g(a)c) = 0$, si ha $g(a) \neq 0$, poiché altrimenti $g(a) = 0 = f_1(a, 0)$, mentre $g(x)$ e $f_1(x, 0)$ sono primi tra loro. Il punto (a, b), con $b = g(a)c$, soddisfa dunque $f_1(x, y) = 0$, $g(x) \neq 0$.

5.6 Il teorema di Tarski

Consideriamo un sistema finito φ di equazioni, condizioni di non annullamento e disuguaglianze della forma

$$F(t_1, \ldots, t_r; x_1, \ldots, x_n) = 0, \quad G(t_1, \ldots, t_r; x_1, \ldots, x_n) \neq 0, \quad H(t_1, \ldots, t_r; x_1, \ldots, x_n) > 0,$$

con F, G, H polinomi a coefficienti interi. Vogliamo mostrare che si può sostituire φ con una famiglia finita di sistemi ψ_j dello stesso tipo, che non contengono le x, con la seguente proprietà per ogni campo reale chiuso R. Assegnati i valori $t_i = c_i \in R$, il sistema φ ha una soluzione nelle x in R se e solo se i c_i soddisfano tutte le condizioni di uno dei sistemi ψ_j. Elimineremo a una

[5] Nell'originale l'esponente totale è indicato con m^2, valore che si ottiene quando $u = v$.

a una tutte le x tranne l'ultima, col metodo del paragrafo precedente. Potremo allora applicare la versione parametrica del teorema di Sturm. Per cominciare, però, procediamo in direzione opposta. Sostituiamo il sistema φ con una sola equazione $F = 0$ al prezzo di introdurre nuove x. Un insieme finito di equazioni $F_i = 0$ si può sostituire con $\sum F_i^2 = 0$ e un insieme finito di condizioni $G_i \neq 0$ con $\prod G_i \neq 0$. La condizione $G \neq 0$ equivale a $G^2 > 0$, e la risolubilità di $H > 0$ equivale alla risolubilità di $Hz^2 - 1 = 0$, dove z è una nuova indeterminata.[6] Con queste riduzioni possiamo supporre φ costituito da una sola equazione $F(t_1, \ldots, t_r; x_1, \ldots, x_n) = 0$. Per il passo induttivo dovremo conservare, insieme all'equazione, una condizione di non annullamento. Si ha infatti il seguente teorema.

Teorema 5.8. *Siano $F(t_i; x, y) \in \mathbb{Z}[t_1, \ldots, t_r; x, y]$, $G(t_i; x) \in \mathbb{Z}[t_1, \ldots, t_r; x]$, dove t_i, x, y sono indeterminate. Si può determinare in un numero finito di passi una famiglia finita di coppie di polinomi $(F_j(t_i; x), G_j(t_i))$, con $F_j \in \mathbb{Z}[t_i; x]$, $G_j \in \mathbb{Z}[t_i]$, $1 \leq j \leq h$, con la seguente proprietà. Per ogni campo reale chiuso R e ogni punto $(c_1, \ldots, c_r) \in R^{(r)}$, il sistema*

$$F(c_i; x, y) = 0, \qquad G(c_i; x) \neq 0 \tag{20}$$

è risolubile nelle incognite x, y in R se e solo se è risolubile in R almeno uno dei sistemi

$$F_j(c_i; x) = 0, \qquad G_j(c_i) \neq 0, \qquad 1 \leq j \leq h. \tag{21}$$

La dimostrazione consisterà in una successione finita di costruzioni di ricoprimenti e di polinomi corrispondenti ai passi del metodo di Seidenberg. Poiché ci interessano esclusivamente campi reali chiusi, possiamo supporre che ogni elemento Δ di un ricoprimento δ di $A = \mathbb{Z}[t_1, \ldots, t_r]$ consista di una sola equazione $d = 0$ e di una sola condizione $l \neq 0$, con $l, d \in A$. Se $\gamma = \{\Gamma_k\}$ è un ricoprimento, esso definisce un raffinamento δ' di δ in cui al termine Δ si sostituiscono $\Delta^{(1)}, \Delta^{(2)}, \ldots$. L'equazione di $\Delta^{(k)}$ è la somma dei quadrati dei membri delle equazioni di Δ e Γ_k, e la condizione di non annullamento è espressa dal prodotto dei rispettivi membri (si veda p. 319 dell'originale). Allora, per ogni campo reale chiuso R, $\Delta^{(k)}(R) = \Delta(R) \cap \Gamma_k(R)$ e $\Delta(R) = \bigcup_k \Delta^{(k)}(R)$. I singoli passi della dimostrazione saranno del tipo seguente. Si parte da un ricoprimento δ e, per $\Delta \in \delta$, da una coppia di polinomi $(F(t_i; x, y), G(t_i; x))$ appartenenti rispettivamente ad $A[x, y]$ e ad $A[x]$. Si costruiscono un ricoprimento δ' come sopra e, per ciascun $\Delta^{(k)}$, un insieme finito di coppie $(F_{kj}(t_i; x, y), G_{kj}(t_i; x))$ tali che, per $(c_1, \ldots, c_r) \in \Delta^{(k)}(R)$, il sistema $F(c_i; x, y) = 0$, $G(c_i; x) \neq 0$ sia risolubile in R se e solo se è risolubile uno dei sistemi $F_{kj}(c_i; x, y) = 0$, $G_{kj}(c_i; x) \neq 0$. Si può così sostituire la terna (Δ, F, G) con le diverse terne $(\Delta^{(k)}, F_{kj}, G_{kj})$. Dopo un numero finito di passi si ottengono un ricoprimento $\omega = \{\Omega_j\}$ e un insieme finito di coppie (F_{jk}, G_{jk}), con $F_{jk} \in A[x]$, $G_{jk} \in A$. Per $(c_1, \ldots, c_r) \in \Omega_j(R)$ il sistema dato in partenza è risolubile se e solo se è risolubile uno dei sistemi $F_{jk}(c_i; x) = 0$, $G_{jk}(c_i) \neq 0$. Poniamo ora $F'_{jk} = F_{jk}^2 + d_j^2$, $G'_{jk} = G_{jk} l_j$, dove $d_j = 0$, $l_j \neq 0$ definiscono Ω_j. È facile vedere che le coppie (F'_{jk}, G'_{jk}) così ottenute possiedono le proprietà richieste nell'enunciato.

Osserviamo poi che, dato un insieme finito $\{F, G, \ldots, H\} \subset A[x]$, possiamo costruire un ricoprimento $\delta = \{\Delta\}$ di A e dei polinomi corrispondenti che rappresentano il massimo comune divisore dell'insieme nel senso seguente.

(i) Per ogni $\Delta \in \delta$ esiste $D(t_i; x) \in A[x]$ tale che, per ogni campo reale chiuso R e ogni $(c_1, \ldots, c_r) \in \Delta(R)$, $D(c_i; x)$ è un massimo comune divisore in $R[x]$ di $F(c_i; x), G(c_i; x), \ldots, H(c_i; x)$.

[6] Nell'originale compare $H^2 z^2 - 1 = 0$. Questa formula esprime $H \neq 0$, mentre $Hz^2 - 1 = 0$ esprime $H > 0$ in un campo reale chiuso.

(ii) Per ogni Δ, o $D(c_i;x) = 0$ per ogni $(c_1,\dots,c_r) \in \Delta(R)$, oppure $D(c_i;x) \neq 0$ per tutti questi punti. Nel secondo caso esistono $F_1, G_1, \dots, H_1 \in A[x]$ tali che, per ciascuno di quei punti, $F_1(c_i;x), G_1(c_i;x), \dots, H_1(c_i;x)$ differiscano rispettivamente da $F(c_i;x)/D(c_i;x), G(c_i;x)/D(c_i;x), \dots, H(c_i;x)/D(c_i;x)$ per un fattore non nullo di R.

Per ottenere (i), osserviamo che, se uno dei polinomi dati è zero, basta applicare l'induzione sul loro numero. Per un solo polinomio il risultato è evidente, e per due polinomi non nulli segue dal lemma di p. 319. Supponiamo ora che i polinomi non nulli siano più di due. Per induzione possiamo ammettere di aver costruito un ricoprimento γ di A e i corrispondenti polinomi E per il massimo comune divisore di tutti i polinomi dati tranne H. Per ciascuna coppia $\{E, H\}$ costruiamo poi un ricoprimento e i polinomi per il suo massimo comune divisore. Un raffinamento come nella dimostrazione del lemma di p. 319 dà (i). Raffinando ulteriormente possiamo supporre che, per ogni Δ del nuovo ricoprimento, o $D(c_i;x) = 0$ per tutti i punti di $\Delta(R)$, oppure

$$D(t_i;x) = v_k(t_1,\dots,t_r)x^k + v_{k-1}(t_1,\dots,t_r)x^{k-1} + \cdots + v_0(t_1,\dots,t_r)$$

e $v_k(c_1,\dots,c_r) \neq 0$ in tutto $\Delta(R)$. Con l'algoritmo della divisione si ottengono un intero $e \geq 0$ e polinomi $F_1, G_1, \dots, H_1, S, T, \dots, U$ in $A[x]$ tali che

$$v_k^e F = F_1 D - S, \quad v_k^e G = G_1 D - T, \quad \dots, \quad v_k^e H = H_1 D - U,$$

dove i gradi in x di $S, T, \dots, U$ sono tutti minori di k. Poiché $D(c_i;x)$ divide $F(c_i;x), G(c_i;x), \dots$, si ha $S(c_i;x) = T(c_i;x) = \cdots = 0$. I polinomi $F_1, G_1, \dots, H_1$ soddisfano perciò la condizione (ii).

Ci occorre un'estensione al caso di due polinomi in due indeterminate x, y oltre ai parametri t_i. Siano dati $F(t_i;x,y), G(t_i;x,y) \in A[x,y]$, $A = \mathbb{Z}[t_1,\dots,t_r]$. Possiamo costruire un ricoprimento $\delta = \{\Delta\}$ e dei polinomi che rappresentano il massimo comune divisore di F, G in $R(x)[y]$ nel senso seguente.

(1) A ogni Δ corrisponde un polinomio $D(t_i;x,y)$ tale che per ogni $(c_1,\dots,c_r) \in \Delta(R)$, $D(c_i;x,y)$ è un massimo comune divisore in $R(x)[y]$ di $F(c_i;x,y)$ e $G(c_i;x,y)$.

(2) Per ogni Δ o $D(c_i;x,y) = 0$ per tutti i punti di $\Delta(R)$, oppure $D(c_i;x,y) \neq 0$ per tutti. In quest'ultimo caso esistono $F_1, G_1 \in A[x,y]$ tali che, a meno di un fattore non nullo di $R[x]$, $F_1(c_i;x,y)$ e $G_1(c_i;x,y)$ coincidono rispettivamente con $F(c_i;x,y)/D(c_i;x,y)$ e $G(c_i;x,y)/D(c_i;x,y)$. Vale a dire, esiste $l(x) \neq 0$ tale che $l(x)F(c_i;x,y) = F_1(c_i;x,y)D(c_i;x,y)$, e analogamente per G.

Per un polinomio $v(t_i;x) \in A[x]$ la condizione $v(c_i;x) = 0$ equivale all'annullarsi per $t_i = c_i$ della somma dei quadrati dei suoi coefficienti, e la condizione $v(c_i;x) \neq 0$ al non annullarsi della stessa somma. Questa osservazione permette di estendere al caso di due indeterminate i risultati sull'algoritmo della divisione contenuti nel lemma di p. 319. Il ragionamento precedente dà allora la proprietà enunciata per $F, G \in A[x,y]$. Lasciamo i particolari al lettore.

Occorre un altro accorgimento formale, corrispondente alla scelta di un $\eta \in R$ con $g(\eta) \neq 0$ per un dato polinomio non nullo $g(x)$. Sia $G(t_i;x) = v_m x^m + \cdots + v_0 \neq 0$, dove $v_j = v_j(t_1,\dots,t_r) \in A$. Formiamo il ricoprimento (9) e prendiamo $k \neq -\infty$. Per $(c_1,\dots,c_r) \in \Gamma_k(R)$ avremo

$$G(c_i;x) = v_k(c_1,\dots,c_r)x^k + \cdots + v_0(c_1,\dots,c_r), \qquad v_k(c_1,\dots,c_r) \neq 0.$$

Sia $L_k(t_1,\dots,t_r)$ l'espressione razionale

$$(k+1) + \sum_{j=0}^{k-1} v_j(t_1,\dots,t_r)^2 v_k(t_1,\dots,t_r)^{-2}.$$

Essa è definita su $\Gamma_k(R)$ e $G(c_i; L_k(c_1, \dots, c_r)) \neq 0$ per la (7). Servirà anche l'analogo per due indeterminate x, y. Se $G(t_i; x, y) \neq 0$, possiamo costruire un ricoprimento $\gamma = \{\Gamma\}$ ed espressioni razionali $L(t_i) \in \mathbb{Q}(t_1, \dots, t_r)$, rapporti di polinomi a coefficienti interi, tali che per ogni Γ valga una delle due alternative. O $G(c_i; x, y) = 0$ su tutto $\Gamma(R)$, oppure, per uno degli $L(t_i)$, $L(c_i)$ è definito e $G(c_i; L(c_i), y) \neq 0$ su tutto $\Gamma(R)$. La dimostrazione estende immediatamente il ragionamento che precede.

Dimostrazione del teorema 5.8. Riduciamo dapprima le due relazioni $F(t_i; x, y) = 0$, $G(t_i; x) \neq 0$ a una sola equazione $K(t_i; x, y) = 0$. Questo passo corrisponde alla parte finale del paragrafo precedente. Costruiamo un ricoprimento δ_1 e i polinomi per il massimo comune divisore dei coefficienti delle potenze di y in F insieme con G. Sia $\Delta_1 \in \delta_1$ e sia $D_1(t_i; x)$ il polinomio associato. Per $(c_1, \dots, c_r) \in \Delta_1(R)$, $D_1(c_i; x)$ è allora un massimo comune divisore in $R[x]$ dei coefficienti delle potenze di y in $F(c_i; x, y)$ e di $G(c_i; x)$. Se $D_1(c_i; x) = 0$, il sistema non ammette soluzioni. Possiamo dunque supporre $D_1(c_i; x) \neq 0$ su tutto $\Delta_1(R)$. Otteniamo allora polinomi $F_1(t_i; x, y) \in A[x, y]$, $G_1(t_i; x) \in A[x]$, tali che $F(c_i; x, y)$ e $G(c_i; x)$ differiscano da $D_1(c_i; x)F_1(c_i; x, y)$ e da $D_1(c_i; x)G_1(c_i; x)$ per un moltiplicatore non nullo di R. L'argomento riduce la coppia alla forma $F_1(c_i; \xi, \eta) = 0$, $G_1(c_i; \xi) \neq 0$, in cui i coefficienti di F_1 rispetto a y e G_1 sono primi tra loro nel loro insieme.[7] Raffiniamo δ_1 in δ_2 applicando a ciascun Δ_1 il ricoprimento associato ai coefficienti in x di G_1, come nella (9). Ci riduciamo così a insiemi $\Delta_2 \in \delta_2$ e polinomi F_2, G_2 tali che, su $\Delta_2(R)$, $F_2(c_i; x, y) = F_1(c_i; x, y)$, $G_2(c_i; x) = G_1(c_i; x)$ e

$$G_2(t_i; x) = v_k x^k + v_{k-1} x^{k-1} + \cdots + v_0, \qquad v_j \in A, \quad v_k(c_1, \dots, c_r) \neq 0.$$

Sia $T(t_i; y)$ il risultante di $F_2(t_i; x, y)$ e $G_2(t_i; x)$ come polinomi in x. Poiché $v_k(c_1, \dots, c_r) \neq 0$, l'uguaglianza $T(c_i; y) = 0$ implicherebbe che $F_2(c_i; x, y)$ e $G_2(c_i; x)$ abbiano in $R(y)[x]$ un fattore comune di grado positivo in x. Possiamo scriverlo nella forma $a(y)b(y)^{-1}h(x, y)$ con $h \in R[x, y]$ primitivo in x su $R[y]$. Allora h divide $F_2(c_i; x, y)$ e $G_2(c_i; x)$, e dal secondo fatto segue $h \in R[x]$. Questo contraddice il fatto che i coefficienti delle potenze di y in $F_2(c_i; x, y)$ e $G_2(c_i; x)$ sono primi tra loro. Dunque $T(c_i; y) \neq 0$ su tutto $\Delta_2(R)$. Passiamo a un raffinamento δ_3 tale che per ogni $\Delta_3 \in \delta_3$ esista un'espressione razionale $L = Q/P$, $P, Q \in A$, con $P(c_i) \neq 0$ e $T(c_i; L(c_i)) \neq 0$ su $\Delta_3(R)$. Sostituiamo al corrispondente F_2 il polinomio

$$F_3(t_i; x, y) = P(t_i)^f F_2(t_i; x, y + L(t_i)), \qquad f = \deg_y F_2(t_i; x, y),$$

e scriviamo G_3 per G_2. Il risultante di F_3 e G_3 rispetto a x ha la forma $P(t_i)^g T(t_i; y + L(t_i))$ e non si annulla per $t_i = c_i$, $y = 0$, su $\Delta_3(R)$. Come nel § 5.5, il sistema $F_3(c_i; x, y) = 0$, $G_3(c_i; x) \neq 0$ è risolubile in R se e solo se lo è $F_4(c_i; x, y) = 0$, dove $F_4(t_i; x, y) = F_3(t_i; x, G_3(t_i; x)y)$. Per ciascun termine del ricoprimento δ_3 siamo così ricondotti a una sola equazione senza condizioni di non annullamento.

Riprendiamo da capo, supponendo di avere la sola equazione $F(t_i; x, y) = 0$: il risultato si applicherà poi ai diversi F_4 e Δ_3 appena considerati. Costruiamo un ricoprimento δ_1 e i polinomi per il massimo comune divisore dei coefficienti di F rispetto a y. Per ogni Δ_1 abbiamo $D_1(t_i; x) \in A[x]$, $F_1(t_i; x, y) \in A[x, y]$ con le proprietà seguenti su $\Delta_1(R)$. Il polinomio $D_1(c_i; x)$ è un massimo comune divisore dei coefficienti di $F(c_i; x, y)$ rispetto a y, e $F(c_i; x, y)$ differisce da $D_1(c_i; x)F_1(c_i; x, y)$ per un fattore non nullo di R. Ne segue che $F(c_i; x, y) = 0$ è risolubile in R se e solo se è risolubile $D_1(c_i; x) = 0$ oppure $F_1(c_i; x, y) = 0$. La prima equazione è già del tipo

[7] Nota alla dimostrazione originale. Nel cancellare D_1 va conservata la condizione $D_1(c_i; \xi) \neq 0$, già implicita in $G(c_i; \xi) \neq 0$. Un modo equivalente per rendere rigorosa la riduzione consiste nel conservare il vincolo originale $G \neq 0$ e dividere ripetutamente i coefficienti di F per il loro massimo comune divisore con G finché questo è costante. L'equazione resta equivalente dove $G \neq 0$ e il grado in x dei coefficienti diminuisce a ogni divisione non banale. I casi $F = 0$ o $G = 0$ si decidono direttamente.

cercato e la conserviamo come alternativa. Resta da esaminare la seconda, dove $F_1(c_i; x, y)$ è primitivo in y su $R[x]$. Per ciascun F_1 costruiamo un ricoprimento relativo al massimo comune divisore di F_1 e $\partial F_1/\partial y$. Ne ricaviamo un ricoprimento δ_2 con la proprietà seguente. Per ogni Δ_2 proveniente da Δ_1 vi sono $D_2, F_2 \in A[x, y]$ tali che, su $\Delta_2(R)$, esiste $l(x) \neq 0$ in $R[x]$ con

$$l(x)F_1(c_i; x, y) = D_2(c_i; x, y)F_2(c_i; x, y),$$

e $D_2(c_i; x, y)$ è un massimo comune divisore di $F_1(c_i; x, y)$ e $\partial F_1(c_i; x, y)/\partial y$ in $R(x)[y]$. Pertanto $F_1(c_i; x, y)$ e $F_2(c_i; x, y)$ hanno in $R[x, y]$ gli stessi fattori irriducibili di grado positivo in y, e nessuno di questi compare con molteplicità maggiore di uno in F_2. Applichiamo il primo passo a F_2. Si ottiene un raffinamento δ_3 di δ_2 tale che, per ogni Δ_3 proveniente da Δ_2, vi è un polinomio $F_3(t_i; x, y) \in A[x, y]$ che su $\Delta_3(R)$ è primitivo in y su $R[x]$ e ha gli stessi fattori irriducibili di grado positivo in y di F_1, tutti con molteplicità uno. Su $\Delta_3(R)$ la risolubilità di $F_1(c_i; x, y) = 0$ equivale perciò a quella di $F_3(c_i; x, y) = 0$. Inoltre $F_3(c_i; x, y)$ e la sua derivata in y non hanno fattori comuni di grado positivo. Poniamo

$$G_3(t_i, t; x, y) = y\,\partial F_3/\partial x - (x - t)\,\partial F_3/\partial y,$$

dove t è una nuova indeterminata. Sia $H(t_i, t; x)$ il risultante rispetto a y di $G_3(t_i, t; x, y)$ e $F_3(t_i; x, y)$. Come nel metodo di decisione, si ha $H(c_i, t; x) \neq 0$. Con un nuovo raffinamento δ_4 otteniamo, per ogni Δ_4, un'espressione razionale $L = Q/P$, $P, Q \in A$, tale che $P(c_i) \neq 0$ e $H(c_i, L(c_i); x) \neq 0$ su $\Delta_4(R)$. Se sostituiamo G_3 con

$$G_4(t_i; x, y) = P(t_i)G_3(t_i, L(t_i); x, y) \in A[x, y]$$

e poniamo $F_4 = F_3$, il risultante $H(t_i; x)$ di F_4, G_4 rispetto a y soddisfa $H(c_i; x) \neq 0$ su $\Delta_4(R)$. Il resto della dimostrazione segue allo stesso modo i passi del metodo di decisione. Lasciamo al lettore il compito di svolgerli. □

Possiamo ora combinare questo teorema di eliminazione con la versione parametrica del teorema di Sturm (teorema 5.5) per dimostrare il nostro risultato principale.

Teorema di Tarski. *Sia φ un insieme finito di equazioni polinomiali, condizioni di non annullamento e disuguaglianze della forma $F(t_1, \ldots, t_r; x_1, \ldots, x_n) = 0$, $G(t_1, \ldots, t_r; x_1, \ldots, x_n) \neq 0$, $H(t_1, \ldots, t_r; x_1, \ldots, x_n) > 0$, con $F, G, H \in \mathbb{Z}[t_1, \ldots, t_r; x_1, \ldots, x_n]$. Si può determinare in un numero finito di passi una famiglia finita di insiemi finiti ψ_j di equazioni polinomiali, condizioni di non annullamento e disuguaglianze dello stesso tipo nei soli parametri t_i, con la seguente proprietà. Per ogni campo reale chiuso R, il sistema φ ammette una soluzione nelle x in R per $t_i = c_i$, $1 \leq i \leq r$, se e solo se i c_i soddisfano tutte le condizioni di uno degli insiemi ψ_j.*

Dimostrazione. Come sopra, possiamo sostituire il sistema dato con una sola equazione in un numero di incognite eventualmente maggiore di n. Possiamo dunque supporre che il sistema abbia la forma

$$F(t_1, \ldots, t_r; x_1, \ldots, x_n) = 0, \qquad G(t_1, \ldots, t_r; x_1, \ldots, x_{n-1}) \neq 0,$$

con $n \geq 1$. Se $n = 1$ si applica il teorema 5.5 e si aggiunge la condizione sui parametri $G(t_1, \ldots, t_r) \neq 0$ a ciascuna condizione Γ_j fornita da quel teorema. Se $n > 1$, trattiamo $x_1, \ldots, x_{n-2}$ come parametri $t_{r+1}, \ldots, t_{r+n-2}$ e applichiamo il teorema 5.8, che sostituisce al sistema un insieme finito di sistemi della forma

$$F_j(t_1, \ldots, t_r; x_1, \ldots, x_{n-1}) = 0, \qquad G_j(t_1, \ldots, t_r; x_1, \ldots, x_{n-2}) \neq 0.$$

La conclusione segue per induzione sul numero delle x. □

Supponiamo ora dati due campi reali chiusi R_1, R_2 con un sottocampo comune F, sul quale gli ordinamenti indotti coincidano. Sia dato un sistema di equazioni, condizioni di non annullamento e disuguaglianze a coefficienti in F avente una soluzione in R_1. Introducendo parametri possiamo intendere questa affermazione nel senso che un certo sistema parametrico a coefficienti interi ha soluzione in R_1 per certi valori $t_i = c_i \in F$. Il teorema di Tarski implica allora che i c_i soddisfano uno di un certo numero di sistemi di equazioni, condizioni di non annullamento e disuguaglianze a coefficienti razionali, determinabili a priori e indipendenti da R_1. Percorrendo a ritroso il ragionamento vediamo che il sistema iniziale ha una soluzione in R_2. In particolare, se un sistema a coefficienti razionali ha una soluzione in un campo reale chiuso R_1, per esempio $\mathbb{R}$, esso ha una soluzione in ogni campo reale chiuso.

Più in generale il teorema di Tarski implica il suo principio metamatematico secondo il quale ogni enunciato «elementare» dell'algebra vero in un campo reale chiuso, per esempio nel campo reale, è vero in ogni campo reale chiuso. Per un'esposizione precisa e dettagliata rimandiamo ai testi di logica matematica.[8] Ci limiteremo a indicare brevemente il significato del principio e a illustrarlo con un'applicazione non banale.

Una *formula atomica* è un'espressione della forma $f > 0$ o $f = 0$, dove f è un polinomio a coefficienti razionali. Una *formula* è un'espressione ottenuta da un numero finito di formule atomiche applicando la congiunzione («e»), la disgiunzione («o»), la negazione («non») e il quantificatore esistenziale («esiste un x tale che»). Altri concetti logici come «implica» e «per ogni x» si possono definire in questi termini. Un *enunciato elementare* è una formula priva di variabili libere.

Per applicare il principio di Tarski occorre saper riconoscere che una determinata asserzione è un enunciato elementare o equivale a uno di essi. A titolo di esempio dimostriamo la seguente estensione del lemma 1 del § 5.5. Sia R un campo reale chiuso e siano $f_1(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)$ polinomi a coefficienti in R. Supponiamo che le equazioni $f_i(x_1, \dots, x_n) = 0$ abbiano una soluzione comune $(a_1, \dots, a_n) \in R^{(n)}$. Allora esiste una soluzione di minima distanza dall'origine, cioè tale che $\sum a_i^2$ sia minimo.

Sostituiamo dapprima il sistema con la sola equazione $f = \sum f_i^2 = 0$. Sostituiamo poi i coefficienti con parametri, ottenendo un polinomio $F(t_1, \dots, t_r; x_1, \dots, x_n)$ a coefficienti interi. La nostra asserzione si presenta ora nella forma elementare seguente. Per ogni $t_i = c_i$ in R, o $F(c_1, \dots, c_r; x_1, \dots, x_n) = 0$ non ha soluzioni, oppure ha una soluzione $x_j = a_j$ tale che $\sum a_j^2 \leq \sum b_j^2$ per ogni soluzione $x_j = b_j$. Ciò si dimostra facilmente in $\mathbb{R}$ col ragionamento che precede il lemma 1 di p. 328, e vale dunque in ogni campo reale chiuso R.

Merita menzione anche un'importante applicazione del teorema di Tarski alle equazioni alle derivate parziali.[9] È un notevole esempio delle connessioni tra i diversi rami della matematica: un risultato nato nella logica matematica ha una conseguenza importante in una delle discipline più applicate. Il teorema di Tarski interviene inoltre nel § 11.4 del volume 2 come elemento importante della dimostrazione di un teorema sulle funzioni razionali definite positive che risponde a un celebre problema di Hilbert.

[8] L'esposizione originale di Tarski è *A Decision Method for Elementary Algebra and Geometry*, RAND Corporation, 1948. Una dimostrazione del principio, detto «eliminazione dei quantificatori» nella teoria dei campi reali chiusi, si trova in G. Kreisel e J. L. Krivine, *Elements of Mathematical Logic*, London, North-Holland, seconda stampa riveduta, 1971, pp. 60–65. Il lavoro di Seidenberg è apparso negli *Annals of Mathematics*, vol. 60 (1954), pp. 365–374.

[9] Si veda A. Friedman, *Generalized Functions and Partial Differential Equations*, Englewood Cliffs, N.J., Prentice Hall, 1963, cap. 7.

Esercizio

1. Completare i dettagli lasciati al lettore nella dimostrazione del teorema 5.8.

Capitolo 6
Spazi vettoriali metrici e gruppi classici

Dal punto di vista analitico la geometria euclidea studia uno spazio vettoriale V di dimensione n su $\mathbb{R}$ rispetto a una particolare forma bilineare simmetrica, che permette di definire sia la lunghezza di un vettore sia il coseno dell'angolo fra due vettori. Prendendo $V = \mathbb{R}^{(n)}$, possiamo scegliere la forma bilineare usuale

$$x \cdot y = \sum_{i=1}^{n} x_i y_i, \qquad x = (x_1, \dots, x_n), \quad y = (y_1, \dots, y_n).$$

Allora $x \cdot x = \sum x_i^2 = |x|^2$ è il quadrato della lunghezza di x e, se θ è l'angolo fra x e y, si ha $\cos\theta = (x \cdot y)/(|x|\,|y|)$. La funzione $x \cdot y$, detta *prodotto scalare* di x e y, è bilineare, nel senso che

$$\begin{aligned} (x + x') \cdot y &= x \cdot y + x' \cdot y, \\ x \cdot (y + y') &= x \cdot y + x \cdot y', \\ (ax) \cdot y &= a(x \cdot y) = x \cdot (ay) \end{aligned}$$

per vettori x, x', y, y' e uno scalare reale a. Il prodotto scalare è simmetrico ($x \cdot y = y \cdot x$) e definito positivo ($x \cdot x > 0$ per $x \neq 0$). Tutto ciò è familiare dalla geometria analitica in due e tre dimensioni, e l'estensione a un qualunque n è immediata e naturale.

Vi sono due modi di generalizzare questa situazione. Anzitutto si può sostituire all'ipotesi di dimensione finita una condizione di completezza: si giunge così agli spazi di Hilbert reali. Sostituendo poi $\mathbb{R}$ con $\mathbb{C}$ e la forma bilineare simmetrica con una forma hermitiana definita positiva, si ottengono gli spazi di Hilbert complessi, tanto importanti in analisi. Qui seguiremo l'altra via: considereremo le estensioni della geometria euclidea che si ottengono sostituendo a $\mathbb{R}^{(n)}$ un qualsiasi spazio vettoriale V di dimensione finita su un campo arbitrario F, e al prodotto scalare una qualunque forma bilineare non degenere $B(x, y)$ (si veda il § 6.1), simmetrica, ossia $B(x, y) = B(y, x)$, oppure alternante, ossia $B(x, x) \equiv 0$. Chiameremo $B(x, y)$ una *metrica* su V. La geometria ottenuta da una forma simmetrica è detta *geometria ortogonale*; quella associata a una forma alternante si chiama *geometria simplettica*. Sono questi gli unici casi in cui l'ortogonalità fra vettori, definita ponendo $x \perp y$ quando $B(x, y) = 0$, è una relazione simmetrica.

A una metrica $B(x, y)$ si associa il gruppo delle trasformazioni lineari η di V per le quali $B(\eta x, \eta y) = B(x, y)$ per ogni $x, y \in V$. Se B è simmetrica si parla di *gruppo ortogonale*; se è alternante, di *gruppo simplettico*. Questi gruppi e il gruppo lineare generale delle trasformazioni lineari biiettive di uno spazio vettoriale di dimensione finita sono i gruppi «classici», nella terminologia di Hermann Weyl. Vedremo in quale senso essi siano prossimi a essere semplici. Dimostrare il risultato preciso al riguardo è uno degli obiettivi principali del capitolo.

Nella prima parte getteremo le basi delle geometrie ortogonale e simplettica. Tratteremo il problema dell'equivalenza delle forme, le forme reali e il teorema di Sylvester sull'inerzia di una forma simmetrica reale (o di una forma quadratica), il teorema di cancellazione di Witt e il teorema di Cartan–Dieudonné sulla generazione dei gruppi ortogonali mediante simmetrie. Passeremo quindi alla struttura dei gruppi geometrici classici. Nell'ultimo paragrafo accenneremo all'estensione della teoria alle forme hermitiane.

6.1 Funzionali lineari e forme bilineari

Sia V uno spazio vettoriale di dimensione finita sul campo F. Ricordiamo che un funzionale lineare su V è un'applicazione di V nel campo di base F tale che

$$f(x+y) = f(x) + f(y), \qquad f(ax) = af(x) \tag{1}$$

per $x, y \in V$, $a \in F$. Questi funzionali costituiscono uno spazio vettoriale V^*, detto *spazio coniugato* di V, nel quale l'addizione e la moltiplicazione per $a \in F$ sono definite da

$$(f+g)(x) = f(x) + g(x), \qquad (af)(x) = af(x). \tag{2}$$

Se $(e_1, e_2, \ldots, e_n)$ è una base di V su F, definiamo il funzionale lineare e_i^* mediante le condizioni

$$e_i^*(e_i) = 1, \qquad e_i^*(e_j) = 0 \quad \text{se } j \neq i. \tag{3}$$

Allora $(e_1^*, \ldots, e_n^*)$ è una base di V^* su F. Infatti, se $x^* \in V^*$ e $x^*(e_i) = a_i \in F$, si ha $(\sum_j a_j e_j^*)(e_i) = a_i = x^*(e_i)$. Poiché un'applicazione lineare è determinata dai valori su una base, ne segue $x^* = \sum_i a_i e_i^*$. Se $\sum_i a_i e_i^* = 0$, allora $a_j = \sum_i a_i e_i^*(e_j) = 0$; pertanto gli e_i^* sono linearmente indipendenti. La base $(e_1^*, \ldots, e_n^*)$ si chiama *duale*, o *complementare*, della base $(e_1, \ldots, e_n)$. Evidentemente V^* ha dimensione n.

Una *forma bilineare* B su V è un'applicazione $(x, y) \mapsto B(x, y)$ da $V \times V$ a F tale che per ogni $y \in V$ l'applicazione

$$y_R : x \longmapsto B(x, y) \tag{4}$$

sia un funzionale lineare su V e per ogni $x \in V$ lo sia l'applicazione

$$x_L : y \longmapsto B(x, y). \tag{5}$$

Queste condizioni equivalgono alle seguenti:

$$\begin{aligned} &B(x + x', y) = B(x, y) + B(x', y), \quad B(ax, y) = aB(x, y), \\ &B(x, y + y') = B(x, y) + B(x, y'), \quad B(x, ay) = aB(x, y). \end{aligned} \tag{6}$$

Possiamo raccoglierle nella sola condizione

$$B(a_1x_1 + a_2x_2, b_1y_1 + b_2y_2) = \sum_{i,j=1}^{2} a_i b_j B(x_i, y_j). \tag{6'}$$

Per induzione la (6′) si estende a

$$B\left(\sum_{i=1}^{m} a_i x_i, \sum_{j=1}^{q} b_j y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{q} a_i b_j B(x_i, y_j). \tag{7}$$

La (7) suggerisce un procedimento generale per costruire forme bilineari. Fissiamo una base $(e_1, \ldots, e_n)$ di V su F e scegliamo per ogni coppia di indici (i, j), con $1 \leq i, j \leq n$, un elemento $b_{ij} \in F$. Per $x = \sum a_i e_i$ e $y = \sum b_i e_i$ poniamo

$$B(x, y) = \sum_{i,j=1}^{n} b_{ij} a_i b_j. \tag{8}$$

Una verifica diretta mostra che $(x, y) \mapsto B(x, y)$ è bilineare. Dalla (7) risulta inoltre che ogni forma bilineare su V si ottiene in questo modo. La matrice $(B(e_i, e_j))$ determinata da B e da

una base $(e_1, \dots, e_n)$ si chiama *matrice di B rispetto alla base* $(e_1, \dots, e_n)$. Il suo determinante $\det(B(e_i, e_j))$ è detto un *discriminante* di B.

Sostituiamo alla base (e_i) una base (f_i), con $f_i = \sum_j p_{ij} e_j$, e sia $q = (q_{ij})$ l'inversa della matrice $p = (p_{ij})$. Abbiamo

$$B(f_i, f_j) = B\left(\sum_k p_{ik} e_k, \sum_l p_{jl} e_l\right) = \sum_{k,l} p_{ik} B(e_k, e_l) p_{jl}.$$

Se $b = (B(e_i, e_j))$, la matrice di B rispetto alla base (f_i) è dunque

$$p b\, {}^t p, \tag{9}$$

dove ${}^t p$ è la trasposta di p. Il determinante della nuova matrice è $(\det p)^2 \det b$: così si trasforma il discriminante $\det b$ al cambiare della base.

Consideriamo le applicazioni $x \mapsto x_L$ e $y \mapsto y_R$ determinate da B come nelle (4) e (5). Poiché $x_L, y_R \in V^*$, esse mandano V in V^*; inoltre sono lineari. Infatti, a $x_1, x_2 \in V$ corrispondono i funzionali $y \mapsto B(x_1, y)$ e $y \mapsto B(x_2, y)$; la loro somma manda y in $B(x_1, y) + B(x_2, y) = B(x_1 + x_2, y)$, ed è dunque $(x_1 + x_2)_L$. Abbiamo perciò $(x_1)_L + (x_2)_L = (x_1 + x_2)_L$. Analogamente $(ax)_L$ manda y in $B(ax, y) = aB(x, y)$, cioè coincide con $a(x_L)$. Ne segue la linearità di $x \mapsto x_L$. Allo stesso modo si verifica la linearità di $y \mapsto y_R$.

Dato un sottospazio U di V, poniamo

$$U^{\perp_L} = \{v \in V \mid B(v, u) = 0 \text{ per ogni } u \in U\},$$
$$U^{\perp_R} = \{v \in V \mid B(u, v) = 0 \text{ per ogni } u \in U\}.$$

Sono sottospazi di V. Dalle definizioni risulta che

$$(U^{\perp_L})^{\perp_R} \supset U, \qquad (U^{\perp_R})^{\perp_L} \supset U, \tag{10}$$

e che, se $U_1 \supset U_2$, allora

$$U_1^{\perp_L} \subset U_2^{\perp_L}, \qquad U_1^{\perp_R} \subset U_2^{\perp_R}. \tag{11}$$

I sottospazi $V^{\perp_L}$ e $V^{\perp_R}$ si chiamano rispettivamente *radicale sinistro* e *radicale destro* di B.

Teorema 6.1. *Per una forma bilineare B sono equivalenti le tre condizioni seguenti: (1) $V^{\perp_R} = 0$; (2) $V^{\perp_L} = 0$; (3) la matrice di B rispetto a una base qualsiasi è invertibile.*

Dimostrazione. Sia $(e_1, \dots, e_n)$ una base e sia $B(e_i, e_j) = b_{ij}$. Per bilinearità, se $B(e_i, z) = 0$ per $1 \le i \le n$, allora $B(x, z) = 0$ per ogni $x \in V$. Dunque $z \in V^{\perp_R}$ se e solo se $B(e_i, z) = 0$ per ogni i; analogamente $z \in V^{\perp_L}$ se e solo se $B(z, e_i) = 0$ per ogni i. Scriviamo $z = \sum_j c_j e_j$. Allora $B(e_i, z) = \sum_j c_j B(e_i, e_j) = \sum_j b_{ij} c_j$. Pertanto $z \in V^{\perp_R}$ se e solo se $(c_1, \dots, c_n)$ soddisfa il sistema omogeneo

$$b_{i1} c_1 + b_{i2} c_2 + \cdots + b_{in} c_n = 0, \qquad 1 \le i \le n. \tag{12}$$

Allo stesso modo $z = \sum_j c_j e_j$ appartiene a $V^{\perp_L}$ se e solo se

$$b_{1i} c_1 + b_{2i} c_2 + \cdots + b_{ni} c_n = 0, \qquad 1 \le i \le n. \tag{13}$$

Dall'algebra lineare sappiamo che la condizione perché il sistema (12), o il (13), abbia una soluzione diversa da $(0, \dots, 0)$ è $\det(b_{ij}) = 0$. Ne segue l'asserto. □

La forma B si dice *non degenere* se soddisfa le condizioni del teorema 6.1.

La condizione su z che $B(x,z)=0$ per ogni x, cioè che $z\in V^{\perp_R}$, equivale a dire che il funzionale z_R è nullo. Dunque $V^{\perp_R}$ è il nucleo dell'applicazione lineare $R:V\to V^*$. Pertanto B è non degenere se e solo se il nucleo di R è nullo, ossia se e solo se R è iniettiva. Poiché $\dim V=\dim V^*$, ciò equivale alla suriettività di R: ogni funzionale su V ha la forma $x_R:y\mapsto B(y,x)$ per qualche $x\in V$. Similmente B è non degenere se e solo se ogni funzionale su V ha la forma $y\mapsto B(x,y)$ per qualche $x\in V$.

Sempre nell'ipotesi che B sia non degenere, mostriamo che le applicazioni $U\mapsto U^{\perp_R}$ e $U\mapsto U^{\perp_L}$ sono l'una inversa dell'altra e perciò biiettive sull'insieme dei sottospazi di V. Dimostreremo a questo scopo la relazione

$$\dim U^{\perp_R}=n-\dim U=\dim U^{\perp_L}. \tag{14}$$

Ricordiamo che, se T è un'applicazione lineare di uno spazio vettoriale V di dimensione finita in un altro spazio vettoriale, $\dim V=\dim T(V)+\dim\ker T$. Infatti, da $V/\ker T$ su $T(V)$ si ha la biiezione indotta $x+\ker T\mapsto Tx$. Siano ora U un sottospazio di V e $x\in V$. Il funzionale x_R si restringe a un funzionale su U. Disponiamo dunque dell'applicazione lineare $x\mapsto x_R|U$ da V allo spazio coniugato U^*. Il suo nucleo è costituito dagli x tali che $B(y,x)=0$ per ogni $y\in U$, ed è quindi $U^{\perp_R}$. Ne segue $n=\dim V=\dim U^{\perp_R}+\dim W$, dove W è lo spazio dei funzionali su U della forma $y\mapsto B(y,x)$, con $x\in V$. Poiché $\dim U^*=\dim U$, per ottenere la (14) sarà sufficiente provare che $W=U^*$. Sia g un funzionale su U. Possiamo estenderlo a un funzionale su V: basta completare una base $(f_1,\dots,f_r)$ di U a una base $(f_1,\dots,f_n)$ di V e definire g' in modo che coincida con g su $f_1,\dots,f_r$ e assuma valori arbitrari in F sui rimanenti vettori della base. Ma abbiamo visto che g' è della forma $y\mapsto B(y,x)$ per qualche $x\in V$; lo stesso vale dunque per g. Questo dimostra $\dim U^{\perp_R}=n-\dim U$; analogamente $\dim U^{\perp_L}=n-\dim U$. Applicando due volte le formule ottenute si trova $\dim(U^{\perp_L})^{\perp_R}=n-\dim U^{\perp_L}=n-(n-\dim U)=\dim U$, e anche $\dim(U^{\perp_R})^{\perp_L}=\dim U$. In virtù delle inclusioni (10), per ogni sottospazio U e per B non degenere si ha quindi

$$(U^{\perp_L})^{\perp_R}=U=(U^{\perp_R})^{\perp_L}. \tag{15}$$

Nella dimostrazione ha avuto una parte importante l'aver determinato la forma dei funzionali su U. Poiché dovremo richiamare questo fatto, lo enunciamo come lemma.

Lemma. *Sia B non degenere e sia U un sottospazio di V. Ogni funzionale lineare su U ha la forma $y\mapsto B(x,y)$ (e anche la forma $y\mapsto B(y,x)$) per un opportuno $x\in V$.*

Se $B(x,y)=0$ diciamo che x è *ortogonale* a y e scriviamo $x\perp y$. È desiderabile che si tratti di una relazione simmetrica, ossia che $x\perp y$ se e solo se $y\perp x$. È facile determinare quando ciò avviene.

Teorema 6.2. *L'ortogonalità definita da una forma bilineare B su V è una relazione simmetrica se e solo se B è simmetrica, ossia $B(x,y)=B(y,x)$ per ogni x,y, oppure B è alternante, ossia $B(x,x)=0$ per ogni $x\in V$.*

Dimostrazione. L'ortogonalità definita da una forma simmetrica è evidentemente simmetrica. Se B è alternante, allora per ogni x,y si ha $B(x,y)+B(y,x)=B(x+y,x+y)-B(x,x)-B(y,y)=0$, e l'antisimmetria di B implica $x\perp y$ se e solo se $y\perp x$. Supponiamo ora che valga quest'ultima proprietà. Per $x,y,z\in V$ arbitrari poniamo $w=B(x,y)z-B(x,z)y$. Allora $x\perp w$, e la condizione $w\perp x$ equivale a

$$B(x,y)B(z,x)=B(x,z)B(y,x). \tag{16}$$

Ponendo $y = x$ otteniamo, per ogni x, z,

$$B(x,x)(B(z,x) - B(x,z)) = 0. \tag{17}$$

Affermiamo che $B(x,y) = B(y,x)$ per ogni x, y, oppure $B(x,x) = 0$ per ogni x. Altrimenti esisterebbero u, v con $B(u,v) \neq B(v,u)$ e un vettore w con $B(w,w) \neq 0$. Per la (17) si avrebbe allora $B(u,u) = B(v,v) = 0$, e inoltre $B(w,u) = B(u,w)$ e $B(w,v) = B(v,w)$. Dalla (16), poiché $B(u,v) \neq B(v,u)$, si ricava anche $B(u,w) = B(w,u) = 0$ e $B(v,w) = B(w,v) = 0$. Ne segue $B(u, w+v) = B(u,v) \neq B(v,u) = B(w+v, u)$, e quindi, per la (17), $B(w+v, w+v) = 0$. Ma $B(w+v, w+v) = B(w,w) + B(w,v) + B(v,w) + B(v,v) = B(w,w)$, in contraddizione con $B(w,w) \neq 0$. □

Due forme bilineari B e B' definite rispettivamente sugli spazi vettoriali V e V' si dicono *equivalenti* se esiste un'applicazione lineare biiettiva $x \mapsto x'$ di V su V' tale che $B(x,y) = B'(x',y')$ per ogni $x, y \in V$. Questa è evidentemente una relazione di equivalenza e comporta che V e V' abbiano la stessa dimensione. Se B è alternante (rispettivamente simmetrica), anche B' ha la stessa proprietà.

Esercizi

1. Sia B una forma bilineare su V. Mostrare per due sottospazi qualsiasi U_1, U_2 che $(U_1 + U_2)^{\perp_L} = U_1^{\perp_L} \cap U_2^{\perp_L}$ e $(U_1 + U_2)^{\perp_R} = U_1^{\perp_R} \cap U_2^{\perp_R}$. Mostrare inoltre che, se B è non degenere, $(U_1 \cap U_2)^{\perp_L} = U_1^{\perp_L} + U_2^{\perp_L}$ e $(U_1 \cap U_2)^{\perp_R} = U_1^{\perp_R} + U_2^{\perp_R}$.

2. Sia B una forma bilineare arbitraria su V e sia U un sottospazio al quale B si restringe in modo non degenere. Mostrare che $V = U \oplus U^{\perp_L} = U \oplus U^{\perp_R}$.

3. Sia B non degenere su V. Mostrare che a ogni forma bilineare C su V corrisponde un'unica trasformazione lineare $L_C : V \to V$ tale che $C(x,y) = B(L_C x, y)$ per ogni $x, y \in V$. Mostrare che C è non degenere se e solo se L_C è biiettiva. Mostrare che esiste un'unica trasformazione lineare biiettiva P di V su sé stesso tale che $B(y,x) = B(Px, y)$ per ogni $x, y \in V$.

4. Se B è non degenere, mostrare che per ogni trasformazione lineare T di V in sé esiste un'unica trasformazione lineare T' tale che $B(Tx, y) = B(x, T'y)$ per ogni $x, y \in V$. Determinare la matrice di T' in funzione di quelle di T e di B rispetto a una base di V. Mostrare che $T \mapsto T'$ è un antiautomorfismo dell'anello delle trasformazioni lineari, e che $(T')' = T$ per ogni T se B è simmetrica o antisimmetrica ($B(y,x) = -B(x,y)$).

5. Date due forme bilineari B_1, B_2 su V, si ponga $(B_1 + B_2)(x,y) = B_1(x,y) + B_2(x,y)$ e, per $a \in F$, $(aB_1)(x,y) = aB_1(x,y)$. Mostrare che queste sono forme bilineari e che, rispetto alle operazioni così definite, le forme bilineari su V costituiscono uno spazio vettoriale su F. Dimostrare che la sua dimensione è n^2.

6. Sia B bilineare simmetrica (rispettivamente alternante) su V, così che $U^{\perp_L} = U^{\perp_R}$ per ogni sottospazio U. Sia W un sottospazio di $V^\perp$. Mostrare che $B(x + W, y + W) = B(x,y)$ definisce su V/W una forma bilineare simmetrica (rispettivamente alternante), non degenere se e solo se $W = V^\perp$.

7. Mostrare che per ogni forma bilineare B esistono basi $(u_1, \ldots, u_n)$ e $(v_1, \ldots, v_n)$ di V tali che $(B(u_i, v_j)) = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$.

8. Sia B una forma bilineare. Fissati i vettori u, v, l'applicazione $x \mapsto B(x,u)v$ è una trasformazione lineare di V in sé. La si indichi con $u \otimes v$. Trovare una formula per la traccia $\mathrm{tr}(u \otimes v)$. Mostrare che, se B è non degenere, ogni trasformazione lineare ha la forma $\sum_i u_i \otimes v_i$.

6.2 Forme alternanti

Le forme bilineari che considereremo nel resto del capitolo sono simmetriche o alternanti. In entrambi i casi $U^{\perp_L} = U^{\perp_R}$ per ogni sottospazio U; indicheremo questo sottospazio con $U^\perp$ e lo chiameremo *complemento ortogonale* di U. Si ha $U \cap U^\perp = 0$ se e solo se la restrizione di B a U è non degenere. In tal caso diremo che U è un *sottospazio non degenere*. Una base $(e_1, \dots, e_n)$ di V determina la matrice $b = (B(e_i, e_j))$ di B. Cambiando base, al posto di b si ottiene $pb\,{}^t p$, con p invertibile. Due matrici $b, c \in M_n(F)$ si dicono *congruenti* se esiste $p \in \mathrm{GL}_n(F)$, il gruppo delle unità di $M_n(F)$, tale che $c = pb\,{}^t p$. È una relazione di equivalenza; a B si associa dunque una classe di congruenza, costituita dalle matrici di B rispetto alle varie basi ordinate di V. Abbiamo chiamato discriminante di B il determinante di una sua matrice b e abbiamo visto che B è non degenere se e solo se $\det b \neq 0$. Precisiamo ora questa nozione: definiamo il discriminante uguale a zero se B è degenere, e altrimenti uguale alla classe laterale $(\det b)F^{*2}$ di $\det b$ nel gruppo F^*/F^{*2}, dove F^* è il gruppo moltiplicativo degli elementi non nulli di F e F^{*2} il sottogruppo dei loro quadrati. Chiameremo $(\det b)F^{*2}$ *il discriminante* di B. I vari discriminanti $\det b$, per b matrice di una stessa forma non degenere, sono precisamente i rappresentanti di questa classe laterale.

Abbiamo visto che, se B è alternante, allora è antisimmetrica: $B(x,y) = -B(y,x)$. Inoltre, se $\mathrm{char}\, F \neq 2$, dall'antisimmetria segue $2B(x,x) = 0$ e quindi $B(x,x) = 0$. In caratteristica diversa da due, pertanto, essere alternante equivale a essere antisimmetrica. Se B è alternante e $(e_1, \dots, e_n)$ è una base, si ha $B(e_i, e_j) = -B(e_j, e_i)$ e $B(e_i, e_i) = 0$. Quindi $b = (B(e_i, e_j))$ è una *matrice alternante*: è antisimmetrica, ossia ${}^t b = -b$, e ha gli elementi diagonali nulli. Viceversa, per ogni matrice alternante $b = (b_{ij})$, la forma bilineare definita da $B(x,y) = \sum_{i,j} b_{ij} a_i b_j$ per $x = \sum_i a_i e_i$ e $y = \sum_i b_i e_i$ è alternante, poiché

$$B(x,x) = \sum_{i \neq j} b_{ij} a_i a_j = \sum_{i<j} (b_{ij} + b_{ji}) a_i a_j = 0.$$

La teoria della struttura delle forme bilineari alternanti è particolarmente semplice. Sia B una tale forma. Dimostreremo che esiste una base

$$(u_1, v_1, u_2, v_2, \dots, u_r, v_r, z_1, \dots, z_{n-2r}) \tag{18}$$

rispetto alla quale la matrice di B è

$$\mathrm{diag}\{S, S, \dots, S, 0, \dots, 0\}, \tag{19}$$

dove

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{20}$$

La notazione indica che sulla diagonale si susseguono blocchi S e poi zeri, mentre tutti gli altri elementi sono nulli. Se $B = 0$, cioè $B(x,y) = 0$ per ogni x, y, il risultato è immediato. Altrimenti possiamo scegliere u, v con $B(u,v) = b \neq 0$. Per $u_1 = u$, $v_1 = b^{-1}v$ si ha allora $B(u_1, v_1) = 1 = -B(v_1, u_1)$. Poiché $B(x, ax) = aB(x,x) = 0$, u_1, v_1 sono linearmente indipendenti. Essi

costituiscono l'inizio della base cercata. Supponiamo di avere già trovato vettori linearmente indipendenti

$$(u_1, v_1, u_2, v_2, \dots, u_k, v_k) \tag{21}$$

per i quali $B(u_i, v_i) = 1 = -B(v_i, u_i)$ e $B(x, y) = 0$ per ogni altra scelta di x, y fra i vettori indicati. Sia V_k il sottospazio di dimensione $2k$ generato dagli u_i, v_i. Affermiamo che $V = V_k \oplus V_k^\perp$. Infatti la matrice della restrizione di B a V_k nella base (21) è $\operatorname{diag}\{S, \dots, S\}$ ed è invertibile. La restrizione è dunque non degenere e $V_k \cap V_k^\perp = 0$. Per $x \in V$ si consideri

$$y = x - \sum_{i=1}^{k} B(x, v_i)u_i + \sum_{i=1}^{k} B(x, u_i)v_i.$$

Si ha

$$\begin{aligned} B(y, u_j) &= B(x, u_j) + B(x, u_j)B(v_j, u_j) = 0, \\ B(y, v_j) &= B(x, v_j) - B(x, v_j)B(u_j, v_j) = 0; \end{aligned}$$

quindi $y \in V_k^\perp$. Dalla relazione $x = y + \sum B(x, v_i)u_i - \sum B(x, u_i)v_i$ segue che $V = V_k + V_k^\perp$, da cui la somma diretta asserita. Consideriamo ora B ristretta a $V_k^\perp$. Se la restrizione è nulla scegliamo una base $(z_1, \dots, z_{n-2k})$ di $V_k^\perp$ e otteniamo la base (18), con $r = k$. Se invece la restrizione è diversa da zero possiamo scegliere $u_{k+1}, v_{k+1} \in V_k^\perp$ in modo che $B(u_{k+1}, v_{k+1}) = 1 = -B(v_{k+1}, u_{k+1})$. Alla successione data si può allora sostituire quella che termina con u_{k+1}, v_{k+1}. Proseguendo così otteniamo il risultato seguente.

Teorema 6.3. *Per ogni forma bilineare alternante B esiste una base (18) di V rispetto alla quale la matrice di B è $s = \operatorname{diag}\{S, S, \dots, S, 0, \dots, 0\}$, con $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.*

Una matrice alternante b determina una forma bilineare alternante B di cui b è la matrice rispetto a una base assegnata (e_i) di V. Se p esprime la base (u_j, v_j, z_k) del teorema 6.3 in funzione degli e_i, si ha $pb\,{}^t p = s$. Posto $q = p^{-1}$ si ottiene $b = qs\,{}^t q$. Le matrici b e s hanno lo stesso rango, poiché moltiplicare a destra o a sinistra per una matrice invertibile conserva il rango. Inoltre $\det s$ vale 0 o 1, sicché $\det b = (\det q)^2 \det s$ vale 0 oppure $(\det q)^2$. Ne segue il primo corollario.

Corollario 1. *Il rango di una matrice alternante a coefficienti in un campo è pari e il suo determinante è un quadrato.*

È chiaro inoltre il seguente risultato.

Corollario 2. *Due matrici alternanti $n \times n$ a coefficienti in un campo sono congruenti se e solo se hanno lo stesso rango.*

Il corollario 1 ammette un importante affinamento. Sia n pari e sia $F = \mathbb{Q}(x_{ij})$ il campo delle funzioni razionali a coefficienti razionali nelle $n(n-1)/2$ indeterminate x_{ij} con $i < j$. Sia X la matrice alternante in $M_n(F)$ che per $i < j$ ha elemento (i, j) uguale a x_{ij}. Gli elementi diagonali sono nulli e, se $i > j$, l'elemento (i, j) è $-x_{ji}$. Per il corollario 1, $\det X$ è il quadrato di un elemento di $\mathbb{Q}(x_{ij})$. Questo campo è il campo delle frazioni del suo sottoanello $\mathbb{Z}[x_{ij}]$, perciò esistono $f, g \in \mathbb{Z}[x_{ij}]$ tali che $\det X = (f/g)^2$. Semplificando i fattori comuni possiamo supporre che f e g non abbiano fattori comuni, eccettuate le unità ± 1, nell'anello fattoriale $\mathbb{Z}[x_{ij}]$. Dalla relazione $g^2 \det X = f^2$ e dalla fattorialità segue che g è un'unità, ossia $g = \pm 1$. Quindi $\det X = f^2$, e tale relazione determina f a meno del segno.

Sia R un anello commutativo e sia $a = (a_{ij})$ una matrice alternante $n \times n$ a coefficienti in R. Esiste un unico omomorfismo di $\mathbb{Z}[x_{ij}]$ in R che manda x_{ij} in a_{ij} per $i < j$. Applicandolo a $\det X = f(x_{12}, x_{13}, \ldots)^2$ si ottiene $\det a = f(a_{12}, a_{13}, \ldots)^2$. In particolare, per $R = \mathbb{Z}$ e $s = \operatorname{diag}\{S, \ldots, S\}$, troviamo $1 = \det s = f(1, \ldots)^2$. Scegliamo il segno di f in modo che $f(1, \ldots) = 1$. Indicheremo il polinomio così determinato con $\operatorname{Pf} X$ e lo chiameremo *pfaffiano* di X. Sostituendo alle x_{ij} le a_{ij} si ottiene il *pfaffiano della matrice alternante* a, indicato con $\operatorname{Pf} a$. Esso soddisfa

$$(\operatorname{Pf} a)^2 = \det a. \tag{22}$$

Abbiamo così stabilito la prima parte del seguente teorema.

Teorema 6.4. *Sia n pari e sia X la matrice alternante $n \times n$ il cui elemento (i, j) è l'indeterminata x_{ij} per $i < j$. Esiste un unico polinomio $\operatorname{Pf} X$ nelle x_{ij}, a coefficienti interi, tale che $(\operatorname{Pf} X)^2 = \det X$ e $\operatorname{Pf} s = 1$ per $s = \operatorname{diag}\{S, \ldots, S\}$, $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Per ogni anello commutativo R e ogni matrice alternante $a \in M_n(R)$ vale la (22). Inoltre, per $q \in M_n(R)$ arbitraria, $qa\,{}^t q$ è alternante e*

$$\operatorname{Pf}(qa\,{}^t q) = (\det q) \operatorname{Pf} a. \tag{23}$$

Dimostrazione. Sia a alternante e q arbitraria. Allora ${}^t(qa\,{}^t q) = -qa\,{}^t q$. Scrivendo $q = (q_{ij})$ e $a = (a_{ij})$ si vede che l'elemento (i, i) di $qa\,{}^t q$ è

$$\begin{aligned} \sum_{j,k} q_{ij} a_{jk} q_{ik} &= \sum_{j<k} q_{ij} a_{jk} q_{ik} + \sum_{j>k} q_{ij} a_{jk} q_{ik} \\ &= \sum_{j<k} q_{ij} a_{jk} q_{ik} - \sum_{j>k} q_{ik} a_{kj} q_{ij} \\ &= \sum_{j<k} q_{ij} a_{jk} q_{ik} - \sum_{j<k} q_{ik} a_{jk} q_{ij} = 0. \end{aligned}$$

Ne segue che $qa\,{}^t q$ è alternante. Per dimostrare la (23) lavoriamo nel campo $\mathbb{Q}(x_{ij}, y_{kl})$, dove le x_{ij} sono le $n(n-1)/2$ indeterminate precedenti e le y_{kl}, con $1 \le k, l \le n$, sono n^2 nuove indeterminate. Siano X come sopra e $Y = (y_{kl})$. La matrice $YX\,{}^t Y$ è alternante e

$$\left(\operatorname{Pf}(YX\,{}^t Y)\right)^2 = \det(YX\,{}^t Y) = (\det Y)^2 \det X = (\det Y)^2 (\operatorname{Pf} X)^2.$$

Pertanto $\operatorname{Pf}(YX\,{}^t Y) = \pm(\det Y) \operatorname{Pf} X$. Ponendo $Y = 1$ si vede che il segno è positivo. Per specializzazione si ottiene quindi la (23). □

Dalla (23) ricaviamo un metodo per calcolare il pfaffiano di una matrice alternante su un campo.

Corollario. *Sia a una matrice alternante a coefficienti in un campo e sia q invertibile tale che $qa\,{}^t q = \operatorname{diag}\{S, \ldots, S, 0, \ldots, 0\}$ come nel teorema 6.3. Allora $\operatorname{Pf} a = (\det q)^{-1}$ se a è invertibile e $\operatorname{Pf} a = 0$ altrimenti.*

Per $n = 2$ e $n = 4$ il calcolo è agevole. Si ottengono rispettivamente

$$\operatorname{Pf} X = x_{12}, \qquad \operatorname{Pf} X = x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23}.$$

Gli esercizi che seguono indicano formule per valori maggiori di n.

Esercizi

1. Mostrare che le matrici

$$b = \begin{pmatrix} 0 & 2 & -1 & 3 \\ -2 & 0 & 4 & -2 \\ 1 & -4 & 0 & 1 \\ -3 & 2 & -1 & 0 \end{pmatrix}, \qquad s = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

   sono congruenti in $M_4(\mathbb{Q})$ e trovare una matrice p tale che $pb\,{}^t p = s$.

2. Sia B una forma bilineare alternante e siano $(u_1, v_1, \ldots, u_k, v_k)$ tali che $B(u_i, v_i) = 1 = -B(v_i, u_i)$, con $B(x, y) = 0$ per tutte le altre coppie dei vettori indicati. Nella notazione dell'esercizio 8 di p. 349, si ponga $E_k = \sum_{i=1}^{k}(u_i \otimes v_i - v_i \otimes u_i)$. Verificare che $E_k^2 = E_k$ e che $B(E_k x, y) = B(x, E_k y)$ per $x, y \in V$.

3. Sia B una forma bilineare alternante non degenere su V e sia T una trasformazione lineare di V in sé. Si definisca *aggiunta di T rispetto a B* l'unica trasformazione lineare T' tale che $B(Tx, y) = B(x, T'y)$ per ogni $x, y \in V$. Determinare l'aggiunta di $u \otimes v$ rispetto a B.

4. Mostrare che $\operatorname{Pf} a$ dipende linearmente da ciascuna riga della matrice alternante a quando si fissano gli elementi della sottomatrice ottenuta eliminando quella riga e la colonna corrispondente.

5. Sia $a = (a_{ij})$ alternante. Posto $\alpha_{ij} = (-1)^{i+j-1} \operatorname{Pf} A_{ij}$, dove A_{ij} è la matrice $(n-2)\times(n-2)$ ottenuta eliminando le righe i, j e le colonne i, j di a, mostrare che $\operatorname{Pf} a = a_{12}\alpha_{12} + a_{13}\alpha_{13} + \cdots + a_{1n}\alpha_{1n}$.

6. Sia $q = (q_{ij})$ una matrice arbitraria $n \times n$ a coefficienti in un anello commutativo, con n pari. Si definisca

$$a_{ij} = \begin{vmatrix} q_{1i} & q_{1j} \\ q_{2i} & q_{2j} \end{vmatrix} + \begin{vmatrix} q_{3i} & q_{3j} \\ q_{4i} & q_{4j} \end{vmatrix} + \cdots + \begin{vmatrix} q_{n-1,i} & q_{n-1,j} \\ q_{ni} & q_{nj} \end{vmatrix}.$$

   Mostrare che $a = (a_{ij})$ è alternante e $\operatorname{Pf} a = \det q$.

7. Sia $s = \operatorname{diag}\{S, S, \ldots, S\}$, con $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Una matrice $a \in M_n(R)$, con R anello commutativo, si dica *simpletticamente simmetrica* se $s^{-1}\,{}^t a s = a$. Mostrare che la condizione equivale all'antisimmetria di sa. Mostrare che a è radice dell'equazione $\operatorname{Pf}(s\lambda - sa) = 0$.

6.3 Forme quadratiche e forme bilineari simmetriche

Sia V uno spazio vettoriale con base $(e_1, e_2, \ldots, e_n)$ e sia $f(x_1, \ldots, x_n)$ un polinomio in n indeterminate con coefficienti nel campo di base F di V. Esso determina da V in F la funzione polinomiale f data da

$$f\left(\sum_i a_i e_i\right) = f(a_1, \ldots, a_n).$$

Se $f(x_1, \ldots, x_n)$ è un polinomio omogeneo di grado r (definizione a p. 138), la funzione corrispondente si chiama *forma di grado r*. In particolare le forme di grado $1, 2, 3, \ldots$ si chiamano rispettivamente lineari, quadratiche, cubiche, e così via. Poiché un polinomio omogeneo di grado uno si scrive $\sum_i c_i x_i$, il concetto di forma lineare coincide con quello di funzione lineare su V.

Un polinomio omogeneo di grado due si scrive $\sum_{i\le j} c_{ij}x_ix_j$; pertanto una forma quadratica è un'applicazione

$$Q\colon x = \sum_i a_ie_i \longmapsto \sum_{i\le j} c_{ij}a_ia_j, \tag{24}$$

dove i c_{ij} sono elementi fissati di F. Queste applicazioni ammettono una semplice caratterizzazione assiomatica, che fornisce la seguente definizione alternativa.

Definizione 6.1. Una *forma quadratica* Q è un'applicazione $x \mapsto Q(x)$ da uno spazio vettoriale V nel suo campo di base F, tale che

1. $Q(ax) = a^2Q(x)$ per $a \in F$, $x \in V$;
2. $B(x, y) = Q(x+y)-Q(x)-Q(y)$ è bilineare, ossia $(x, y) \mapsto B(x, y)$ è una forma bilineare, evidentemente simmetrica.

Le due definizioni date sono equivalenti. Supponiamo anzitutto che Q sia definita dalla (24). Allora $ax = \sum_i aa_ie_i$ e

$$Q(ax) = \sum_{i\le j} c_{ij}(aa_i)(aa_j) = a^2Q(x).$$

Inoltre, se $y = \sum_i b_ie_i$, si ha $x + y = \sum_i(a_i + b_i)e_i$ e

$$\begin{aligned} Q(x + y) - Q(x) - Q(y) &= \sum_{i\le j} c_{ij}((a_i + b_i)(a_j + b_j) - a_ia_j - b_ib_j) \\ &= \sum_{i\le j} c_{ij}(a_ib_j + b_ia_j) = \sum_{i,j} d_{ij}a_ib_j, \end{aligned}$$

dove $d_{ii} = 2c_{ii}$, $d_{ij} = c_{ij}$ per $i < j$ e $d_{ij} = c_{ji}$ per $i > j$. È chiaro che B è bilineare; dunque Q definita dalla (24) soddisfa gli assiomi 1 e 2. Viceversa, supponiamo che Q soddisfi 1 e 2. Per $a, b \in F$ e $x, y \in V$ si ha

$$Q(ax + by) = a^2Q(x) + b^2Q(y) + abB(x, y).$$

Per induzione otteniamo

$$Q\left(\sum_i a_ie_i\right) = \sum_i a_i^2Q(e_i) + \sum_{i<j} a_ia_jB(e_i, e_j);$$

dunque $Q(x) = \sum_{i\le j} c_{ij}a_ia_j$, dove $c_{ii} = Q(e_i)$ e $c_{ij} = B(e_i, e_j)$ per $i < j$. Pertanto Q ha la forma (24).

La forma bilineare B associata a Q, ossia $B(x, y) = Q(x + y) - Q(x) - Q(y)$, è simmetrica e soddisfa $B(x, x) = 2Q(x)$. Se $\operatorname{char} F \neq 2$, B determina Q perché $Q(x) = \frac{1}{2}B(x, x)$. Se $\operatorname{char} F = 2$, si ha $B(x, x) = 0$ e quindi B è alternante. Se B è una forma bilineare, $Q(x) = B(x, x)$ è una forma quadratica la cui forma bilineare associata è $B(x, y) + B(y, x)$.

Per una forma bilineare simmetrica B abbiamo definito il radicale di B, $\operatorname{rad} B = V^\perp$; B è non degenere se e solo se $\operatorname{rad} B = 0$. Per una forma quadratica Q definiamo il *radicale bilineare*, $\operatorname{bilrad} Q$, come il radicale della forma bilineare associata B. Definiamo invece il *radicale di* Q ponendo

$$\operatorname{rad} Q = \{z \mid Q(x + z) = Q(x) \text{ per ogni } x \in V\}.$$

Poiché $Q(x+z) = Q(x)+Q(z)+B(x, z)$, si vede che $z \in \operatorname{rad} Q$ se e solo se $Q(z) = 0$ e $B(x, z) = 0$ per ogni $x \in V$. Quindi $\operatorname{rad} Q \subseteq \operatorname{bilrad} Q$, e $\operatorname{rad} Q$ è l'insieme degli $z \in \operatorname{bilrad} Q$ tali che $Q(z) = 0$.

Da questo fatto, o direttamente dalla definizione iniziale, si vede che $\operatorname{rad} Q$ è un sottospazio di V. Se $\operatorname{char} F \neq 2$, la relazione $B(x,z) = 0$ per ogni x implica $Q(z) = \frac{1}{2}B(z,z) = 0$, sicché in questo caso $\operatorname{rad} Q = \operatorname{bilrad} Q$.

In generale chiamiamo *difetto* di Q la dimensione dello spazio quoziente $\operatorname{bilrad} Q/\operatorname{rad} Q$, ossia la differenza $\dim \operatorname{bilrad} Q - \dim \operatorname{rad} Q$. Essa è zero se $\operatorname{char} F \neq 2$. Se invece $\operatorname{char} F = 2$, allora $F^2 = \{a^2 \mid a \in F\}$ è un sottocampo di F, e possiamo considerare la dimensione $[F : F^2]$ di F come spazio vettoriale su F^2. Supponiamola finita e sia $r > [F : F^2]$. Siano $z_1, z_2, \ldots, z_r \in \operatorname{bilrad} Q$. Poiché $r > [F : F^2]$, possiamo scegliere a_i non tutti nulli tali che $\sum_i a_i^2 Q(z_i) = 0$. Allora $Q(\sum_i a_i z_i) = \sum_i a_i^2 Q(z_i) = 0$, quindi $\sum_i a_i z_i \in \operatorname{rad} Q$. Vediamo così che $r > [F : F^2]$ elementi qualsiasi di $\operatorname{bilrad} Q$ sono linearmente dipendenti modulo $\operatorname{rad} Q$. Ne segue che il difetto di Q non supera $[F : F^2]$. In particolare, se F è perfetto, $F = F^2$ e il difetto di Q è zero o uno.

La teoria delle forme quadratiche in caratteristica arbitraria è interessante. Come si vede da quanto precede, il caso di caratteristica due presenta però qualche complicazione. Ci limiteremo quindi al caso $\operatorname{char} F \neq 2$.[1] In questo caso le nostre osservazioni mostrano che la teoria equivale a quella delle forme bilineari simmetriche. Presenteremo i risultati talvolta in termini di forme quadratiche, talvolta in termini di forme bilineari simmetriche. Dimostriamo anzitutto il seguente teorema di diagonalizzazione.

Teorema 6.5. *Sia B una forma bilineare simmetrica su uno spazio vettoriale V su un campo F di caratteristica diversa da due. Esiste allora una base $(u_1, \ldots, u_r, z_1, \ldots, z_{n-r})$ rispetto alla quale la matrice di B è della forma*

$$\operatorname{diag}\{b_1, \ldots, b_r, 0, \ldots, 0\}, \qquad b_i \neq 0. \tag{25}$$

Una base $(u_1, u_2, \ldots, u_n)$ tale che $B(u_i, u_j) = 0$ per ogni $i \neq j$ si chiama *base ortogonale* di V (rispetto a B oppure a una forma quadratica assegnata Q). Il Teorema 6.5 afferma che tale base esiste.

Dimostrazione. Daremo una dimostrazione costruttiva, dovuta a Lagrange. Osserviamo anzitutto che, se $B = 0$, qualsiasi base $(z_1, z_2, \ldots, z_n)$ soddisfa la condizione. Supponiamo dunque $B \neq 0$. Affermiamo che esiste $u \neq 0$ tale che $B(u,u) \neq 0$. In caso contrario, per ogni u, v si avrebbe

$$2B(u,v) = B(u+v, u+v) - B(u,u) - B(v,v) = 0,$$

contro l'ipotesi $B \neq 0$. Scegliamo quindi u_1 tale che $B(u_1, u_1) = b_1 \neq 0$. Questo avvia la costruzione induttiva della base cercata. Supponiamo di avere già determinato vettori linearmente indipendenti $(u_1, \ldots, u_k)$ tali che $B(u_i, u_j) = \delta_{ij} b_i$, con $b_i \neq 0$, $\delta_{ii} = 1$ e $\delta_{ij} = 0$ per $i \neq j$. Sia V_k il sottospazio generato da questi u_i. Mostriamo che $V = V_k \oplus V_k^\perp$. Poiché la matrice della restrizione di B a V_k è $\operatorname{diag}\{b_1, \ldots, b_k\}$, questa forma è non degenere; dunque $V_k \cap V_k^\perp = 0$. Sia $x \in V$ e si ponga

$$y = x - \sum_{i=1}^{k} B(x, u_i) b_i^{-1} u_i.$$

Allora $B(y, u_j) = B(x, u_j) - B(x, u_j) b_j^{-1} B(u_j, u_j) = 0$. Perciò $y \in V_k^\perp$ e $x = y + \sum_i B(x, u_i) b_i^{-1} u_i \in V_k + V_k^\perp$. Ne segue $V = V_k \oplus V_k^\perp$. Se la restrizione di B a $V_k^\perp$ è nulla, prendiamo $r = k$ e scegliamo come $(z_1, \ldots, z_{n-k})$ una base qualsiasi di $V_k^\perp$. Altrimenti scegliamo $u_{k+1} \in V_k^\perp$ tale che $b_{k+1} = B(u_{k+1}, u_{k+1}) \neq 0$. Allora $(u_1, \ldots, u_{k+1})$ è un sistema linearmente indipendente con le stesse proprietà di $(u_1, \ldots, u_k)$. Ripetendo il procedimento si ottiene infine una base del tipo richiesto. □

[1] Per il caso di caratteristica due si vedano C. Chevalley, *The Algebraic Theory of Spinors*, New York, Columbia University Press, 1954, e J. Dieudonné, *La Géométrie des Groupes Classiques*, 2ª ed., Springer, 1963.

La dimostrazione suggerisce alcune osservazioni. Anzitutto essa è effettivamente costruttiva. Per indicare un modo meccanico di eseguirla, supponiamo di avere un sistema di vettori $\{e_i\}$ che genera V; potrebbe trattarsi, ad esempio, di una base. Se qualche $B(e_i, e_i) \neq 0$, possiamo scegliere $u_1 = e_i$. Altrimenti, supposto $B \neq 0$, si ha $B(e_i, e_j) \neq 0$ per qualche coppia $e_i \neq e_j$. Allora $B(e_i + e_j, e_i + e_j) = 2B(e_i, e_j) \neq 0$ e possiamo prendere $u_1 = e_i + e_j$. Supponiamo ora di avere già determinato $(u_1, \dots, u_k)$ e V_k come nella dimostrazione. Per ogni e_j del sistema di generatori assegnato si ponga

$$f_j = e_j - \sum_{i=1}^{k} B(e_j, u_i) b_i^{-1} u_i.$$

Si vede allora che gli f_j generano $V_k^\perp$, e possiamo ripetere il procedimento applicato a V.

Un altro punto di notevole interesse teorico è che b_1 può essere scelto uguale a un qualunque elemento non nullo di F *rappresentato* da B, nel senso che l'equazione $B(u_1, u_1) = b_1$ ammette soluzione. Analogamente b_{k+1} è un qualsiasi elemento non nullo rappresentato dalla restrizione di B a $V_k^\perp$. Questa libertà nella scelta dei b_i permetterà in certi casi di porre parecchi di essi uguali a uno. Osserviamo inoltre che, scelta la base come nel teorema, il radicale è generato da $z_1, \dots, z_{n-r}$. Infatti, se $z = \sum_i c_i u_i + \sum_j d_j z_j$, si ha $B(z, z_j) = 0$ per ogni z_j, e $B(u_i, z) = 0$ se e solo se $c_i = 0$. Quindi $z \in V^\perp$ se e solo se $z = \sum_j d_j z_j$.

Osserviamo infine che dalla dimostrazione ricaviamo un'altra prova del fatto che, se la restrizione di B a un sottospazio U è non degenere, allora $V = U \oplus U^\perp$ (esercizio 2, p. 348). Possiamo infatti scegliere una base ortogonale $(u_1, \dots, u_k)$ di U. La dimostrazione mostra che possiamo completarla a una base ortogonale $(u_1, \dots, u_k, u_{k+1}, \dots, u_n)$ di V. Si avrà evidentemente $U^\perp = Fu_{k+1} + \cdots + Fu_n$ e $V = U \oplus U^\perp$.

Dal punto di vista matriciale, il Teorema 6.5 fornisce una matrice diagonale congruente a una matrice simmetrica assegnata. La difficoltà è che tale matrice diagonale non è unica. Ad esempio, sostituendo u_i con $c_i u_i \neq 0$, si sostituisce b_i con $c_i^2 b_i$. Si potrebbe quindi sperare, al più, che i b_i siano determinati a meno di quadrati. Ma non vale neppure questo, poiché al posto di b_1 si può prendere qualsiasi elemento non nullo della forma $\sum_i b_i c_i^2$, e questo può non essere il prodotto di un quadrato per uno dei b_i (si veda l'esercizio 2 alla fine del paragrafo). Il problema di classificare le matrici simmetriche rispetto alla congruenza, o equivalentemente le forme bilineari simmetriche rispetto all'equivalenza, è in generale assai difficile e dipende dalle proprietà aritmetiche del campo di base. Nel caso di $\mathbb{Q}$, o più generalmente di un'estensione algebrica F di $\mathbb{Q}$, si ha una soluzione completa dovuta a Minkowski e Hasse. Per $\mathbb{Q}$, il risultato di Minkowski afferma che la congruenza vale se e solo se vale su $\mathbb{R}$ e su certe estensioni di $\mathbb{Q}$, i campi p-adici $\mathbb{Q}_p$, definiti per ogni numero primo p.[2] Il teorema di Minkowski e Hasse è piuttosto profondo. Esistono tuttavia alcuni importanti tipi di campi per i quali il problema della congruenza si risolve facilmente. Fra questi figurano (1) i campi algebricamente chiusi, (2) i campi reali chiusi e (3) i campi finiti. Li consideriamo ora nell'ordine.

1. Campi algebricamente chiusi. In questi campi ogni b_i si può sostituire con 1, cosicché ogni matrice simmetrica è congruente a una matrice della forma $\operatorname{diag}\{1, \dots, 1, 0, \dots, 0\}$. Il numero degli 1 è il rango della matrice diagonale. Poiché per p invertibile le matrici $ps\,{}^t p$ e s hanno lo stesso rango, esso è anche il rango della matrice assegnata s (ed è uguale a $n - \dim \operatorname{rad} B$). Dal nostro risultato segue evidentemente il teorema seguente.

Teorema 6.6. *Se F è algebricamente chiuso e di caratteristica diversa da due, due matrici simmetriche di $M_n(F)$ sono congruenti se e solo se hanno lo stesso rango.*

[2] Questi campi sono trattati nel volume II.

2. Campi reali chiusi.[3] Supponiamo che $F = R$ sia reale chiuso. Poiché gli elementi positivi ammettono radici quadrate, nella (25) i b_i positivi possono essere sostituiti con 1 e quelli negativi con -1. Riordinando gli u_i, possiamo supporre che la matrice canonica sia $\operatorname{diag}\{1,\dots,1,-1,\dots,-1,0,\dots,0\}$. Mostreremo che il numero dei $+1$, e quindi anche quello dei -1, è un invariante. Equivalentemente sarà un invariante la *segnatura*, definita come $p-q$, dove p è il numero dei $+1$ e q quello dei -1. Ciò seguirà dal teorema seguente.

Teorema 6.7 (Sylvester). *Sia F un campo ordinato e supponiamo che le matrici diagonali*

$$\operatorname{diag}\{b_1,\dots,b_r,0,\dots,0\}, \qquad \operatorname{diag}\{b'_1,\dots,b'_r,0,\dots,0\}, \qquad b_i, b'_i \neq 0,$$

siano congruenti. Il numero dei b_i positivi è allora uguale al numero dei b'_i positivi.

Dimostrazione. Possiamo supporre che esse siano matrici di una forma bilineare simmetrica su uno spazio vettoriale di dimensione n sul campo F e che siano positivi i primi p fra i b_i e i primi p' fra i b'_i, mentre i rimanenti sono negativi. Siano $(u_1,\dots,u_r,z_1,\dots,z_{n-r})$ e $(u'_1,\dots,u'_r,z'_1,\dots,z'_{n-r})$ le basi che danno le matrici del teorema. Se $z \in \sum_{i=1}^{p} Fu_i$, si ha $z = \sum_{i=1}^{p} a_i u_i$; dunque $B(z,z) = \sum_{i=1}^{p} a_i^2 b_i > 0$ per $z \neq 0$. Se invece $z \in \sum_{i=p'+1}^{r} Fu'_i + \sum_j Fz'_j$, si ha analogamente $B(z,z) \leq 0$. I due spazi $\sum_{i=1}^{p} Fu_i$ e $\sum_{i=p'+1}^{r} Fu'_i + \sum_j Fz'_j$ hanno perciò in comune soltanto il vettore nullo. La somma delle loro dimensioni non supera quindi n, in virtù della nota formula $\dim(U_1+U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$ per due sottospazi U_i di V. Otteniamo così $p + (n-p') \leq n$, da cui $p \leq p'$. Per simmetria $p' \leq p$, e dunque $p = p'$. □

Da questo risultato segue il teorema seguente.

Teorema 6.8. *Due matrici diagonali di $M_n(R)$, con R campo reale chiuso, sono congruenti se e solo se hanno lo stesso rango e la stessa segnatura (numero degli elementi positivi meno numero degli elementi negativi).*

Prima di passare alle forme bilineari simmetriche su un campo finito, diamo alcune definizioni e osservazioni di interesse generale. Una forma bilineare simmetrica non degenere B si dice *isotropa*, o *forma nulla*, se esiste un vettore $u \neq 0$ tale che $B(u,u) = 0$. Un tale vettore si dice *isotropo*. Una forma che non è isotropa si dice *anisotropa*. Se B è isotropa, essa è *universale*, nel senso che $B(v,v) = b$ ammette soluzione per ogni $b \neq 0$ in F. Infatti, se $B(u,u) = 0$ per $u \neq 0$, la non degenerazione di B assicura l'esistenza di w tale che $B(u,w) = \frac{1}{2}$. Per $v = au + w$ si ha allora

$$B(v,v) = a^2 B(u,u) + 2aB(u,w) + B(w,w) = a + B(w,w).$$

Scegliendo $a = b - B(w,w)$ otteniamo dunque $B(v,v) = b$. Passiamo ora al terzo caso.

3. Campi finiti. Mostreremo che queste forme possono essere classificate mediante i loro discriminanti. Dimostriamo dapprima il lemma seguente.

Lemma. *Ogni forma bilineare simmetrica non degenere B su uno spazio vettoriale V di dimensione almeno due su un campo finito F di caratteristica diversa da due è universale.*

Dimostrazione. Basta dimostrare il risultato per le forme binarie, ossia nel caso $\dim V = 2$; e poiché nel caso isotropo abbiamo già dimostrato l'universalità, possiamo supporre B anisotropa. Possiamo anche supporre di avere per la forma una matrice diagonale. Il problema si riduce allora a dimostrare che, se $ab \neq 0$ e $ax^2+by^2 \neq 0$ per ogni $(x,y) \neq (0,0)$, l'equazione $ax^2+by^2 = c$ è risolubile per ogni $c \neq 0$. Dividendo per a possiamo prendere $a = 1$. Ora, dalla condizione $x^2 + by^2 \neq 0$ segue che $-b$ non è un quadrato; inoltre $x^2 + by^2$ è la norma di $x + y\sqrt{-b}$ nell'estensione quadratica K/F, con $K = F(\sqrt{-b})$. Se $|F| = q$, si ha $|K| = q^2$, e l'applicazione $u \mapsto u^q$ è un automorfismo di K/F diverso dall'identità. Dunque $N_{K/F}(u) = uu^q = u^{q+1}$. Dobbiamo quindi mostrare che per ogni $c \neq 0$ di F esiste $u \in K^*$ tale che $u^{q+1} = c$. Ora

[3] Sono stati definiti nel § 5.1, p. 308. Il caso particolare più importante è il campo $\mathbb{R}$ dei numeri reali.

K^* è ciclico di ordine q^2-1, e $u \mapsto u^{q+1}$ è un omomorfismo di K^* in F^*. Il suo nucleo è il sottogruppo degli u tali che $u^{q+1}=1$. Dato che il gruppo è ciclico, questo sottogruppo ha ordine $q+1$. L'immagine ha dunque ordine $(q^2-1)/(q+1)=q-1$, sicché l'omomorfismo è suriettivo. La dimostrazione è completa (cfr. l'esercizio 1, p. 300, dove si enuncia un risultato più generale). □

Possiamo ora dimostrare il teorema seguente.

Teorema 6.9. *Ogni forma bilineare simmetrica non degenere su uno spazio vettoriale V su un campo finito di caratteristica diversa da due ammette una matrice della forma* $\operatorname{diag}\{1,1,\dots,1,d\}$. *Equivalentemente, ogni matrice simmetrica invertibile a elementi in un tale campo finito è congruente a una della forma* $\operatorname{diag}\{1,1,\dots,1,d\}$. *Inoltre, due matrici simmetriche invertibili a elementi in un tale campo finito sono congruenti se e solo se hanno lo stesso discriminante.*

Dimostrazione. Il procedimento di diagonalizzazione di Lagrange e il lemma precedente mostrano che si possono prendere $b_1=b_2=\cdots=b_{n-1}=1$. Ciò dimostra la prima affermazione e quella equivalente sulle matrici. Poiché matrici congruenti hanno lo stesso discriminante (il discriminante della forma bilineare associata), per l'ultima affermazione basterà provare che le matrici diagonali $\operatorname{diag}\{1,1,\dots,1,d_i\}$, $i=1,2$, $d_i\neq 0$, sono congruenti quando i loro discriminanti coincidono. Ma il discriminante è d_iF^{*2}, e $d_1F^{*2}=d_2F^{*2}$ implica $d_2=a^2d_1$ con $a\in F^*$, da cui segue la congruenza delle due matrici diagonali. □

Esercizi

1. Trovare una matrice diagonale d congruente in $M_3(\mathbb{Q})$ a

$$s=\begin{pmatrix}-2 & 3 & 5\\ 3 & 1 & -1\\ 5 & -1 & 4\end{pmatrix}.$$

   Determinare anche una matrice p tale che $ps\,{}^tp=d$.

2. Mostrare che $\operatorname{diag}\{1,1\}$ e $\operatorname{diag}\{5,5\}$ sono congruenti in $M_2(\mathbb{Q})$.

3. Mostrare che la forma bilineare simmetrica B su uno spazio V su $\mathbb{R}$ è *definita positiva*, nel senso che $B(u,u)>0$ per ogni $u\neq 0$, se e solo se fra le sue matrici figura la matrice identità 1. Usare la riduzione di Lagrange (che in questo caso si chiama *procedimento di ortogonalizzazione di Schmidt*) per dimostrare che, se s è una matrice di una forma bilineare simmetrica definita positiva, esiste una matrice triangolare p, con zeri al di sopra della diagonale principale, tale che $ps\,{}^tp=1$, ossia $s=q\,{}^tq$ con $q=p^{-1}$.

4. Nelle ipotesi dell'esercizio 3, chiamare *cartesiana* una base $(u_1,u_2,\dots,u_n)$ se $B(u_i,u_j)=\delta_{ij}$. Mostrare che, se $(v_1,v_2,\dots,v_n)$ è un'altra base cartesiana, la matrice che collega le due basi è ortogonale ($o\,{}^to=1$). Usare il risultato dell'esercizio 3 per mostrare che ogni matrice invertibile $m\in M_n(\mathbb{R})$ si scrive nella forma po, con p triangolare e o ortogonale.

5. Dimostrare che l'insieme delle funzioni polinomiali su V può essere definito come il sottoanello dell'anello delle applicazioni da V in F generato dalle funzioni lineari. Qui l'addizione e la moltiplicazione delle applicazioni da V in F sono le usuali: $(f+g)(x)=f(x)+g(x)$, $(fg)(x)=f(x)g(x)$. Si ottiene così una definizione intrinseca delle funzioni polinomiali.

6. Sia Q una forma quadratica non degenere su uno spazio vettoriale di dimensione $n\geq 3$ su un campo finito. Mostrare che Q è isotropa.

6.4 Nozioni fondamentali di geometria ortogonale

Introduciamo le definizioni fondamentali della geometria ortogonale, ossia dello studio di uno spazio vettoriale rispetto a una forma bilineare simmetrica non degenere. Supponiamo $\operatorname{char} F \neq 2$, sicché è indifferente trattare forme bilineari simmetriche B o forme quadratiche Q. A una forma quadratica Q associamo la forma bilineare simmetrica $B(x,y) = Q(x+y) - Q(x) - Q(y)$, e a una forma bilineare simmetrica B associamo la forma quadratica $Q(x) = \frac{1}{2}B(x,x)$. Diremo che Q è *non degenere* se B è non degenere.

Siano (V_i, Q_i), $i = 1, 2$, due coppie costituite da uno spazio vettoriale V_i e da una forma quadratica Q_i su V_i. Un'*isometria* η di (V_1, Q_1) su (V_2, Q_2) è un'applicazione lineare biiettiva di V_1 su V_2 tale che $Q_2(\eta x) = Q_1(x)$ per ogni $x \in V_1$. Indicando con B_i la forma bilineare simmetrica associata a Q_i, ne segue $B_2(\eta x, \eta y) = B_1(x,y)$ per $x, y \in V_1$, poiché

$$\begin{aligned} B_2(\eta x, \eta y) &= Q_2(\eta(x+y)) - Q_2(\eta x) - Q_2(\eta y) \\ &= Q_1(x+y) - Q_1(x) - Q_1(y) = B_1(x,y). \end{aligned}$$

Anche il viceversa è immediato, dal momento che $Q_i(x) = \frac{1}{2}B_i(x,x)$. Se B_1 è non degenere, la condizione di iniettività è superflua. Infatti, se $\eta\colon V_1 \to V_2$ è lineare e soddisfa $Q_2(\eta x) = Q_1(x)$ per $x \in V_1$, allora è necessariamente iniettiva: da $B_2(\eta x, \eta y) = B_1(x,y)$ e $\eta x = 0$ segue $B_1(x,y) = 0$ per ogni y, e dunque $x = 0$ per la non degenerazione di B_1.

Se esiste un'isometria di (V_1, Q_1) su (V_2, Q_2), si dicono *equivalenti* le forme quadratiche Q_1 e Q_2, come pure le forme bilineari simmetriche associate B_1 e B_2.

Se Q è una forma quadratica non degenere su V, un'isometria di V su se stesso si chiama *trasformazione ortogonale* di V, o di (V, Q). Ogni endomorfismo lineare η di V tale che $Q(\eta x) = Q(x)$ per $x \in V$ è ortogonale. Infatti, abbiamo visto che questa condizione implica l'iniettività di η, e poiché supponiamo sempre che V abbia dimensione finita, η è anche suriettiva. Se η è ortogonale, lo è anche η^{-1}; se η_1 e η_2 sono ortogonali, lo è anche $\eta_1\eta_2$. Perciò l'insieme $\mathrm{O}(V, Q)$ delle trasformazioni ortogonali è un sottogruppo del gruppo delle trasformazioni lineari biiettive di V. Esso si chiama *gruppo ortogonale* di V rispetto a Q.

Sia $(e_1, e_2, \ldots, e_n)$ una base di V e sia $\eta \in \mathrm{O}(V, Q)$. Si ha $B(\eta e_i, \eta e_j) = B(e_i, e_j)$ per ogni $i, j = 1, \ldots, n$. Viceversa, se un endomorfismo lineare η soddisfa queste condizioni, allora per ogni $x = \sum_i a_i e_i$ si ha

$$Q(\eta x) = \tfrac{1}{2} \sum_{i,j} a_i a_j B(\eta e_i, \eta e_j) = \tfrac{1}{2} \sum_{i,j} a_i a_j B(e_i, e_j) = Q(x).$$

Pertanto un endomorfismo lineare η di V è ortogonale se e solo se

$$B(\eta e_i, \eta e_j) = B(e_i, e_j), \qquad 1 \leq i, j \leq n. \tag{26}$$

Sia ora $b = (B(e_i, e_j))$ la matrice di B rispetto alla base $(e_1, e_2, \ldots, e_n)$ e sia $\eta e_i = \sum_j h_{ij} e_j$. Le condizioni (26) diventano $\sum_{k,l} h_{ik} b_{kl} h_{jl} = b_{ij}$ per ogni i, j. In forma matriciale esse sono

$$h b\, {}^t h = b, \qquad h = (h_{ij}). \tag{27}$$

Questa è dunque la condizione necessaria e sufficiente sulla matrice h di η rispetto alla base $(e_1, \ldots, e_n)$ affinché η sia ortogonale. Prendendo i determinanti nella (27), si ottiene $(\det h)^2 \det b = \det b$; poiché $\det b \neq 0$, la matrice di una trasformazione ortogonale rispetto a qualsiasi base soddisfa $\det h = \pm 1$. Dato che il determinante della matrice di una trasformazione lineare è invariante per cambiamento di base, se vale 1 o -1 in una base, avrà lo stesso valore in ogni altra base. Una trasformazione ortogonale si dice *propria*, o *rotazione*, se $\det h = 1$;

altrimenti si dice *impropria*. Rispetto a una base ortogonale di V la matrice b di B è diagonale. Qualunque matrice diagonale i cui elementi diagonali siano 1 o -1 soddisfa allora la (27) e determina una trasformazione ortogonale. Essa è propria o impropria secondo che il numero dei -1 sia pari o dispari. L'insieme $\mathrm{O}^+(V,Q)$ delle rotazioni è evidentemente un sottogruppo normale di indice due in $\mathrm{O}(V,Q)$.

A ogni vettore u tale che $Q(u) \neq 0$ possiamo associare una trasformazione ortogonale S_u definita da

$$S_u x = x - \frac{B(x,u)}{Q(u)} u. \tag{28}$$

Poiché $x \mapsto x$ e $x \mapsto B(x,u)v$ sono lineari per ogni u, v, S_u è lineare. Inoltre

$$Q(S_u x) = Q(x) - \frac{B(x,u)^2}{Q(u)} + \frac{B(x,u)^2}{Q(u)^2} Q(u) = Q(x).$$

Quindi S_u è ortogonale. Ora $S_u u = u - (B(u,u)/Q(u))u = u - 2u = -u$, e per $v \perp u$ si ha $S_u v = v$. Poiché $B(u,u) \neq 0$, si ha la decomposizione $V = Fu \oplus (Fu)^\perp$. Le proprietà appena indicate descrivono dunque completamente S_u: questa trasformazione lineare coincide con l'identità su $(Fu)^\perp$ e manda u in $-u$. Chiameremo S_u la *simmetria* determinata da u. Rispetto a una base di V costituita da u e da una base di $(Fu)^\perp$, la matrice di S_u è $\mathrm{diag}\{-1, 1, \ldots, 1\}$. Di conseguenza S_u è impropria. È anche chiaro che $S_u^2 = 1$.

Per ogni trasformazione ortogonale η vale

$$\eta S_u \eta^{-1} = S_{\eta u}. \tag{29}$$

Infatti

$$\begin{aligned} \eta S_u \eta^{-1} x &= \eta \left(\eta^{-1} x - \frac{B(\eta^{-1}x, u)}{Q(u)} u \right) \\ &= \eta \left(\eta^{-1} x - \frac{B(\eta\eta^{-1}x, \eta u)}{Q(u)} u \right) = x - \frac{B(x, \eta u)}{Q(u)} \eta u \\ &= x - \frac{B(x, \eta u)}{Q(\eta u)} \eta u = S_{\eta u} x. \end{aligned}$$

Abbiamo usato l'identità $B(\eta x, y) = B(\eta^{-1}\eta x, \eta^{-1} y) = B(x, \eta^{-1} y)$, valida per ogni trasformazione ortogonale. Abbiamo definito l'aggiunta T' di una trasformazione lineare T rispetto a B mediante la condizione $B(Tx, y) = B(x, T'y)$ (esercizio 4, p. 349). La relazione ottenuta afferma dunque che l'aggiunta di una trasformazione ortogonale coincide con la sua inversa. È immediato che questa proprietà, ossia $TT' = 1$, implica a sua volta che T sia ortogonale.

Un'osservazione molto importante sulle aggiunte è la seguente. Se una trasformazione lineare T lascia stabile un sottospazio U, ossia $T(U) \subseteq U$, allora T' lascia stabile $U^\perp$. Infatti, se $v \in U^\perp$, si ha $B(u,v) = 0$ per ogni $u \in U$, e quindi $B(u, T'v) = B(Tu, v) = 0$. Ne segue che, se η è ortogonale e lascia stabile U, allora $\eta U = U$ e quindi $\eta^{-1} U = U$. Poiché $\eta' = \eta^{-1}$, η' lascia stabile U; di conseguenza $\eta = (\eta')'$ lascia stabile $U^\perp$.

Diremo che V è *somma diretta ortogonale* dei sottospazi $U_1, U_2, \ldots, U_r$ se $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$ e $U_i \perp U_j$ per ogni $i \neq j$. In tal caso scriveremo

$$V = U_1 \perp U_2 \perp \cdots \perp U_r. \tag{30}$$

Se $x = x_1 + \cdots + x_r$, con $x_i \in U_i$, si ha $Q(x) = Q(x_1) + \cdots + Q(x_r)$. Se ora abbiamo una seconda decomposizione $V = U_1' \perp U_2' \perp \cdots \perp U_r'$ e isometrie $\eta_i \colon U_i \to U_i'$, la trasformazione lineare η

tale che $\eta|_{U_i} = \eta_i$ soddisfa

$$Q(\eta x) = Q\left(\sum_i \eta_i x_i\right) = \sum_i Q(\eta_i x_i) = \sum_i Q(x_i) = Q(x).$$

Dunque η è ortogonale. È pure chiaro che i sottospazi U_i della (30) sono non degeneri, ossia la restrizione di B a U_i è non degenere.

Un sottospazio U è *isotropo* se contiene un vettore isotropo ($u \neq 0$, $Q(u) = 0$). Si dice *totalmente isotropo* se la restrizione di Q a U è nulla o, equivalentemente, se $U \subseteq U^\perp$.

Uno spazio bidimensionale V non degenere e isotropo si chiama *piano iperbolico*. Il seguente teorema dice pressoché tutto ciò che si può dire sui piani iperbolici.

Teorema 6.10. *(1) Per uno spazio vettoriale bidimensionale V munito di una forma quadratica Q sono equivalenti le seguenti condizioni: (i) V è un piano iperbolico; (ii) V ha una base (u, v) costituita da una* coppia iperbolica *di vettori, nel senso che*

$$B(u,u) = 0 = B(v,v), \qquad B(u,v) = 1 = B(v,u); \tag{31}$$

(iii) il discriminante di B è $-1F^{*2}$. (2) Due piani iperbolici qualsiasi sono isometrici. (3) Ogni piano iperbolico contiene esattamente due sottospazi unidimensionali totalmente isotropi. (4) Il gruppo delle rotazioni di un piano iperbolico V è isomorfo al gruppo moltiplicativo F^* del campo F, e ogni trasformazione ortogonale impropria di V è una simmetria.

Dimostrazione. (1) Se V è un piano iperbolico, contiene un vettore $u \neq 0$ tale che $Q(u) = 0$ e, poiché $V^\perp = 0$, anche un vettore v tale che $B(u, v) \neq 0$. Poiché $B(u, u) = 0$, v non è multiplo di u; quindi (u, v) è una base. Sostituendo v con un suo multiplo, possiamo supporre $B(u, v) = 1$. Inoltre per $a \in F$ si ha $Q(v + au) = Q(v) + a$. Sostituendo v con $v - Q(v)u$, otteniamo $Q(v) = 0$ oltre a $B(u, v) = 1$. La base (u, v) soddisfa ora la (31), dunque (i) implica (ii). Supponiamo (ii). La matrice definita dalla (31) ha determinante -1, quindi il discriminante di B è $-1F^{*2}$. Supponiamo ora che esso sia $-1F^{*2}$. Esiste allora una base (u_1, u_2) rispetto alla quale la matrice di B è $\operatorname{diag}\{b_1, b_2\}$ e $b_1 b_2 = -c^2 \neq 0$, $c \in F$. Ponendo $x = cu_1 + b_1 u_2$, si ha $Q(x) = \frac{1}{2}(c^2 b_1 + b_1^2 b_2) = 0$. Pertanto V è un piano iperbolico e (iii) implica (i).

(2) Due piani iperbolici hanno basi (u, v) e (u', v') che sono coppie iperboliche. È evidente che l'applicazione lineare $u \mapsto u'$, $v \mapsto v'$ è un'isometria.

(3) Sia (u, v) una base costituita da una coppia iperbolica. Allora $Q(au + bv) = ab$. Quindi $au + bv$ è isotropo se e solo se $a = 0$, $b \neq 0$, oppure $a \neq 0$, $b = 0$. Dunque Fu e Fv sono gli unici sottospazi unidimensionali totalmente isotropi di V.

(4) Sia η una trasformazione ortogonale del piano iperbolico V e siano Fu, Fv i due sottospazi unidimensionali totalmente isotropi. Allora o $\eta(Fu) = Fu$ e $\eta(Fv) = Fv$, oppure $\eta(Fu) = Fv$ e $\eta(Fv) = Fu$. Nel primo caso $\eta u = au$ e $\eta v = bv$. Poiché $abB(u, v) = B(\eta u, \eta v) = B(u, v)$ e $B(u, v) \neq 0$, segue $ab = 1$. Dunque $\eta u = au$ e $\eta v = a^{-1}v$, e η è evidentemente una rotazione. Nel secondo caso $\eta u = av$ e $\eta v = bu$, con $b = a^{-1}$ come prima. Questa volta η è impropria. D'altra parte, per ogni $a \neq 0$, le applicazioni lineari $u \mapsto au$, $v \mapsto a^{-1}v$ e $u \mapsto av$, $v \mapsto a^{-1}u$ sono rispettivamente una rotazione e una trasformazione ortogonale impropria. L'applicazione che a $a \in F^*$ associa la rotazione $u \mapsto au$, $v \mapsto a^{-1}v$ è quindi un isomorfismo di F^* su $\mathrm{O}^+(V, Q)$. Infine, se η è una trasformazione ortogonale impropria, cosicché $\eta u = av$ e $\eta v = a^{-1}u$, si ha $\eta(u + av) = u + av$ e $\eta(u - av) = -(u - av)$. Pertanto η è la simmetria S_{u-av}. □

Esercizi

1. Mostrare che, se η è una trasformazione ortogonale e $V_1 = \{x \mid \eta x = x\}$, allora $\dim V = \dim V_1 + \dim(1 - \eta)V$. Mostrare inoltre che $V_1 = ((1 - \eta)V)^\perp$ e quindi $V_1^\perp = (1 - \eta)V$.

2. Sia η una trasformazione ortogonale con $\dim V_1 \geq \dim V - 1$, dove V_1 è definito come nell'esercizio 1. Mostrare che $\eta = 1$ oppure η è una simmetria.

3. Sia (u, v) una coppia iperbolica e sia $w \in (Fu + Fv)^\perp$ non isotropo. Verificare che la trasformazione lineare ρ definita da

$$u \mapsto u, \qquad v \mapsto v - Q(w)u - w, \qquad x \mapsto x + B(x, w)u \quad (x \in (Fu + Fv)^\perp)$$

coincide con $S_w S_{w-Q(w)u}$. (Si osservi che $Q(w - Q(w)u) \neq 0$.)

4. Sia V munito di una forma quadratica non degenere Q e sia Σ un insieme di endomorfismi lineari di V chiuso rispetto al passaggio all'aggiunta: se $T \in \Sigma$, la sua aggiunta T' rispetto a B appartiene a Σ. Mostrare che, se U è un sottospazio stabile per Σ ($TU \subseteq U$ per ogni $T \in \Sigma$), anche $U^\perp$ è stabile per Σ.

5. Sia Q anisotropa. Mostrare che 0 è l'unico endomorfismo lineare nilpotente di V autoaggiunto rispetto a Q.

6. Una trasformazione lineare T si dice *unipotente* se $T - 1 = Z$ è nilpotente. Mostrare che, se Q è anisotropa, 1 è l'unica trasformazione ortogonale unipotente di V rispetto a Q. Verificare che la trasformazione ρ definita nell'esercizio 3 è unipotente. Dedurne che, se $\dim V \geq 3$ e Q è isotropa, $\mathrm{O}(V, Q)$ contiene trasformazioni ortogonali unipotenti diverse dall'identità.

7. (Malcev.) Sia T una trasformazione lineare nilpotente di V autoaggiunta rispetto a B (non degenere). Mostrare che V è somma diretta ortogonale di sottospazi V_i, dove V_i ha una base $(z_i, Tz_i, \ldots, T^{n_i-1}z_i)$ e la matrice di B rispetto a questa base ha la forma

$$\begin{pmatrix} 0 & \cdots & 0 & b \\ \vdots & \cdots & b & 0 \\ 0 & b & \cdots & \vdots \\ b & 0 & \cdots & 0 \end{pmatrix}, \qquad b \neq 0.$$

Dedurne che in (V, Q) esistono trasformazioni lineari nilpotenti autoaggiunte diverse da zero se e solo se Q è isotropa.

8. Sia T un endomorfismo lineare di uno spazio vettoriale V di dimensione finita. Mostrare che esiste su V una forma bilineare simmetrica non degenere B rispetto alla quale T è autoaggiunto. (*Suggerimento.* Basta supporre T «ciclico», ossia $V = F[T]u$ per un opportuno vettore u. Si ha allora una base $(u_1, \ldots, u_n)$ di V su F tale che $Tu_i = u_{i+1}$, $1 \leq i \leq n-1$, e $Tu_n = \sum_{i=1}^n \alpha_i u_i$. Sia B la forma bilineare su V tale che $B(u_i, u_j) = \lambda_{i+j-n} \in F$, dove $\lambda_1 = 1$ e $\lambda_k = 0$ per $k \leq 0$. Mostrare che B è simmetrica e non degenere e che si possono scegliere le λ in modo che T sia autoaggiunto rispetto a B.)

6.5 Il teorema di cancellazione di Witt

Dimostreremo ora un teorema fondamentale sulle forme quadratiche su uno spazio vettoriale V su F, con $\mathrm{char}\, F \neq 2$. Curiosamente, malgrado la sua importanza e il suo carattere elementare, esso fu scoperto piuttosto tardi nello sviluppo della teoria. Fra le sue conseguenze più importanti figurano la riduzione del problema di classificazione delle forme quadratiche al caso delle forme anisotrope e la definizione di un invariante numerico, detto indice di Witt, che generalizza la nozione di segnatura di una forma quadratica reale. Il risultato implica anche un teorema sul

prolungamento delle isometrie a trasformazioni ortogonali. I preparativi del paragrafo precedente ci permettono di passare subito alla dimostrazione.

Teorema di cancellazione di Witt. *Sia Q una forma quadratica non degenere su uno spazio vettoriale V su un campo F di caratteristica diversa da due. Siano U_1 e U_2 sottospazi non degeneri isometrici, ossia tali che esista un'isometria fra loro. Allora $U_1^\perp$ e $U_2^\perp$ sono isometrici.*

(Poiché $V = U_1 \oplus U_1^\perp = U_2 \oplus U_2^\perp$, si tratta davvero di un teorema di «cancellazione».)

Dimostrazione. Indichiamo l'isometria con $\sim$. Per ipotesi $U_1 \sim U_2$ e le restrizioni di Q a U_1 e a U_2 sono non degeneri. Vogliamo mostrare che $U_1^\perp \sim U_2^\perp$. Procediamo per induzione su $\dim U_i$. Supponiamo anzitutto $U_i = Fu_i$ con $Q(u_i) \neq 0$. Possiamo supporre $Q(u_1) = Q(u_2)$. Si ha $Q(u_1 \pm u_2) = 2Q(u_1) \pm B(u_1, u_2)$, perciò $Q(u_1 + u_2) \neq 0$ oppure $Q(u_1 - u_2) \neq 0$. Supponiamo dapprima $Q(u_1 + u_2) \neq 0$ e consideriamo la simmetria $S_{u_1+u_2}$. Poiché $B(u_1 + u_2, u_1 - u_2) = 2Q(u_1) - 2Q(u_2) = 0$, si ha $(u_1 - u_2) \perp (u_1 + u_2)$, e pertanto $S_{u_1+u_2}(u_1 - u_2) = u_1 - u_2$. D'altra parte $S_{u_1+u_2}(u_1 + u_2) = -(u_1 + u_2)$. Quindi $S_{u_1+u_2}u_1 = -u_2$ e di conseguenza $S_{u_1+u_2}(Fu_1)^\perp = (Fu_2)^\perp$, ossia $U_1^\perp \sim U_2^\perp$. Analogamente, se $Q(u_1 - u_2) \neq 0$, si usa $S_{u_1-u_2}$, che manda $u_1 + u_2$ in sé e $u_1 - u_2$ in $u_2 - u_1$. Allora $S_{u_1-u_2}u_1 = u_2$ e $S_{u_1-u_2}(Fu_1)^\perp = (Fu_2)^\perp$.

Supponiamo ora che il risultato valga per i sottospazi di dimensione $\dim U_i - 1 \geq 1$. Possiamo scegliere in U_1 un vettore non isotropo u_1 e scrivere $U_1 = Fu_1 \perp W_1$. Allora W_1 è non degenere. Applicando un'isometria di U_1 su U_2 otteniamo $U_2 = Fu_2 \perp W_2$ con $Fu_1 \sim Fu_2$ e $W_1 \sim W_2$. Si ha dunque $V = Fu_1 \perp W_1 \perp U_1^\perp = Fu_2 \perp W_2 \perp U_2^\perp$. Applicando a Fu_1 e Fu_2 il risultato unidimensionale, concludiamo che esiste un'isometria η di $W_1 \perp U_1^\perp$ su $W_2 \perp U_2^\perp$. Si ha pertanto $W_2 \perp U_2^\perp = \eta(W_1) \perp \eta(U_1^\perp)$ e $W_2 \sim W_1 \sim \eta(W_1)$. L'ipotesi induttiva applicata ai sottospazi W_2 e $\eta(W_1)$ di $W_2 \perp U_2^\perp$ dà $U_2^\perp \sim \eta(U_1^\perp) \sim U_1^\perp$. □

Supponiamo, come nel teorema, che U_1, U_2 siano sottospazi non degeneri e sia η un'isometria di U_1 su U_2. Il teorema fornisce un'isometria ζ di $U_1^\perp$ su $U_2^\perp$. Poiché $V = U_1 \perp U_1^\perp = U_2 \perp U_2^\perp$, l'applicazione lineare di V in sé che coincide con η su U_1 e con ζ su $U_1^\perp$ è una trasformazione ortogonale che prolunga l'isometria assegnata.

Mostriamo che il risultato vale per sottospazi arbitrari: ogni isometria fra sottospazi di V si può prolungare a una trasformazione ortogonale di V. Baseremo la dimostrazione su un'immersione canonica di un sottospazio degenere in uno non degenere, che ricondurrà il problema al caso non degenere. Ci occorre il seguente teorema di immersione.

Teorema 6.11. *Sia V munito di una forma quadratica non degenere e sia U un sottospazio tale che* $\operatorname{rad} U = U \cap U^\perp \neq 0$. *Si scriva $U = \operatorname{rad} U \oplus U'$, dove U' è un sottospazio, e sia $(z_1, \dots, z_r)$ una base di* $\operatorname{rad} U$. *Possiamo allora immergere U in un sottospazio non degenere $U \oplus W$, dove W ha una base $(w_1, \dots, w_r)$ tale che (z_i, w_i) è una coppia iperbolica per $1 \leq i \leq r$ e $U + W = U' \perp H_1 \perp H_2 \perp \cdots \perp H_r$, con $H_i = Fz_i + Fw_i$ piano iperbolico.*

Dimostrazione. Sia f la funzione lineare su U tale che $f(z_1) = 1$, $f(z_i) = 0$ per $i > 1$ e $f(u') = 0$ per $u' \in U'$. Per il lemma di p. 347 esiste $w_1 \in V$ tale che $f(u) = B(u, w_1)$ per ogni $u \in U$. Dunque $B(z_1, w_1) = 1$, $B(z_i, w_1) = 0$ per $i > 1$ e $B(u', w_1) = 0$ per $u' \in U'$. Sostituendo w_1 con un opportuno $w_1 + az_1$, possiamo supporre $Q(w_1) = 0$. Allora (z_1, w_1) è una coppia iperbolica, in particolare linearmente indipendente. Si ha

$$V = (Fz_1 + Fw_1) \oplus (Fz_1 + Fw_1)^\perp, \qquad U_1 = U' + \sum_{j>1} Fz_j \subseteq V_1 = (Fz_1 + Fw_1)^\perp.$$

Il radicale di U_1 è $\sum_{j>1} Fz_j$. Se $r = 1$, prendiamo $W = Fw_1$ e otteniamo $U + W = U' \perp H_1$, dove H_1 è il piano iperbolico $Fz_1 + Fw_1$. Se $r > 1$, sostituiamo alla coppia V, U la coppia V_1, U_1 e osserviamo che la dimensione del radicale di U_1 è $r - 1$. Procedendo per induzione sulla

dimensione del radicale, otteniamo dunque vettori $w_2, \dots, w_r \in V_1$ tali che

$$U_1 + \sum_{j=2}^{r} Fw_j = U' \perp H_2 \perp \cdots \perp H_r, \qquad H_j = Fz_j + Fw_j$$

è un piano iperbolico. Allora $W = \sum_{i=1}^{r} Fw_i$ soddisfa le condizioni richieste. □

Sia ora η un'isometria di un sottospazio U_1 su un sottospazio U_2. Se U_1 è non degenere, anche U_2 lo è e abbiamo già visto che η si prolunga a una trasformazione ortogonale. Nel caso degenere scriviamo $U_1 = \operatorname{rad} U_1 \oplus U_1'$, dove U_1' è un sottospazio.[4] Per il Teorema 6.11 esiste un sottospazio non degenere $U_1 + W_1 = U_1' \perp H_1 \perp \cdots \perp H_r$, dove $H_i = Fz_i + Fw_i$ e (z_i, w_i) è una coppia iperbolica. Possiamo immergere $U_2 = \eta U_1$ in $U_2 + W_2 = \eta(U_1') \perp H_1' \perp \cdots \perp H_r'$, dove $H_i' = F(\eta z_i) + Fw_i'$ e $(\eta z_i, w_i')$ è una coppia iperbolica. L'applicazione lineare di $U_1 + W_1$ su $U_2 + W_2$ che coincide con η su U_1 e manda w_i in w_i', $1 \le i \le r$, è evidentemente un'isometria che prolunga η. Poiché $U_1 + W_1$ è non degenere, essa si prolunga a sua volta a una trasformazione ortogonale. Dunque η si prolunga a una trasformazione ortogonale, e abbiamo dimostrato il teorema seguente.

Teorema di prolungamento di Witt. *Se V è munito di una forma quadratica non degenere Q, ogni isometria di un sottospazio U_1 su un sottospazio U_2 si può prolungare a una trasformazione ortogonale.*

Questo vale in particolare per sottospazi totalmente isotropi ($Q|_{U_1} = 0$). Se U_1 e U_2 sono due tali sottospazi della stessa dimensione, il teorema di prolungamento assicura l'esistenza di una trasformazione ortogonale che manda U_1 su U_2. Se U è un sottospazio totalmente isotropo della massima dimensione possibile e U_1 è un qualsiasi sottospazio totalmente isotropo, esiste una trasformazione ortogonale η che manda U_1 in U. Allora $\eta^{-1}U$ è un sottospazio totalmente isotropo contenente U_1. Ne segue che tutti i sottospazi totalmente isotropi massimali hanno la stessa dimensione. Tale dimensione comune si chiama *indice di Witt* di Q e si indica con $\nu(Q)$.

Sia ora U un sottospazio totalmente isotropo di dimensione $\nu = \nu(Q)$. Per il Teorema 6.11 possiamo immergerlo in un sottospazio $U + W$ che è somma diretta ortogonale di ν piani iperbolici. Dunque $2\nu \le n = \dim V$, e quindi $\nu(Q) \le \lfloor n/2 \rfloor$. Possiamo anche scrivere $V = (U+W) \oplus (U+W)^\perp$, ed è chiaro che il sottospazio $X = (U+W)^\perp$ è anisotropo, ossia non contiene vettori isotropi. Abbiamo la decomposizione

$$V = H_1 \perp H_2 \perp \cdots \perp H_\nu \perp X, \tag{32}$$

dove H_i è un piano iperbolico, $1 \le i \le \nu$, e X è anisotropo. Supponiamo poi che $V = H_1' \perp \cdots \perp H_r' \perp Y$ sia una decomposizione in somma diretta ortogonale di piani iperbolici H_i' e di un sottospazio anisotropo Y. Scegliendo $z_i' \neq 0$ in H_i' con $Q(z_i') = 0$, si ottiene un sottospazio totalmente isotropo $\sum_{i=1}^{r} Fz_i'$ di dimensione r. Dunque $r \le \nu$. Inoltre esiste una trasformazione ortogonale che manda $H_1 \perp \cdots \perp H_r$ su $H_1' \perp \cdots \perp H_r'$. Essa manda $(\sum_{i=1}^{r} H_i)^\perp = H_{r+1} \perp \cdots \perp H_\nu \perp X$ su $(\sum_{i=1}^{r} H_i')^\perp = Y$. Essendo Y anisotropo, deve aversi $r = \nu$ e $X \sim Y$. Si chiama *nucleo anisotropo* di V ogni sottospazio anisotropo Y tale che V sia somma diretta ortogonale di Y e di piani iperbolici. Abbiamo visto che due nuclei anisotropi qualsiasi sono isometrici. Ne segue che due forme quadratiche non degeneri della stessa dimensione sono equivalenti se e solo se sono equivalenti i loro nuclei anisotropi, nel senso evidente. Il problema della classificazione delle forme quadratiche a meno di equivalenza si riduce così al caso anisotropo.

Esercizi

[4] In questo punto il testo a stampa, p. 369, presenta una frase interrotta e un frammento ripetuto dal capoverso successivo. Il passaggio è ricostruito in base al Teorema 6.11.

1. Si dica che Q ha *indice di Witt massimo* se $\nu(Q) = \lfloor n/2 \rfloor$. Mostrare che due forme quadratiche non degeneri di indice di Witt massimo sono sempre equivalenti se n è pari, e che, se n è dispari, sono equivalenti se e solo se le forme bilineari associate hanno lo stesso discriminante.

2. Mostrare che, se Q è una forma quadratica non degenere su uno spazio vettoriale V su $\mathbb{R}$, allora $\nu(Q) = \frac{1}{2}(n - |\operatorname{sig} Q|)$, dove $\operatorname{sig} Q$ è la segnatura di Q.

3. (Cayley.) Sia η un endomorfismo lineare di V, munito di una forma quadratica non degenere Q, e si indichi con η' l'aggiunta rispetto alla forma bilineare simmetrica associata B. Sia η ortogonale (quindi $\eta' = \eta^{-1}$) e si supponga $\det(\eta + 1) \neq 0$. Si definisca la *trasformata di Cayley* di η ponendo

$$\sigma = (1 - \eta)(1 + \eta)^{-1} = (1 + \eta)^{-1}(1 - \eta).$$

   Mostrare che σ è antisimmetrica rispetto a B, nel senso che $\sigma' = -\sigma$, e che $\det(\sigma + 1) \neq 0$. Mostrare che $\eta = (1 - \sigma)(1 + \sigma)^{-1} = (1 + \sigma)^{-1}(1 - \sigma)$.

4. Usare l'esercizio 3 per dimostrare che, se V ha dimensione dispari, ogni trasformazione ortogonale propria ha un punto fisso non nullo ($\eta x = x$), e che, se V ha dimensione pari, ogni trasformazione ortogonale impropria ha un punto fisso non nullo.

5. Sia Q una forma quadratica di indice di Witt massimo ν su uno spazio V di dimensione $n = 2\nu$. Sia U un sottospazio totalmente isotropo di V con base $(u_1, \dots, u_\nu)$. Per il Teorema 6.11 esiste una base $(u_1, \dots, u_\nu, w_1, \dots, w_\nu)$ di V rispetto alla quale la matrice di B è

$$\begin{pmatrix} 0 & 1_\nu \\ 1_\nu & 0 \end{pmatrix}, \tag{33}$$

   dove 1_ν è la matrice identità di ordine ν. Mostrare che $(u_1, \dots, u_\nu, w'_1, \dots, w'_\nu)$ è una base con la stessa matrice (33) se e solo se $w'_i = w_i + v_i$, con $v_i \in U$ e $B(u_i, v_j) = -B(u_j, v_i)$ per $1 \leq i, j \leq \nu$. Osservare che quest'ultima condizione equivale a $v_i = \sum_j s_{ij} u_j$, dove $S = (s_{ij})$ è antisimmetrica.

6. Siano Q, V, U come nell'esercizio 5. Sia G_U il sottogruppo di $\mathrm{O}(V, Q)$ degli η che fissano ogni $u \in U$. Mostrare che G_U è isomorfo al gruppo additivo delle matrici antisimmetriche di ordine ν a elementi in F e che è quindi abeliano. Mostrare che $G_U \subseteq \mathrm{O}^+(V, Q)$.[5]

7. Siano $Q, V, U, (u_1, \dots, u_\nu), (w_1, \dots, w_\nu)$ come nell'esercizio 5 e si ponga $W = \sum_i F w_i$. Sia $H_{U,W}$ il sottogruppo delle trasformazioni ortogonali tali che $\eta(U) = U$ e $\eta(W) = W$. Mostrare che $H_{U,W} \cong \mathrm{GL}_\nu(F)$, il gruppo delle matrici invertibili di ordine ν a elementi in F.

8. Con le notazioni dell'esercizio 7, sia $\eta \in G_U$, come definito nell'esercizio 6. Determinare il sottospazio V_1 dei vettori fissati da η. Mostrare che esiste η tale che $V_1 = U$ se e solo se ν è pari, nel qual caso n è divisibile per quattro.

[5] Negli esercizi 5 e 6 l'originale reca n al posto di ν nell'intervallo degli indici e nell'ordine delle matrici antisimmetriche.

6.6 Il teorema di Cartan–Dieudonné

Per le forme quadratiche sui reali o sui complessi, É. Cartan ha dimostrato che ogni trasformazione ortogonale è prodotto di al più n simmetrie, dove n è la dimensione dello spazio vettoriale. Dieudonné ha generalizzato il risultato a forme quadratiche su campi arbitrari. Dimostriamo anzitutto una versione più debole del teorema.

Teorema 6.12. *Ogni trasformazione ortogonale è prodotto di simmetrie.*

Dimostrazione. Sia η ortogonale e sia u un vettore con $Q(u) \neq 0$. Come nella dimostrazione del teorema di cancellazione di Witt, esiste una simmetria S_w, $w = u + \varepsilon\eta u$, tale che $\eta' u = -\varepsilon u$, dove $\eta' = S_w\eta$ ed ε vale 1 o -1. Allora η' lascia stabile $(Fu)^\perp$, sottospazio non degenere di dimensione $n-1$. Per induzione sulla dimensione, la restrizione $\eta'|_{(Fu)^\perp}$ è uguale a $\overline{S}_{w_1}\overline{S}_{w_2}\cdots\overline{S}_{w_k}$, dove $\overline{S}_{w_i}$ è la simmetria di $(Fu)^\perp$ determinata da $w_i \in (Fu)^\perp$. Si ha $\overline{S}_{w_i} = S_{w_i}|_{(Fu)^\perp}$ e, poiché $u \perp w_i$, S_{w_i} fissa u. Allora $\eta'' = S_{w_k}S_{w_{k-1}}\cdots S_{w_1}\eta'$ è l'identità su $(Fu)^\perp$, perché $(S_{w_1}\cdots S_{w_k})^{-1} = S_{w_k}\cdots S_{w_1}$, essendo $S_w^2 = 1$. Inoltre $\eta'' u = \eta' u = \pm u$. Se $\eta'' u = u$, si ha $\eta'' = 1$; se $\eta'' u = -u$, si ha $\eta'' = S_u$. In entrambi i casi η' è prodotto di simmetrie, e lo stesso vale per $\eta = S_w\eta'$. □

Prima di dimostrare il più preciso teorema di Cartan–Dieudonné, osserviamo un fatto già enunciato in un esercizio (esercizio 1, p. 366). Se η è ortogonale, il sottospazio V_1 dei suoi punti fissi è il complemento ortogonale dell'immagine di $1-\eta$. In altri termini $\eta x = x$ se e solo se $x \perp (1-\eta)y$ per ogni y. Ora $\eta x = x$ se e solo se $\eta^{-1}x = x$, ossia se e solo se $\eta' x = x$. Poiché B è non degenere, questo equivale a $B((1-\eta')x, y) = 0$ per ogni y. Il risultato segue dall'identità $B((1-\eta')x, y) = B(x, (1-\eta)y)$.

Abbiamo chiamato unipotente una trasformazione lineare T tale che $T-1 = Z$ sia nilpotente. Questa condizione implica $\det T = 1$, dove $\det T$ indica il determinante di una matrice qualsiasi di T. Lo si vede usando la matrice canonica di Jordan di T come a p. 199, ma si può anche dare il seguente argomento più elementare. Si osservi anzitutto che T ha un punto fisso non nullo u. (Preso $v \neq 0$, basta scegliere per u l'ultimo vettore non nullo nella successione $v, Zv, Z^2v, \ldots$.) Allora Fu è stabile per T, che induce una trasformazione lineare $\overline{T}$ su $\overline{V} = V/Fu$. Essa è unipotente, quindi, per induzione sulla dimensione, $\det\overline{T} = 1$. Calcolando la matrice di T rispetto a una base $(u_1, u_2, \ldots, u_n)$ con $u_1 = u$ e con $(u_2 + Fu, \ldots, u_n + Fu)$ base di $\overline{V}$, si vede che $\det T = 1$. In particolare, ogni trasformazione ortogonale unipotente è una rotazione.

Teorema di Cartan–Dieudonné. *Se $\dim V = n$, ogni trasformazione ortogonale η di V è prodotto di al più n simmetrie.*

Dimostrazione. La dimostrazione che daremo è dovuta ad Artin. Osserviamo anzitutto che il risultato vale se il sottospazio V_1 dei punti fissi di η non è totalmente isotropo. In tal caso V_1 contiene un vettore u con $Q(u) \neq 0$ e η lascia stabile $(Fu)^\perp$. Procedendo per induzione su n, possiamo supporre che $\eta|_{(Fu)^\perp}$ sia prodotto di al più $n-1$ simmetrie definite da elementi di $(Fu)^\perp$. Ne segue che anche η è prodotto di al più $n-1$ simmetrie. Osserviamo poi che il risultato vale se esiste u con $Q(u) \neq 0$ e $Q(u - \eta u) \neq 0$. Infatti, come nella dimostrazione del Teorema 6.12 o del teorema di cancellazione di Witt, esiste S_w tale che $\eta' = S_w\eta$ fissi u. Allora η' è prodotto di al più $n-1$ simmetrie e $\eta = S_w\eta'$ è prodotto di al più n simmetrie.

Consideriamo il caso bidimensionale. Se Q è anisotropa, il risultato segue da quanto appena dimostrato. Possiamo quindi supporre che V sia un piano iperbolico. Allora V ha una base iperbolica (u, v) e η manda $u \mapsto au$, $v \mapsto a^{-1}v$ per qualche $a \in F^*$, oppure $u \mapsto av$, $v \mapsto a^{-1}u$ (Teorema 6.10, p. 365). Nel primo caso possiamo supporre $a \neq 1$, altrimenti $\eta = 1$ e il risultato è banale. Per $w = u + v$ si ha $w - \eta w = (1-a)u + (1-a^{-1})v$, con $Q(w) \neq 0$ e $Q(w - \eta w) \neq 0$, sicché il risultato segue come prima. Se invece $\eta u = av$ e $\eta v = a^{-1}u$, il vettore $w = u + av$ è fissato da η e $Q(w) \neq 0$. Si applica quindi la nostra prima osservazione.

Rimane un solo caso possibile: $\dim V \geq 3$, il sottospazio V_1 dei punti fissi è totalmente isotropo e $Q(u-\eta u) = 0$ per ogni u tale che $Q(u) \neq 0$. Dimostriamo che da $\dim V \geq 3$ e da quest'ultima condizione segue $Q(u-\eta u) = 0$ per ogni u. È sufficiente verificarlo per $w \neq 0$ con $Q(w) = 0$. Lo spazio $(Fw)^\perp$ ha dimensione $n-1$, e da $n \geq 3$ segue $n-1 > \lfloor n/2 \rfloor$, cosicché $(Fw)^\perp$ non è totalmente isotropo. Esiste allora $u \neq 0$ tale che $u \perp w$ e $Q(u) \neq 0$. Si ha anche $(w \pm u) \perp w$ e $Q(w \pm u) = Q(u) \neq 0$. Ponendo $\zeta = 1 - \eta$, abbiamo dunque le tre equazioni

$$Q(\zeta u) = 0, \qquad Q(\zeta w + \zeta u) = 0, \qquad Q(\zeta w - \zeta u) = 0.$$

Esse implicano $Q(\zeta w) = Q(w - \eta w) = 0$. Questo vale pertanto per ogni vettore di V, e quindi $(1-\eta)V$ è totalmente isotropo. Inoltre, come abbiamo visto, $V_1 = ((1-\eta)V)^\perp$ e perciò $(1-\eta)V = V_1^\perp$. Poiché V_1 e $(1-\eta)V$ sono totalmente isotropi, si ha $V_1 \subseteq V_1^\perp = (1-\eta)V$ e $(1-\eta)V \subseteq ((1-\eta)V)^\perp = V_1$. Dunque $V_1 = (1-\eta)V$, per cui $(1-\eta)^2 x = 0$ per ogni x. Ne segue che η è unipotente e quindi è una rotazione. Inoltre $n = \dim V_1 + \dim V_1^\perp$ e $V_1^\perp = (1-\eta)V = V_1$, sicché $n = 2 \dim V_1$ e V ha dimensione pari.

Per completare la dimostrazione nel caso appena considerato, basta porre $\eta' = S_w \eta$, con S_w una simmetria qualsiasi. Poiché η' è impropria, essa rientra nei casi precedenti ed è prodotto di $k \leq n$ simmetrie. Ogni simmetria è impropria, quindi k è dispari; e poiché n è pari, $k \leq n-1$. Ne segue che $\eta = S_w \eta'$ è prodotto di al più n simmetrie. □

Dal teorema di Cartan–Dieudonné si ricava subito che ogni rotazione di uno spazio di dimensione dispari e ogni trasformazione ortogonale impropria di uno spazio di dimensione pari hanno un punto fisso non nullo. Abbiamo già indicato questi risultati come conseguenze della parametrizzazione di Cayley delle trasformazioni ortogonali mediante trasformazioni lineari antisimmetriche (esercizi 3 e 4, p. 370). Per dedurli dal teorema di Cartan–Dieudonné, osserviamo che la nota formula di algebra lineare $\dim(U_1 \cap U_2) = \dim U_1 + \dim U_2 - \dim(U_1 + U_2)$ per i sottospazi di V permette di dimostrare, per induzione su k, che l'intersezione di k iperpiani, cioè di k sottospazi di dimensione $n-1$, ha dimensione almeno $n-k$. Sia $\eta = S_{u_1} S_{u_2} \cdots S_{u_k}$, con S_{u_i} simmetria determinata dal vettore u_i. Poiché S_{u_i} ha un iperpiano di punti fissi e ogni punto dell'intersezione di questi iperpiani è fisso per η, vediamo che η ha un punto fisso non nullo se è prodotto di $k \leq n-1$ simmetrie. Ogni simmetria è ortogonale impropria, perciò un prodotto di k simmetrie è proprio o improprio secondo che k sia pari o dispari. Dal teorema di Cartan–Dieudonné segue dunque che ogni rotazione è prodotto di un numero pari, al più n, di simmetrie; e ogni trasformazione ortogonale impropria è prodotto di un numero dispari, al più n, di simmetrie. Una rotazione in dimensione dispari o una trasformazione ortogonale impropria in dimensione pari è pertanto prodotto di $k \leq n-1$ simmetrie e ha un punto fisso non nullo.

Sia V_1 l'insieme dei punti fissi della trasformazione ortogonale η. L'argomento appena usato mostra che, se η è prodotto di k simmetrie, allora $\dim V_1 \geq n-k$. Poiché $\dim V_1 = n-r$, dove r è il rango di $1-\eta$, ossia $\dim(1-\eta)V$, η non può essere prodotto di meno di r simmetrie. Si può scrivere η come prodotto di r simmetrie? Scherk ha dimostrato che, in generale, la risposta è affermativa. Precisamente, se r è il rango di $1-\eta$, η si scrive come prodotto di r simmetrie, salvo il caso in cui $1-\eta$ è antisimmetrica rispetto alla forma bilineare B, in cui il numero minimo necessario è $r+2$.[6]

Un'altra importante conseguenza del teorema di Cartan–Dieudonné, o anche del più debole Teorema 6.12, è la seguente.

Teorema 6.13. *Se* $\dim V \geq 3$*, il gruppo dei commutatori* $(\mathrm{O}(V,Q), \mathrm{O}(V,Q))$ *coincide con* $(\mathrm{O}^+(V,Q), \mathrm{O}^+(V,Q))$.

[6] P. Scherk, «On the decomposition of orthogonalities into symmetries», *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 481–491.

Dimostrazione. Osserviamo che $(\mathrm{O}(V,Q),\mathrm{O}(V,Q))$ è generato dai commutatori $S_u^{-1}S_v^{-1}S_uS_v = (S_uS_v)^2$ di simmetrie S_u, S_v. Infatti, poiché il coniugato $\eta S_u \eta^{-1}$ è $S_{\eta u}$, il sottogruppo $\mathrm{O}'(V,Q)$ generato da tutti gli $(S_uS_v)^2$ è normale in $\mathrm{O}(V,Q)$. Il gruppo quoziente è generato dalle classi laterali $S_u\mathrm{O}'(V,Q)$. Poiché questi generatori commutano, $\mathrm{O}(V,Q)/\mathrm{O}'(V,Q)$ è commutativo. Dunque $\mathrm{O}'(V,Q) \supseteq (\mathrm{O}(V,Q),\mathrm{O}(V,Q))$. L'inclusione inversa segue dalla definizione di sottogruppo dei commutatori. Quindi $\mathrm{O}'(V,Q) = (\mathrm{O}(V,Q),\mathrm{O}(V,Q))$. Per concludere basterà mostrare che ogni $(S_uS_v)^2$ è prodotto di commutatori di rotazioni.

Se $n = \dim V$ è dispari, la trasformazione lineare -1 è ortogonale impropria e appartiene al centro di $\mathrm{O}(V,Q)$. Allora $-S_u = (-1)S_u$ è una rotazione, e il commutatore di S_u e S_v coincide con il commutatore delle due rotazioni $-S_u$ e $-S_v$. Possiamo quindi supporre n pari, e dunque $n \geq 4$. In questo caso affermiamo che $U^\perp$, dove $U = Fu + Fv$, contiene un vettore w con $Q(w) \neq 0$. Altrimenti $U^\perp$ sarebbe totalmente isotropo e $U^\perp \subseteq U^{\perp\perp} = U$. Poiché $\dim U^\perp = n - \dim U \geq n-2 \geq 2$, si avrebbe $U = U^\perp$ totalmente isotropo, contro il fatto che U contiene u e v, con $Q(u) \neq 0$ e $Q(v) \neq 0$. Scegliamo dunque $w \in U^\perp$ con $Q(w) \neq 0$. Allora $S_uS_wS_u^{-1} = S_{S_uw} = S_w$, per cui $S_uS_w = S_wS_u$. Analogamente S_v e S_w commutano. Il commutatore di S_u e S_v coincide pertanto con quello delle due rotazioni S_uS_w e S_vS_w. □

6.7 Struttura del gruppo lineare generale $\mathrm{GL}_n(F)$

Nel resto del capitolo studieremo la struttura dei gruppi geometrici «classici». Con questo nome intendiamo il gruppo lineare generale $\mathrm{GL}_n(F)$, definito come gruppo degli automorfismi lineari di uno spazio vettoriale V di dimensione n su un campo F; i gruppi ortogonali di V definiti da forme quadratiche non degeneri; e il gruppo simplettico, definito come gruppo delle isometrie di uno spazio vettoriale munito della metrica data da una forma bilineare alternante non degenere. Camille Jordan, nel *Traité des Substitutions* (1870), studiò per primo i gruppi $\mathrm{GL}_n(F)$ e i gruppi simplettici su F per $F = \mathbb{Z}/(p)$, con p primo. Dickson trattò la generalizzazione di questi risultati a un campo di base finito arbitrario e studiò i gruppi ortogonali e unitari su campi finiti nel suo libro *Linear Groups with an Exposition of Galois Field Theory*, apparso nel 1900. Poco più tardi, nel 1901, Dickson avviò lo studio dei gruppi classici su un campo arbitrario F e determinò la struttura di $\mathrm{GL}_n(F)$, del gruppo simplettico $\mathrm{Sp}_n(F)$ e di certi gruppi ortogonali.[7] Iwasawa diede nel 1941 una dimostrazione semplice dei risultati su $\mathrm{GL}_n(F)$.[8] In quel lavoro Iwasawa delineò anche una dimostrazione della semplicità dei gruppi simplettici proiettivi su campi arbitrari. Nel 1948 Dieudonné, nella monografia *Sur les Groupes Classiques*, dimostrò un risultato sorprendente: per un campo di base arbitrario la struttura dei gruppi ortogonali è profondamente diversa nei due casi di indice di Witt positivo e di indice di Witt nullo, ossia per le forme anisotrope. Nel primo caso si ha per $n \geq 5$ un teorema generale: il quoziente del gruppo dei commutatori rispetto al suo centro è semplice. Questo vale per qualsiasi campo di base. Per le forme anisotrope, invece, la struttura dipende dal campo di base ed esistono casi in cui il gruppo dei commutatori modulo il suo centro non è semplice. Su un campo finito l'indice di Witt è sempre positivo per $n \geq 3$. I gruppi classici su campi finiti fornirono i primi esempi, oltre ai gruppi alterni, di gruppi finiti semplici non abeliani.

Cominciamo da $\mathrm{GL}_n(F)$, il gruppo degli automorfismi lineari di uno spazio vettoriale V di dimensione n su F. Usando la corrispondenza fra trasformazioni lineari e matrici rispetto a una base di V, possiamo identificare $\mathrm{GL}_n(F)$ con il gruppo delle matrici invertibili $n \times n$ a elementi in F. Per iniziare adotteremo il punto di vista matriciale. Abbiamo l'omomorfismo determinante

[7] L. E. Dickson, «The theory of linear groups in an arbitrary field», *Transactions of the American Mathematical Society*, vol. 2 (1901), pp. 363–394.

[8] K. Iwasawa, «Über die Einfachheit der speziellen projektiven Gruppen», *Proceedings of the Imperial Academy of Tokyo*, vol. 17 (1941), pp. 57–59.

$a \mapsto \det a$ di $\mathrm{GL}_n(F)$ nel gruppo moltiplicativo F^* degli elementi non nulli di F. Il suo nucleo è il *gruppo unimodulare*, o *gruppo lineare speciale*, $\mathrm{SL}_n(F)$ delle matrici di determinante 1. Il nostro risultato principale sarà che, eccettuati i casi $n = 2$ con F campo di due o tre elementi, $\mathrm{SL}_n(F)$ modulo il suo centro è semplice. Determiniamo anzitutto un sistema di generatori di $\mathrm{SL}_n(F)$.

Lemma 1. *Il gruppo* $\mathrm{SL}_n(F)$ *è generato dalle matrici elementari* $T_{ij}(b) = 1 + be_{ij}$, $i \neq j$, $b \in F$.

(Come di consueto, e_{ij} è la matrice il cui unico elemento non nullo è un 1 nella posizione (i, j). Si ha $e_{ij}e_{kl} = \delta_{jk}e_{il}$.)

Dimostrazione. Dimostreremo un risultato più generale, sostituendo al campo F un dominio euclideo D. Per ogni b e $i \neq j$, la matrice $T_{ij}(b)$ ha evidentemente determinante 1. Mostreremo che ogni $A \in M_n(D)$ con $\det A = 1$ è prodotto di matrici $T_{ij}(b)$. Dalla dimostrazione del Teorema 3.8 (p. 182) risulta che esistono matrici P, Q, prodotti di matrici della forma $T_{ij}(b)$ e di matrici $P_{ij} = 1 + e_{ij} + e_{ji} - e_{ii} - e_{jj}$, tali che $PAQ = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$. Poiché

$$P_{ij} = (1 + e_{ij})(1 - e_{ji})(1 + e_{ij})(1 - 2e_{ii}),$$

possiamo sostituire ai P_{ij} matrici $T_{ij}(b)$ e matrici $1 - 2e_{ii}$. Inoltre

$$T_{ij}(b)(1 - 2e_{ii}) = (1 - 2e_{ii})T_{ij}(-b), \qquad T_{ji}(b)(1 - 2e_{ii}) = (1 - 2e_{ii})T_{ji}(-b),$$

e $1 - 2e_{ii}$ commuta con ogni $T_{jk}(b)$ per $j, k \neq i$. Possiamo quindi raccogliere i fattori $1 - 2e_{ii}$ all'estrema sinistra di P e all'estrema destra di Q. Moltiplicando per i loro inversi si modifica la matrice diagonale ottenendo ancora una matrice diagonale. Possiamo dunque supporre che P e Q siano prodotti di matrici della forma $T_{ij}(b)$. Il problema si riduce così ad $A = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$. Dall'ipotesi $\det A = 1$ segue $d_1 d_2 \cdots d_n = 1$, perciò ogni d_i è invertibile. Mediante una successione di trasformazioni elementari, che il lettore riconoscerà facilmente, si passa nell'ordine attraverso le matrici

$$\begin{pmatrix} d^{-1} & 0 \\ 0 & d \end{pmatrix} \longrightarrow \begin{pmatrix} d^{-1} & 0 \\ d^{-1} & d \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & -d \\ d^{-1} & d \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & -d \\ d^{-1} & 0 \end{pmatrix}$$
$$\longrightarrow \begin{pmatrix} 1 & -d \\ d^{-1} & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ d^{-1} & 1 \end{pmatrix}.$$

Quindi $\mathrm{diag}\{d^{-1}, d\}$ è prodotto di matrici elementari $T_{ij}(b)$ con $i, j = 1, 2$, $i \neq j$. Lo stesso vale evidentemente per $\mathrm{diag}\{d_1^{-1}, d_1, 1, \ldots, 1\}$. Moltiplicando $A = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$ a destra per questa matrice, otteniamo $\mathrm{diag}\{1, d_1 d_2, d_3, \ldots, d_n\}$. Ripetendo il procedimento si giunge a $\mathrm{diag}\{1, 1, \ldots, 1, d_1 \cdots d_n\} = 1$, il che conclude la dimostrazione. □

Lemma 2. *Per* $n \geq 2$, *salvo i casi* $n = 2$ *e* F *campo di due o tre elementi,* $\mathrm{SL}_n(F)$ *coincide con il proprio gruppo dei commutatori.*

Dimostrazione. Basta mostrare che i generatori $T_{ij}(b)$ appartengono al gruppo dei commutatori. Se $n \geq 3$, scegliamo $k \neq i, j$. Il risultato segue dal calcolo

$$\begin{aligned} T_{ij}(b) &= T_{ik}(b)T_{kj}(1)T_{ik}(-b)T_{kj}(-1) \\ &= T_{ik}(b)T_{kj}(1)T_{ik}(b)^{-1}T_{kj}(1)^{-1}. \end{aligned} \tag{34}$$

Se $n = 2$, si ha

$$\begin{pmatrix} d & 0 \\ 0 & d^{-1} \end{pmatrix} \begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d^{-1} & 0 \\ 0 & d \end{pmatrix} \begin{pmatrix} 1 & -c \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & c(d^2 - 1) \\ 0 & 1 \end{pmatrix}. \tag{35}$$

Se F ha più di tre elementi possiamo scegliere $d \neq 0$ con $d^2 \neq 1$ e quindi, per ogni b assegnato, $c = (d^2-1)^{-1}b$. La (35) mostra allora che $T_{12}(b)$ appartiene al gruppo dei commutatori di $\mathrm{SL}_2(F)$. Analogamente vi appartiene $T_{21}(b)$. □

Poiché $\mathrm{GL}_n(F)/\mathrm{SL}_n(F) \cong F^*$ è abeliano, $\mathrm{SL}_n(F)$ contiene il gruppo dei commutatori $\mathrm{GL}_n(F)'$ di $\mathrm{GL}_n(F)$. D'altra parte, nelle ipotesi del Lemma 2, $\mathrm{SL}_n(F) = \mathrm{SL}_n(F)' \subseteq \mathrm{GL}_n(F)'$. Pertanto $\mathrm{SL}_n(F) = \mathrm{GL}_n(F)'$. Poiché ogni $T_{ij}(1) = 1 + e_{ij}$, $i \neq j$, appartiene a $\mathrm{SL}_n(F)$, una matrice che commuta con tutti gli elementi di $\mathrm{SL}_n(F)$ commuta con ogni matrice. Essa ha dunque la forma $d1$. Ne segue che il centro C di $\mathrm{SL}_n(F)$ è $F^*1 \cap \mathrm{SL}_n(F)$, insieme finito delle matrici $d1$ con $d^n = 1$.

Indichiamo il gruppo quoziente $\mathrm{SL}_n(F)/C$ con $\mathrm{PSL}_n(F)$ e lo chiamiamo *gruppo unimodulare proiettivo*. Mostreremo che è semplice per $n \geq 2$, eccettuati i due casi $n = 2$, $|F| = 2$ o 3. La dimostrazione, dovuta a Iwasawa, si basa su un'azione naturale di $\mathrm{GL}_n(F)$ su un insieme $P_{n-1}(F)$ detto *spazio proiettivo di dimensione* $n-1$ *su* F. Esso è semplicemente l'insieme dei sottospazi unidimensionali Fx, $x \neq 0$, dello spazio vettoriale V di dimensione n su F. Se T è un automorfismo lineare di V, definiamo la sua azione su Fx mediante $T(Fx) = F(Tx)$. Otteniamo così un'azione di $\mathrm{GL}_n(F)$ su $P_{n-1}(F)$. Il suo nucleo è costituito dai T tali che $F(Tx) = Fx$ per ogni $x \neq 0$ di V. Affermiamo che esso è l'insieme dei $T = a1$, $a \neq 0$. La condizione perché T appartenga al nucleo equivale a richiedere che per ogni $x \neq 0$ esista $a_x \in F^*$ tale che $Tx = a_x x$. Per $b \neq 0$ di F si ha $T(bx) = a_{bx}(bx)$ e $T(bx) = bTx = ba_x x = a_x(bx)$, quindi $a_{bx} = a_x$. Allora $T = a1$, $a = a_x$, se $\dim V = 1$. Supponiamo ora $\dim V \geq 2$ e sia $(e_1, \ldots, e_n)$ una base di V su F. Si ha $Te_i = a_{e_i}e_i$ e, per $i \neq j$, $T(e_i + e_j) = a_{e_i+e_j}(e_i + e_j)$. Poiché $T(e_i + e_j) = Te_i + Te_j = a_{e_i}e_i + a_{e_j}e_j$, si ha $a_{e_i+e_j} = a_{e_i} = a_{e_j}$. Di nuovo $T = a1$, $a \neq 0$. Viceversa, ogni applicazione di questa forma agisce come l'identità su $P_{n-1}(F)$. Ponendo $\mathrm{PGL}_n(F) = \mathrm{GL}_n(F)/F^*1$, otteniamo perciò un'azione fedele su $P_{n-1}(F)$, in cui la classe laterale $[T] = F^*T$ agisce su Fx mediante $[T](Fx) = F(Tx)$. Il gruppo $\mathrm{PGL}_n(F)$ si chiama *gruppo proiettivo*. Esso contiene il sottogruppo $F^*\mathrm{SL}_n(F)/F^*1 \cong \mathrm{SL}_n(F)/(F^*1 \cap \mathrm{SL}_n(F))$, che abbiamo chiamato gruppo unimodulare proiettivo. Anch'esso agisce fedelmente su $P_{n-1}(F)$.

Richiamiamo alcuni concetti e risultati sulle azioni di gruppi, introdotti nel § 1.12, che serviranno in questa dimostrazione di semplicità e nelle successive. L'azione di un gruppo G su un insieme S è *transitiva* se, dati $x_1, x_2 \in S$, esiste $g \in G$ tale che $gx_1 = x_2$. Per un intero positivo k l'azione si dice k *volte transitiva* se, date due k-uple ordinate di elementi distinti $(x_1, \ldots, x_k)$ e $(y_1, \ldots, y_k)$ di S, esiste $g \in G$ tale che $gx_i = y_i$ per $1 \leq i \leq k$. La transitività semplice, cioè per $k = 1$, coincide naturalmente con la transitività. L'azione di G è *primitiva* se le sole partizioni di S stabili per l'azione indotta di G sull'insieme delle parti $\mathcal{P}(S)$ sono le due banali: (1) la partizione costituita dal solo S, (2) la partizione in singoletti $\{x\}$, $x \in S$. L'imprimitività equivale all'esistenza di un sottoinsieme proprio A di S, contenente almeno due elementi, tale che per ogni $g \in G$ o $gA = A$ oppure $gA \cap A = \varnothing$. Ricordiamo anche il criterio seguente: se G agisce transitivamente su S, l'azione è primitiva se e solo se $\mathrm{Stab}\, x$, per un qualsiasi $x \in S$, è un sottogruppo massimale di G (Teorema 1.12, p. 77).

Lemma 3. *(1) Un'azione doppiamente transitiva di un gruppo G su un insieme S è primitiva. (2) Se G agisce primitivamente su S e $H \trianglelefteq G$ non è contenuto nel nucleo, allora H agisce transitivamente su S. (3) Se un sottogruppo H agisce transitivamente su S, allora $G = H\,\mathrm{Stab}\, x$ per ogni $x \in S$, dove $\mathrm{Stab}\, x$ è lo stabilizzatore di x in G.*

Dimostrazione. (1) Sia A un sottoinsieme proprio di S contenente due elementi distinti x, y. Per la doppia transitività esiste $g \in G$ con $gx = x$ e $gy \notin A$. Allora $gA \neq A$, ma $gA \cap A$ contiene x e quindi non è vuoto. L'azione è perciò primitiva. (2) Consideriamo la partizione di S nelle orbite di H. Essendo $H \trianglelefteq G$, per ogni $g \in G$ e $x \in S$ si ha $g(Hx) = H(gx)$. Dunque G lascia stabile la partizione di S nelle orbite di H. Per la primitività di G e poiché H non è contenuto

nel nucleo dell'azione, si ha una sola orbita di H. Pertanto H agisce transitivamente su S. (3) Siano H un sottogruppo transitivo di G, $x \in S$, $g \in G$. Esiste allora $h \in H$ tale che $hx = gx$. Ne segue $h^{-1}g \in \operatorname{Stab} x$ e $g \in H \operatorname{Stab} x$. □

Il criterio fondamentale di semplicità che useremo è il seguente.

Lemma 4. *Sia data un'azione di G su un insieme S e sia K il suo nucleo. Allora G/K è semplice se G soddisfa le tre condizioni seguenti.*

1. G agisce primitivamente su S.
2. $G = G'$, dove G' è il gruppo dei commutatori di G.
3. Esiste $x \in S$ tale che $\operatorname{Stab} x$ contenga un sottogruppo normale abeliano A_x i cui coniugati gA_xg^{-1}, $g \in G$, generano G.

Dimostrazione. Sia $H \trianglelefteq G$, con $H \not\subseteq K$. Per il Lemma 3(2), H è transitivo su S. Sia x come nella condizione 3. Per il Lemma 3(3), $G = H \operatorname{Stab} x$. Consideriamo $G^* = HA_x$. È normale in G, dunque contiene tutti i gA_xg^{-1}. Per la condizione 3 segue $G^* = G$. Quindi $G = HA_x$ e $G/H \cong A_x/(H \cap A_x)$ è abeliano. Perciò H contiene $G' = G$ e $H = G$. Ne segue la semplicità di G/K. □

Per applicare questo criterio a $\mathrm{PSL}_n(F)$ occorrono il Lemma 2 e altri due risultati che ora dimostreremo.

Lemma 5. *Se $n \geq 2$, $\mathrm{SL}_n(F)$ è doppiamente transitivo sullo spazio proiettivo $P_{n-1}(F)$.*

Dimostrazione. Dati $Fx_1 \neq Fx_2$ e $Fy_1 \neq Fy_2$, con $x_i, y_i \neq 0$, dobbiamo trovare una trasformazione lineare T di determinante 1 tale che $Tx_1 = a_1y_1 \neq 0$ e $Tx_2 = a_2y_2 \neq 0$, $a_i \in F$. Per ipotesi sia x_1, x_2 sia y_1, y_2 sono coppie linearmente indipendenti. Possiamo scegliere una base $(x_1, x_2, \ldots, x_n)$ e scrivere $y_1 = \sum_j a_{1j}x_j$, $y_2 = \sum_j a_{2j}x_j$. Se $n > 2$, alla matrice

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}$$

possiamo aggiungere $n-2$ righe in modo da ottenere una matrice (a_{ij}) di determinante 1. Sia $y_i = \sum_j a_{ij}x_j$ per $1 \leq i \leq n$ e sia T la trasformazione lineare $x_i \mapsto y_i$. Essa soddisfa le condizioni richieste. Se $n = 2$, si ha $\det(a_{ij}) = a \neq 0$ per $i, j = 1, 2$, e possiamo prendere la trasformazione lineare $x_1 \mapsto y_1$, $x_2 \mapsto a^{-1}y_2$. □

Sia $(e_1, \ldots, e_n)$ una base di V e consideriamo in $\mathrm{SL}_n(F)$ lo stabilizzatore del punto proiettivo Fe_1, che indicheremo con $\operatorname{Stab} e_1$. Esso è l'insieme delle trasformazioni lineari le cui matrici sono della forma

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & & & \\ \vdots & & A_{n-1} & \\ a_{n1} & & & \end{pmatrix}, \qquad a_{11} \det A_{n-1} = 1.$$

L'applicazione che associa a una tale trasformazione la matrice A_{n-1} è un omomorfismo. Il suo nucleo A_{e_1} è l'insieme delle trasformazioni le cui matrici sono

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21} & & & \\ \vdots & & 1_{n-1} & \\ a_{n1} & & & \end{pmatrix}.$$

Moltiplicando queste matrici si vede che A_{e_1} è abeliano. Esso è pertanto un sottogruppo normale abeliano di $\operatorname{Stab} e_1$. È chiaro che contiene tutte le trasformazioni di matrice $T_{21}(b)$, $b \in F^*$. Le formule

$$(1 - e_{11} - e_{22} + e_{12} - e_{21})T_{21}(b)(1 - e_{11} - e_{22} + e_{12} - e_{21})^{-1} = T_{12}(-b),$$
$$T_{jk}(1)T_{ij}(b)T_{jk}(-1)T_{ij}(-b) = T_{ik}(-b)$$

e il fatto che $\mathrm{SL}_n(F)$ sia generato dai $T_{ij}(b)$ implicano che i coniugati di A_{e_1} generano $\mathrm{SL}_n(F)$. Enunciamo il risultato ottenuto.

Lemma 6. *Sia* $\operatorname{Stab} e_1$ *lo stabilizzatore del punto determinato da* $e_1 \neq 0$ *in* $\mathrm{SL}_n(F)$. *Esso contiene un sottogruppo normale abeliano* A_{e_1} *i cui coniugati generano* $\mathrm{SL}_n(F)$.

Dai Lemmi 2, 5 e 6 segue che $\mathrm{SL}_n(F)$, salvo i casi $n = 2$, $|F| = 2$ o 3, soddisfa le ipotesi del criterio di semplicità del Lemma 4. Otteniamo quindi il risultato principale.

Teorema 6.14. *Il gruppo* $\mathrm{PSL}_n(F)$ *è semplice per* $n \geq 2$, *eccettuati i casi* $n = 2$, $|F| = 2$ *o* 3.

Supponiamo ora F finito e determiniamo l'ordine di $\mathrm{PSL}_n(F)$. Sia $|F| = q = p^r$, dove p è la caratteristica di F. Contiamo anzitutto gli elementi di $\mathrm{GL}_n(F)$. Fissata una base $(e_1, e_2, \ldots, e_n)$, a ogni altra base $(f_1, f_2, \ldots, f_n)$ corrisponde una e una sola trasformazione lineare tale che $e_i \mapsto f_i$ per $1 \leq i \leq n$, e così si ottengono tutti gli automorfismi lineari. Pertanto $|\mathrm{GL}_n(F)|$ è il numero delle basi di V. Per il primo vettore f_1 si può prendere qualunque vettore non nullo di V. Poiché ogni vettore si scrive in modo unico nella forma $\sum_{i=1}^n a_i e_i$, i vettori di V sono q^n, e per f_1 abbiamo $q^n - 1$ scelte. Fissato f_1, per f_2 possiamo prendere qualunque vettore che non sia suo multiplo, con $q^n - q$ possibilità. Per scegliere f_3 dobbiamo evitare le q^2 combinazioni lineari $a_1 f_1 + a_2 f_2$ e abbiamo quindi $q^n - q^2$ scelte. Continuando si ottiene

$$|\mathrm{GL}_n(F)| = N = (q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}).$$

L'omomorfismo $A \mapsto \det A$ di $\mathrm{GL}_n(F)$ in F^* è suriettivo, perciò l'immagine ha $|F^*| = q - 1$ elementi. Il nucleo è $\mathrm{SL}_n(F)$. Dunque

$$|\mathrm{SL}_n(F)| = \frac{(q^n - 1) \cdots (q^n - q^{n-1})}{q - 1} = (q^n - 1) \cdots (q^n - q^{n-2})q^{n-1}.$$

Infine vogliamo determinare $|\mathrm{PSL}_n(F)| = |\mathrm{SL}_n(F)|/|C|$. Qui $|C|$ è il numero delle soluzioni di $x^n = 1$ in F^*. Poiché $|F^*| = q - 1$, per ogni $x \in F^*$ si ha $x^{q-1} = 1$. Quindi $|C|$ è il numero delle soluzioni di $x^d = 1$, dove $d = (n, q - 1)$. Essendo F^* ciclico, tali elementi sono d. Abbiamo dunque

$$|\mathrm{PSL}_n(F)| = (q^n - 1)(q^n - q) \cdots (q^n - q^{n-2})q^{n-1}/d, \qquad d = (n, q - 1).$$

Esercizi

1. Determinare la struttura di $\mathrm{PSL}_2(F)$ nei casi $|F| = 2$ e $|F| = 3$.
2. Mostrare che $\mathrm{PSL}_n(F)$ è il gruppo dei commutatori di $\mathrm{PGL}_n(F)$.
3. Mostrare che $\mathrm{GL}_n(F) \trianglerighteq \mathrm{SL}_n(F) \trianglerighteq C \trianglerighteq 1$ è una serie normale in cui tutti i fattori, tranne $\mathrm{SL}_n(F)/C$, sono abeliani.
4. Una trasformazione lineare T si chiama *trasvezione* se esiste un iperpiano U tale che $T|_U = 1_U$ e $Tx - x \in U$ per ogni x. Mostrare che le trasformazioni lineari corrispondenti alle matrici $T_{ij}(b)$, $i \neq j$, $b \in F$, sono trasvezioni. Mostrare che ogni trasvezione τ ha la forma $x \mapsto x + f(x)u$, dove f è una funzione lineare e u è un vettore tale che $f(u) = 0$. Dedurne che, per una trasvezione diversa dall'identità, esiste una base $(e_1, \ldots, e_n)$ di V rispetto alla quale la matrice di τ è $T_{12}(1)$.

5. Sia $|F| = q = p^r$. Mostrare che le matrici triangolari superiori della forma

$$\begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

costituiscono un p-sottogruppo di Sylow di $\mathrm{GL}_n(F)$ e di $\mathrm{SL}_n(F)$.

6. Determinare il normalizzatore N, in $\mathrm{GL}_n(F)$, del sottogruppo H delle matrici diagonali. Mostrare che $N/H \cong S_n$, gruppo simmetrico su n elementi.

6.8 Struttura dei gruppi ortogonali

In questo paragrafo supponiamo che Q sia una forma quadratica non degenere di indice di Witt positivo sullo spazio vettoriale V. Supponiamo dapprima $\dim V = 2$, cosicché V è un piano iperbolico. Dal Teorema 6.10 (p. 365) e dalla sua dimostrazione sappiamo che, se (u, v) è una coppia iperbolica in V, le rotazioni di V sono le applicazioni lineari η_a, $a \in F^*$, tali che $u \mapsto au$, $v \mapsto a^{-1}v$. L'applicazione $a \mapsto \eta_a$ è un isomorfismo del gruppo moltiplicativo F^* sul gruppo delle rotazioni $\mathrm{O}^+(V, Q)$. Le trasformazioni ortogonali improprie sono simmetrie della forma τ_b, dove τ_b manda $u \mapsto bv$, $v \mapsto b^{-1}u$. Verificando sulla base (u, v), si vede che $\tau_b \eta_a \tau_b^{-1} = \eta_a^{-1}$. Poiché inoltre $\tau_b^2 = 1$, il gruppo $\mathrm{O}(V, Q) = \mathrm{O}^+(V, Q) \cup \mathrm{O}^+(V, Q)\tau_b$ è isomorfo a un prodotto semidiretto di un gruppo ciclico di ordine due e di F^* (si veda l'esercizio 9, p. 79).

D'ora in poi supponiamo $\dim V \geq 3$. Sia x un vettore isotropo. Definiremo un sottogruppo H_x di $\mathrm{Stab}\, x$ che svolgerà il ruolo del sottogruppo A_x nel criterio di semplicità del Lemma 4 del paragrafo precedente. Per $u \in (Fx)^\perp$, consideriamo l'applicazione

$$\widetilde{\rho}_{x,u} \colon z \longmapsto z + B(z, u)x, \qquad z \in (Fx)^\perp. \tag{36}$$

È lineare e manda $(Fx)^\perp$ in sé perché $x \in (Fx)^\perp$. Inoltre

$$Q(z + B(z, u)x) = Q(z) + B(z, B(z, u)x) + B(z, u)^2 Q(x) = Q(z),$$

e per $u_1, u_2 \in (Fx)^\perp$ si ha

$$\begin{aligned} \widetilde{\rho}_{x,u_1+u_2} z &= z + B(z, u_1 + u_2)x = z + B(z, u_1)x + B(z, u_2)x \\ &= z + B(z, u_1)x + B(z + B(z, u_1)x, u_2)x = \widetilde{\rho}_{x,u_2}(\widetilde{\rho}_{x,u_1} z). \end{aligned}$$

Poiché $\widetilde{\rho}_{x,0} = 1$ su $(Fx)^\perp$, ne segue che $\widetilde{\rho}_{x,u}$ è invertibile, con $\widetilde{\rho}_{x,-u} = \widetilde{\rho}_{x,u}^{-1}$. È quindi un'isometria di $(Fx)^\perp$ su se stesso. Per il teorema di prolungamento di Witt essa si prolunga a una trasformazione ortogonale di V. Indipendentemente da tale teorema, costruiremo esplicitamente un prolungamento e mostreremo che è unico. Scegliamo, come già più volte, un vettore y tale che $B(x, y) = 1$ e $Q(y) = 0$. Allora $V = (Fx)^\perp \oplus Fy = Fx + Fy + U$, dove $U = (Fx + Fy)^\perp$. Perciò $(Fx)^\perp = U + Fx$. Per $u \in (Fx)^\perp$ possiamo scrivere $u = ax + u'$, $u' \in U$. Poiché $\widetilde{\rho}_{x,ax} = 1$ per la (36), si ha $\widetilde{\rho}_{x,u} = \widetilde{\rho}_{x,u'}$ e possiamo supporre $u \in U$. Definiamo $\rho_{x,u}$ come la trasformazione lineare di V che coincide con $\widetilde{\rho}_{x,u}$ su $(Fx)^\perp$ e manda

$$y \longmapsto ax + by + v, \qquad a, b \in F, \quad v \in U,$$

cercando di determinare a, b, v in modo che $\rho_{x,u} \in \mathrm{O}(V, Q)$. Dato che $V = (Fx)^\perp \oplus Fy$, le condizioni sono

$$Q(\rho_{x,u} y) = Q(y) = 0, \qquad B(\rho_{x,u} y, \rho_{x,u} z) = B(y, z) \quad (z \in (Fx)^\perp).$$

La seconda può essere sostituita dalle due condizioni

$$B(\rho_{x,u}y, x) = B(y, x) = 1, \qquad B(\rho_{x,u}y, \rho_{x,u}z) = 0 \quad (z \in U).$$

Nel complesso si ottengono le tre condizioni $ab + Q(v) = 0$, $b = 1$, $B(z, v) + B(z, u) = 0$ per ogni $z \in U$. Esse hanno un'unica soluzione, $b = 1$, $a = -Q(v)$, $v = -u$, ossia $b = 1$, $a = -Q(u)$, $v = -u$. La definizione di $\rho_{x,u}$ diventa così

$$\begin{aligned} &\rho_{x,u}z = z + B(z, u)x, \qquad z \in (Fx)^\perp, \\ &\rho_{x,u}y = -Q(u)x + y - u, \end{aligned} \tag{37}$$

dove $B(x, y) = 1$, $Q(y) = 0$ e $u \in (Fx + Fy)^\perp$.[9]

Si vede facilmente che $\rho_{x,u}$ è il solo prolungamento ortogonale di $\widetilde{\rho}_{x,u}$. Una volta scelto y (e non occorre cambiarlo), il rappresentante di u che appartiene a U è unico, poiché $(Fx)^\perp = Fx \oplus U$. L'analisi precedente mostra allora che la forma (37) di $\rho_{x,u}$ è unica. Dall'unicità seguono alcune importanti conseguenze.

Abbiamo visto che, per $u_1, u_2 \in (Fx)^\perp$, $\widetilde{\rho}_{x,u_1+u_2} = \widetilde{\rho}_{x,u_1}\widetilde{\rho}_{x,u_2}$. Ora $\rho_{x,u_1}\rho_{x,u_2}$ è un prolungamento ortogonale di $\widetilde{\rho}_{x,u_1}\widetilde{\rho}_{x,u_2} = \widetilde{\rho}_{x,u_1+u_2}$. Poiché anche ρ_{x,u_1+u_2} è un tale prolungamento, dall'unicità segue

$$\rho_{x,u_1}\rho_{x,u_2} = \rho_{x,u_1+u_2}. \tag{38}$$

È chiaro inoltre che $\rho_{x,u} = 1$ se e solo se $\widetilde{\rho}_{x,u} = 1$ su $(Fx)^\perp$. Dalla definizione (36) ciò accade se e solo se $u \in \mathrm{rad}(Fx)^\perp$. Poiché questo radicale è Fx, si ha

$$\rho_{x,u} = 1 \quad \Longleftrightarrow \quad u \in Fx. \tag{39}$$

Analogamente, per $\eta \in \mathrm{O}(V, Q)$, si ha

$$\eta\rho_{x,u}\eta^{-1} = \rho_{\eta x,\eta u}. \tag{40}$$

Abbiamo visto che possiamo scegliere $u \in U = (Fx + Fy)^\perp$. Con tale scelta la (39) dà $\rho_{x,u} = 1$ se e solo se $u = 0$. Per la (38), $u \mapsto \rho_{x,u}$ è quindi un monomorfismo del gruppo additivo di U in $\mathrm{O}(V, Q)$. Indichiamo con H_x l'immagine $\{\rho_{x,u}\}$. Dalla definizione di $\widetilde{\rho}_{x,u}$ risulta $\widetilde{\rho}_{ax,u} = \widetilde{\rho}_{x,au}$ per $a \in F^*$; dunque anche

$$\rho_{ax,u} = \rho_{x,au}, \qquad a \in F^*. \tag{41}$$

Ne segue $H_{ax} = H_x$ per ogni $a \in F^*$. Chiaramente H_x è un sottogruppo abeliano di $\mathrm{O}(V, Q)$. Inoltre $\rho_{x,u}x = x + B(x, u)x = x$, e perciò $H_x \subseteq \mathrm{Stab}\, x$. Per $\eta \in \mathrm{Stab}\, x$ la (40) diventa $\eta\rho_{x,u}\eta^{-1} = \rho_{x,\eta u}$, da cui $H_x \trianglelefteq \mathrm{Stab}\, x$. Riuniamo parte di questi risultati nel lemma seguente.

Lemma 1. *Sia x un vettore isotropo di V e sia H_x l'insieme delle trasformazioni lineari di V definite dalla (37), dove u varia in $U = (Fx + Fy)^\perp$ e y è tale che $B(x, y) = 1$, $Q(y) = 0$. Allora H_x è un sottogruppo normale abeliano di* $\mathrm{Stab}\, x$ *in* $\mathrm{O}(V, Q)$ *e $u \mapsto \rho_{x,u}$ è un isomorfismo del gruppo additivo di U su H_x.*

Osserviamo inoltre che le trasformazioni $\rho_{x,u}$ sono unipotenti. Infatti, posto $\nu_{x,u} = \rho_{x,u} - 1$, dalla definizione segue che $\nu_{x,u}x = 0$, $\nu_{x,u}z \in Fx$ per $z \in (Fx)^\perp = Fx + U$, e $\nu_{x,u}y \in (Fx)^\perp$. Pertanto $\nu_{x,u}^2 z = 0 = \nu_{x,u}^3 y$. Dunque $\nu_{x,u}^3 = 0$ e $\rho_{x,u} = 1 + \nu_{x,u}$ è unipotente. Abbiamo visto che una trasformazione ortogonale unipotente è necessariamente una rotazione; perciò $H_x \subseteq \mathrm{O}^+(V, Q)$.

[9] Queste trasformazioni sembrano essere state introdotte per la prima volta da C. L. Siegel in «Über die Zetafunktionen indefiniter quadratischer Formen II», *Mathematische Zeitschrift*, vol. 44 (1938), pp. 398–426.

Introduciamo il gruppo Ω, definito come sottogruppo di $\mathrm{O}(V,Q)$ generato da tutti gli H_x con x isotropo. Se η è ortogonale, dalla (40) segue $\eta H_x \eta^{-1} = H_{\eta x}$, dunque Ω è normale in $\mathrm{O}(V,Q)$. Mostreremo che Ω coincide con il gruppo dei commutatori di $\mathrm{O}(V,Q)$ e vedremo che, salvo il caso $n = 4$ e indice di Witt $\nu = 2$, con l'ulteriore eccezione indicata nel Teorema 6.15, Ω modulo il suo centro è semplice. La dimostrazione che daremo è dovuta a Tamagawa[10] e segue il metodo di Iwasawa, basato sul criterio di semplicità del § 6.7. Per applicarlo useremo un'azione di Ω su un cono quadrico dello spazio proiettivo $P_{n-1}(F)$. Nello spazio vettoriale V sia C l'insieme dei vettori x con $Q(x) = 0$. Se $(e_1, \dots, e_n)$ è una base di V e $x = \sum_i a_i e_i$, la condizione $Q(x) = 0$ equivale all'equazione quadratica $\sum_{i,j} b_{ij} a_i a_j = 0$ nelle coordinate (a_i) di x, dove $b_{ij} = B(e_i, e_j)$. Quindi C è un cono quadrico in V. Sia PC il cono quadrico corrispondente in $P_{n-1}(F)$, cioè l'insieme dei sottospazi unidimensionali Fx determinati dai vettori isotropi x di V. Ogni $\eta \in \mathrm{O}(V,Q)$ permuta i punti di C; abbiamo così un'azione $Fx \mapsto F(\eta x)$ di $\mathrm{O}(V,Q)$ su PC. Mostreremo che il suo nucleo è $\{1, -1\}$. Dimostriamo anzitutto il lemma seguente.

Lemma 2. *Il centro di* $\mathrm{O}(V,Q)$ *è* $\{1, -1\}$.

Dimostrazione. Sia γ nel centro e sia S_u la simmetria determinata dal vettore non isotropo u. Poiché Fu è l'insieme degli x tali che $S_u x = -x$ e $S_u(\gamma u) = \gamma(S_u u) = -\gamma u$, si ha $\gamma u \in Fu$. Essendo γ ortogonale, $\gamma u = \pm u$ per ogni u non isotropo. Sia $(u_1, \dots, u_n)$ una base ortogonale di V. Si ha allora $\gamma u_i = \varepsilon_i u_i$, con $\varepsilon_i = \pm 1$. Per $i \neq j$, supponiamo dapprima che $u_i + u_j$ non sia isotropo. Allora $\gamma(u_i + u_j) = \pm(u_i + u_j)$ e anche $\gamma(u_i + u_j) = \varepsilon_i u_i + \varepsilon_j u_j$, da cui $\varepsilon_i = \varepsilon_j$. Se invece $u_i + u_j$ è isotropo, scegliamo $k \neq i, j$ e consideriamo $u_i + u_j + u_k$. Poiché tale vettore non è isotropo, si ha

$$\gamma(u_i + u_j + u_k) = \pm(u_i + u_j + u_k) = \varepsilon_i u_i + \varepsilon_j u_j + \varepsilon_k u_k.$$

Di nuovo $\varepsilon_i = \varepsilon_j$. Dunque $\gamma u_i = \varepsilon u_i$ per ogni i, con $\varepsilon = \pm 1$. Gli u_i costituiscono una base, dunque $\gamma = \pm 1$. □

Poiché 1 e -1 inducono l'identità su PC, appartengono al nucleo dell'azione di $\mathrm{O}(V,Q)$ su PC. Il lemma seguente mostra che $\{1, -1\}$ è l'intero nucleo.

Lemma 3. *Se* $\eta \in \mathrm{O}(V,Q)$ *soddisfa* $\eta x \in Fx$ *per ogni* x *isotropo, allora* $\eta = \pm 1$.

Dimostrazione. Sia (u, v) una coppia iperbolica e sia $z \in (Fu + Fv)^\perp$. Allora $x = z - Q(z)u + v$ è isotropo, perché $Q(x) = Q(z) - Q(z)B(u,v) = Q(z) - Q(z) = 0$. Si ha dunque $\eta u = c_u u$, $\eta v = c_v v$, $\eta x = c_x x$, con c_u, c_v, c_x elementi non nulli di F. Pertanto

$$c_x(z - Q(z)u + v) = \eta x = \eta z - c_u Q(z) u + c_v v.$$

Poiché $\eta z \in (Fu + Fv)^\perp$, segue $\eta z = c_x z$ e $c_x = c_v$; scegliendo z con $Q(z) \neq 0$ si ottiene anche $c_v = c_u$. Dunque $\eta = c1$ con $c = c_v$, e poiché η è ortogonale, $c = \pm 1$. □

Restringendo a Ω l'azione di $\mathrm{O}(V,Q)$ su PC si ottiene un'azione con nucleo $\Omega \cap \{1, -1\}$. Poiché $\Omega \subseteq \mathrm{O}^+(V,Q)$, se $n = \dim V$ è dispari $-1 \notin \Omega$. In tal caso l'azione di Ω su PC è fedele. Esaminiamola ora più da vicino.

Lemma 4. *Siano* $T_x = C \cap (Fx)^\perp$ *e* $PT_x = \{Fy \neq 0 \mid y \in T_x\}$. *Allora* H_x *agisce transitivamente sul complemento di* PT_x *in* PC.

Dimostrazione. Se y, z sono vettori isotropi non ortogonali a x, dobbiamo trovare $\rho_{x,u} \in H_x$ tale che $\rho_{x,u} y \in Fz$. Possiamo supporre $B(y,x) = 1 = B(z,x)$. Si ha $V = Fy + (Fx)^\perp = Fy \oplus Fx \oplus U$, dove $U = (Fx + Fy)^\perp$. Dunque $z = ay + bx + u$, $u \in U$. Da $B(z,x) = 1$ segue $a = 1$, e da $Q(z) = 0$ segue $b + Q(u) = 0$. Quindi $z = y - Q(u)x + u$, e per definizione $\rho_{x,-u} y = z$. □

[10] T. Tamagawa, «On the structure of orthogonal groups», *American Journal of Mathematics*, vol. 80 (1958), pp. 191–197. Una versione migliorata della dimostrazione compare in alcune dispense ciclostilate di Tamagawa.

Una coppia di punti (Fx, Fy) di PC si dice *iperbolica* se $B(x, y) \neq 0$. In tal caso possiamo supporre che (x, y) sia una coppia iperbolica di vettori di V. Il lemma precedente mostra che, se (Fx, Fy) e (Fx, Fz) sono iperboliche, esiste un elemento di Ω che fissa Fx e manda Fy in Fz.

Lemma 5. *Il gruppo Ω è transitivo su PC e sull'insieme delle coppie iperboliche ordinate di punti di PC.*

Dimostrazione. Siano Fx, Fy punti distinti di PC. Affermiamo che esiste $Fz \in PC$ tale che entrambe le coppie (Fz, Fx) e (Fz, Fy) siano iperboliche. Supponiamo dapprima (Fx, Fy) iperbolica; possiamo allora scegliere (x, y) coppia iperbolica. Sia u un vettore non isotropo di $U = (Fx + Fy)^\perp$ e si ponga $z = x - Q(u)y + u$. Allora $Q(z) = -Q(u)B(x, y) + Q(u) = 0$, $B(z, x) = -Q(u) \neq 0$ e $B(z, y) = 1$. Quindi Fz soddisfa la richiesta. Supponiamo ora $B(x, y) = 0$. Poiché x, y sono linearmente indipendenti, esiste una funzione lineare che manda entrambi in 1. Esiste perciò $z \in V$ tale che $B(x, z) = 1 = B(y, z)$. Sottraendo a z un opportuno multiplo di x, possiamo ottenere $Q(z) = 0$. Allora (Fz, Fx) e (Fz, Fy) sono iperboliche e Fz ha di nuovo la proprietà richiesta. Possiamo adesso applicare il Lemma 4 per ottenere $\eta \in \Omega$ tale che $\eta(Fx) = Fy$. Questo dimostra la transitività di Ω su PC. Siano infine (Fx, Fy) e (Fx', Fy') coppie iperboliche. Esiste $\eta \in \Omega$ con $\eta(Fx) = Fx'$. Anche $(Fx', \eta(Fy))$ è iperbolica. Come osservato sopra, esiste $\zeta \in \Omega$ con $\zeta(Fx') = Fx'$ e $\zeta(\eta(Fy)) = Fy'$. Allora $\zeta\eta$ manda (Fx, Fy) in (Fx', Fy'), e ciò dimostra la seconda affermazione. □

Possiamo ora dimostrare il risultato sull'azione di Ω su PC che più ci interessa.

Lemma 6. *Il gruppo Ω agisce primitivamente su PC, salvo il caso* $\dim V = 4$ *e* $\nu(Q) = 2$.

Dimostrazione. Supponiamo anzitutto $\nu(Q) = 1$. In tal caso ogni coppia di punti distinti (Fx, Fy) di PC è iperbolica e per il Lemma 5 Ω è doppiamente transitivo su PC. La sua azione è quindi primitiva. Supponiamo ora $\nu \geq 2$. Escludendo il caso $\dim V = 4$, $\nu = 2$, si ha $\dim V \geq 5$. Sia S una parte contenente più di un punto di una partizione di PC stabile per Ω. Per la primitività basta mostrare $S = PC$. Supponiamo dapprima che S contenga due punti distinti Fx, Fy tali che $B(x, y) = 0$. Possiamo trovare un vettore isotropo z tale che $B(x, z) = 1$, $B(y, z) = 0$. Si ha $V = (Fx + Fz) \oplus U$, dove $U = (Fx + Fz)^\perp$ è non degenere e ha dimensione almeno tre. Poiché $y \in U$, esiste $w \in U$ tale che (y, w) è una coppia iperbolica. Dato che $\dim U \geq 3$ e U contiene vettori isotropi, possiamo applicare il Lemma 5 a U rispetto alla restrizione di Q. Esiste perciò una trasformazione ortogonale η, prodotto di trasformazioni $\rho_{u,v}$ con $u, v \in U$, tale che $\eta(Fy) = Fw$. Poiché $x \in (Fu + Fv)^\perp$, la (37) mostra che $\eta(Fx) = Fx$. Da $Fx \in S$ segue $\eta S = S$, e da $Fy \in S$ segue $Fw \in S$. Quindi S contiene Fy e Fw, che formano una coppia iperbolica.

Possiamo dunque supporre che S contenga due punti Fx, Fy con (Fx, Fy) iperbolica. Sia Fz un punto qualsiasi di PC distinto da Fx. Abbiamo visto che esiste Fw tale che (Fx, Fw) e (Fz, Fw) sono iperboliche. Applicando il Lemma 5 a (Fx, Fy) e (Fx, Fw), otteniamo $\eta \in \Omega$ tale che $\eta(Fx) = Fx$ e $\eta(Fy) = Fw$. Allora $\eta S = S$ e $Fw \in S$, poiché $Fy \in S$. Applichiamo ora il Lemma 5 a (Fx, Fw) e (Fz, Fw): esiste $\zeta \in \Omega$ tale che $\zeta(Fx) = Fz$ e $\zeta(Fw) = Fw$. Poiché $Fx, Fw \in S$, ne segue $\zeta S = S$ e $Fz \in S$. Abbiamo scelto Fz arbitrariamente in PC fuori di Fx, dunque $S = PC$. □

Per dimostrare il teorema di semplicità ci occorre un ultimo lemma.

Lemma 7. *Il gruppo Ω contiene il sottogruppo dei commutatori* $(\mathrm{O}(V, Q), \mathrm{O}(V, Q))$ *di* $\mathrm{O}(V, Q)$.

Dimostrazione. Siano (x, y) una coppia iperbolica e u un vettore non isotropo. Affermiamo che esiste $\rho \in \Omega$ tale che $\rho u \in Fx + Fy$. Osserviamo che $u_1 = x + Q(u)y$ soddisfa $Q(u_1) = Q(u)$. Esiste dunque $\eta \in \mathrm{O}(V, Q)$ tale che $\eta u_1 = u$. Per il Lemma 5 esiste $\rho \in \Omega$ tale che $\rho(F(\eta x)) = Fx$ e $\rho(F(\eta y)) = Fy$. Poiché $u_1 \in Fx + Fy$, si ha $u = \eta u_1 \in \eta(Fx) + \eta(Fy)$, quindi $\rho u \in \rho F(\eta x) + \rho F(\eta y) = Fx + Fy$, come volevamo. Ne segue che per ogni simmetria S_u esiste

$\rho \in \Omega$ tale che $\rho S_u \rho^{-1} = S_{u'}$, dove $u' = \rho u \in Fx + Fy$.

Sia $O_{x,y}$ il sottogruppo di $\mathrm{O}(V,Q)$ generato dalle simmetrie $S_{u'}$, $u' \in Fx + Fy$. Poiché la restrizione di $S_{u'}$ a $U = (Fx + Fy)^\perp$ è l'identità, si vede che $\eta' \mapsto \eta'|_{Fx+Fy}$ è un isomorfismo di $O_{x,y}$ su $\mathrm{O}(Fx + Fy, Q)$. Esso manda il sottogruppo $O^+_{x,y}$ generato dai prodotti di coppie di simmetrie determinate da $u' \in Fx+Fy$ su $\mathrm{O}^+(Fx+Fy,Q)$. Sia ora ζ una rotazione qualsiasi di V e si scriva $\zeta = S_{u_1} S_{u_2} \cdots S_{u_{2k}}$, con gli u_i non isotropi. Per quanto appena dimostrato esistono $\rho_i \in \Omega$ tali che $u'_i = \rho_i u_i \in Fx + Fy$. Dunque

$$\zeta = S_{u_1} S_{u_2} \cdots S_{u_{2k}} = (\rho_1^{-1} S_{u'_1} \rho_1) \cdots (\rho_{2k}^{-1} S_{u'_{2k}} \rho_{2k}). \tag{42}$$

Poiché Ω è normale in $\mathrm{O}(V,Q)$, questa relazione implica $\zeta = \rho S_{u'_1} \cdots S_{u'_{2k}}$ per un $\rho \in \Omega$. Quindi

$$\mathrm{O}^+(V,Q) \subseteq \Omega O^+_{x,y}. \tag{43}$$

D'altra parte abbiamo già visto che $\Omega \subseteq \mathrm{O}^+(V,Q)$, perciò

$$\mathrm{O}^+(V,Q) = \Omega O^+_{x,y}. \tag{44}$$

Segue

$$\mathrm{O}^+(V,Q)/\Omega \cong O^+_{x,y}/(O^+_{x,y} \cap \Omega).$$

Poiché $O^+_{x,y} \cong \mathrm{O}^+(Fx + Fy, Q)$ e $Fx + Fy$ è un piano iperbolico, $O^+_{x,y}$ è abeliano. Quindi $\mathrm{O}^+(V,Q)/\Omega$ è abeliano e

$$\Omega \supseteq (\mathrm{O}^+(V,Q), \mathrm{O}^+(V,Q)). \tag{45}$$

Abbiamo però già dimostrato (Teorema 6.13, p. 374) che

$$(\mathrm{O}(V,Q), \mathrm{O}(V,Q)) = (\mathrm{O}^+(V,Q), \mathrm{O}^+(V,Q)).$$

Dunque

$$\Omega \supseteq (\mathrm{O}(V,Q), \mathrm{O}(V,Q)). \tag{46}$$

□

Possiamo ora dimostrare il teorema principale sulla struttura dei gruppi ortogonali.

Teorema 6.15 (Dickson–Dieudonné). *Sia Q una forma quadratica non degenere di indice di Witt positivo ν su uno spazio vettoriale V di dimensione $n \geq 3$. Allora il quoziente del gruppo dei commutatori del gruppo ortogonale $\mathrm{O}(V,Q)$ rispetto al suo centro è semplice, salvo i casi $n = 4$, $\nu = 2$, e $n = 3$, $|F| = 3$.*

Dimostrazione. Mostreremo dapprima che $\Omega = (\mathrm{O}(V,Q), \mathrm{O}(V,Q))$ e $(\Omega, \Omega) = \Omega$. Poiché $\mathrm{O}(V,Q) \supseteq \Omega \supseteq (\mathrm{O}(V,Q), \mathrm{O}(V,Q))$, la prima uguaglianza seguirà dalla seconda. Per provare $\Omega = (\Omega, \Omega)$ basta mostrare che $\rho_{x,u} \in (\Omega, \Omega)$ per ogni x isotropo e $u \in (Fx)^\perp$. Scegliamo y tale che (x, y) sia una coppia iperbolica e sia $O_{x,y}$ il sottogruppo definito nella dimostrazione del Lemma 7. Si ha $O_{x,y} \cong \mathrm{O}(Fx + Fy, Q)$. Per ogni $a \in F^*$ esiste $\eta_a \in O_{x,y}$ tale che $\eta_a x = ax$, $\eta_a y = a^{-1} y$, e vi è $\tau \in O_{x,y}$ con $\tau x = y$, $\tau y = x$. Allora $\tau^{-1} \eta_a^{-1} \tau \eta_a = \eta_a^2 \in (\mathrm{O}(V,Q), \mathrm{O}(V,Q)) \subseteq \Omega$. Nel considerare $\rho_{x,u}$ possiamo supporre $u \in U = (Fx + Fy)^\perp$ (cfr. p. 383 per questa e altre proprietà delle trasformazioni $\rho_{x,u}$). Si ha

$$\eta_{a^2} \rho_{x,u} \eta_{a^2}^{-1} \rho_{x,u}^{-1} = \rho_{a^2 x, u} \rho_{x,-u} = \rho_{x, a^2 u} \rho_{x,-u} = \rho_{x,(a^2-1)u}.$$

Se $|F| \geq 4$, possiamo scegliere $a \in F^*$ con $a^2 \neq 1$. Sostituendo u con $(a^2 - 1)^{-1} u$, si vede che $\rho_{x,u}$ è un commutatore di elementi di Ω. In questo caso $\Omega = (\Omega, \Omega) = (\mathrm{O}(V,Q), \mathrm{O}(V,Q))$.

Supponiamo ora $|F| = 3$. Tenendo conto dei casi esclusi, $n \geq 4$ e, se $n = 4$, $\nu = 1$. Per la formula $\rho_{x,u_1}\rho_{x,u_2} = \rho_{x,u_1+u_2}$ e per l'esistenza di una base ortogonale di U, basta mostrare che ogni $\rho_{x,u}$, con $u \in U$ non isotropo, appartiene a (Ω, Ω). Se $n = 4$, $\nu = 1$, lo spazio U ha dimensione due ed è anisotropo; possiamo quindi completare u a una base ortogonale (u, v) di U. Poiché $Q(u) = -Q(v)$ implicherebbe che U fosse iperbolico, si ha $Q(u) = 1 = Q(v)$ oppure $Q(u) = -1 = Q(v)$. Se $n \geq 5$, U ha dimensione almeno tre ed è non degenere. Il complemento ortogonale $(Fu)^\perp \cap U$ di Fu in U è non degenere e ha dimensione almeno due. La restrizione di Q a tale spazio è dunque universale, cosicché esiste di nuovo un vettore $v \perp u$ con $Q(v) = Q(u)$. Esiste allora una trasformazione ortogonale τ tale che $\tau x = x$, $\tau u = -v$, $\tau v = u$. Si ha $\tau^2 x = x$, $\tau^2 u = -u$, $\tau^2 v = -v$, e pertanto

$$\tau^2 \rho_{x,u} \tau^{-2} \rho_{x,u}^{-1} = \rho_{x,-u}\rho_{x,-u} = \rho_{x,-2u} = \rho_{x,u}.$$

Osserviamo ora che $\tau^2 \in (\mathrm{O}(V,Q), \mathrm{O}(V,Q))$. Infatti, in ogni gruppo G generato da elementi g_i di ordine due, il quadrato di ciascun elemento appartiene al gruppo dei commutatori. Sia $\Gamma = \{g_i\}$ un insieme di generatori con $g_i^2 = 1$, cosicché $g_i^{-1} = g_i$. Per $g = g_1 \cdots g_k$ si ha

$$g^2 = g_1 \cdots g_k g_1 \cdots g_k = g_1 \cdots g_{k-1} g_1 \cdots g_{k-1} \big((g')^{-1} g_k^{-1} g' g_k\big),$$

dove $g' = g_1 \cdots g_{k-1}$. Per induzione su k si conclude $g^2 \in (G, G)$. Essendo $\mathrm{O}(V,Q)$ generato da simmetrie, si ha $\tau^2 \in (\mathrm{O}(V,Q), \mathrm{O}(V,Q)) \subseteq \Omega$. Ne segue $\rho_{x,u} \in (\Omega, \Omega)$, e di nuovo $\Omega = (\Omega, \Omega) = (\mathrm{O}(V,Q), \mathrm{O}(V,Q))$.

Per completare la dimostrazione verifichiamo il criterio di semplicità (Lemma 4, p. 379) per $\Omega = (\mathrm{O}(V,Q), \mathrm{O}(V,Q))$. Abbiamo visto che Ω agisce primitivamente sul cono quadrico proiettivo PC (Lemma 6) e che $\Omega = (\Omega, \Omega)$. Inoltre, per ogni $Fx \in PC$, $\mathrm{Stab}(Fx)$ contiene il sottogruppo normale abeliano H_x. Poiché Ω è transitivo su PC (Lemma 5), due sottogruppi H_x, H_y con x, y isotropi sono sempre coniugati in Ω. Quindi Ω, generato da tutti gli H_x, è generato dai coniugati di uno di questi sottogruppi. Sono così verificate tutte le condizioni del criterio e Ω/K è semplice, dove K è il nucleo dell'azione. Per il Lemma 3 si ha $K = \Omega \cap \{1, -1\}$, e la dimostrazione è conclusa. □

Il teorema precedente è il risultato fondamentale sulla struttura dei gruppi ortogonali. Ci si può chiedere se l'ipotesi che Q abbia indice di Witt positivo sia necessaria per la semplicità di $P\Omega = \Omega/(\Omega \cap \{1, -1\})$ e se occorra escludere i casi $n = 4$, $\nu = 2$, e $n = 3$, $|F| = 3$. Negli esercizi si vedrà che queste restrizioni sono necessarie per garantire la semplicità. Un'altra domanda naturale riguarda l'intersezione $\Omega \cap \{1, -1\}$: si ha $-1 \in \Omega$? Se n è dispari, -1 è impropria e quindi $-1 \notin \Omega$. In tal caso Ω è semplice, nelle ipotesi del teorema. Se n è pari, si può mostrare che $-1 \in \Omega$ se e solo se il discriminante della forma bilineare associata a Q è 1, ossia F^{*2}. Questo si può dimostrare usando le algebre di Clifford, uno strumento importante nello studio dei gruppi ortogonali. La dimostrazione si trova nel § 4.8 del volume II.

Esercizi

1. Sia $F = \mathbb{R}((x))$ il campo delle serie formali $\sum_{k=r}^{\infty} a_k x^k$, $a_k \in \mathbb{R}$, $r \in \mathbb{Z}$, definito come campo delle frazioni di $\mathbb{R}[[x]]$ (si veda l'esercizio 7, p. 127). Si chiami *ordine* di una tale serie il numero r se $a_r \neq 0$, e si dica *intera* la serie se $r \geq 0$. Si ponga uguale a ∞ l'ordine di 0. Allora

   $$\mathrm{ord}(uv) = \mathrm{ord}\, u + \mathrm{ord}\, v, \qquad \mathrm{ord}(u+v) \geq \min(\mathrm{ord}\, u, \mathrm{ord}\, v)$$

   per $u, v \in F$, dove ord indica l'ordine. Mostrare che l'insieme I delle serie di ordine almeno zero è un sottoanello di F e che, per ogni $k > 0$, l'insieme P_k delle serie di ordine almeno k è un ideale di I.

2. Sia F come nell'esercizio 1 e sia Q la forma quadratica definita in coordinate da $Q = \sum_{i=1}^{n} x_i^2$. Allora $\mathrm{O}(V,Q)$ è isomorfo al gruppo delle matrici ortogonali nel senso usuale, $A\,{}^tA = 1$. Mostrare che ne segue l'integralità di A, cioè $A \in M_n(I)$. Per $k > 0$ sia G_k l'insieme delle matrici ortogonali della forma $1 + B$ con $B \in M_n(P_k)$. Verificare che G_k è un sottogruppo normale del gruppo delle matrici ortogonali e che $\bigcap_k G_k = 1$. Dimostrare che G_k/G_{k+1} è abeliano e $G_k \supsetneq G_{k+1}$. Dedurne che $(\mathrm{O}(V,Q), \mathrm{O}(V,Q))$ modulo il proprio centro non è semplice.

3. Mostrare che Ω non agisce primitivamente su PC se $n = 4$ e $\nu = 2$, e dedurne che in questo caso $P\Omega$ non è semplice.

6.9 Geometria simplettica. Il gruppo simplettico

Si chiama *geometria simplettica* lo studio di uno spazio vettoriale V di dimensione finita rispetto a una forma bilineare alternante non degenere B. Sappiamo che la dimensione di un tale spazio è pari e abbiamo chiamato *gruppo simplettico* il gruppo delle trasformazioni lineari biiettive η di V tali che $B(\eta x, \eta y) = B(x,y)$ per $x, y \in V$. Poiché due forme alternanti non degeneri su spazi della stessa dimensione sono equivalenti, è superfluo indicare la dipendenza da B. Denotiamo quindi il gruppo simplettico con $\mathrm{Sp}_n(F)$, dove F è il campo di base e $n = 2r$ la dimensione di V. Lo studio di questo gruppo e della geometria associata è analogo, e più semplice, di quello dei gruppi ortogonali. Possiamo dunque trattare il caso simplettico con relativa brevità.

Stabiliamo anzitutto l'analogo del teorema di prolungamento di Witt. Abbiamo mostrato (§ 6.2, p. 351) che V ha una base $(u_1, v_1, u_2, v_2, \dots, u_r, v_r)$ tale che

$$B(u_i, u_j) = 0 = B(v_i, v_j), \qquad B(u_i, v_j) = \delta_{ij} = -B(v_j, u_i). \tag{47}$$

Una tale base si chiamerà *base simplettica* di V. Se (u_i, v_i) è una base simplettica e $\eta \in \mathrm{Sp}_n(F)$, anche $(\eta u_i, \eta v_i)$ è una base simplettica. Viceversa, date due basi simplettiche (u_i, v_i) e (u_i', v_i'), la trasformazione lineare η tale che $u_i \mapsto u_i'$, $v_i \mapsto v_i'$, $1 \le i \le r$, è simplettica. Se U è un sottospazio di V, possiamo trovarne una base $(u_1, v_1, \dots, u_k, v_k, u_{k+1}, \dots, u_m)$ tale che $(u_{k+1}, \dots, u_m)$ sia una base del radicale $U \cap U^\perp$ di U e gli u_i, v_i, $1 \le i \le k$, soddisfino la (47). Come nel caso ortogonale (p. 368), possiamo trovare vettori $v_{k+1}, \dots, v_m$ tali che i $2m$ vettori u_j, v_j, $1 \le j \le m$, soddisfino la (47). L'argomento usato nel caso ortogonale dimostra ora che, se η è un'applicazione lineare da U in un altro sottospazio U' ed è un'isometria, ossia è biiettiva e soddisfa $B(\eta x, \eta y) = B(x,y)$ per $x, y \in U$, si può prolungare η a una trasformazione simplettica di V.

Introduciamo un particolare sistema di generatori di $\mathrm{Sp}_n(F)$. Sia u un vettore non nullo di V e sia $c \in F$. Definiamo

$$\tau_{u,c} \colon x \longmapsto x + cB(x,u)u. \tag{48}$$

Una verifica diretta mostra che $B(\tau_{u,c}x, \tau_{u,c}y) = B(x,y)$ per $x, y \in V$ e che

$$\tau_{u,c_1}\tau_{u,c_2} = \tau_{u,c_1+c_2}. \tag{49}$$

Inoltre $\tau_{u,c} = 1$ se e solo se $c = 0$. Pertanto $c \mapsto \tau_{u,c}$ è un monomorfismo del gruppo additivo di F in $\mathrm{Sp}_n(F)$. Chiamiamo $\tau_{u,c}$ una *trasvezione simplettica* di direzione u. Per $\eta \in \mathrm{Sp}_n(F)$ si ha

$$\eta\tau_{u,c}\eta^{-1} = \tau_{\eta u,c}, \tag{50}$$

e per $a \in F^*$ si ha

$$\tau_{au,c} = \tau_{u,a^2c}. \tag{51}$$

Dalla (48) è chiaro che $\tau_{u,c}x = x$ per $x \in (Fu)^\perp$; in particolare $\tau_{u,c}u = u$. Inoltre $\zeta_{u,c} = \tau_{u,c} - 1$ manda ogni x in Fu; dunque $\zeta_{u,c}^2 = 0$. Perciò $\tau_{u,c}$ è unipotente e ha determinante 1.

Lemma 1. *Il gruppo* $\mathrm{Sp}_n(F)$ *è generato dalle trasvezioni simplettiche.*

Dimostrazione. Basterà mostrare che, date due coppie di vettori (u, v), (u', v') iperboliche nel senso che $B(u, v) = 1 = B(u', v')$, esiste un prodotto di trasvezioni simplettiche che manda $u \mapsto u'$, $v \mapsto v'$. Supponiamo infatti di avere questa proprietà e sia $\eta \in \mathrm{Sp}_n(F)$. Se (u, v) è una coppia iperbolica, anche $(\eta u, \eta v)$ lo è, perciò esiste un prodotto di trasvezioni ζ tale che $\zeta u = \eta u$, $\zeta v = \eta v$. Allora $\eta' = \zeta^{-1}\eta$ fissa u e v ed essendo simplettica lascia stabile $U = (Fu + Fv)^\perp$. La restrizione $\eta'|_U$ è simplettica e, poiché $\dim U = n - 2$, si può supporre per induzione che sia prodotto di trasvezioni di direzioni appartenenti a U. Il prodotto delle trasvezioni di V determinate dagli stessi vettori è una trasformazione simplettica che coincide con l'identità su $Fu + Fv$ e con η' su U. Poiché $V = Fu + Fv + U$, essa coincide con η'. Dunque η' e di conseguenza $\eta = \zeta\eta'$ sono prodotti di trasvezioni simplettiche. Rimane così da mostrare che esiste un prodotto di trasvezioni simplettiche che mandi una coppia iperbolica (u, v) in (u', v'). Procederemo in due passi:

$$(u, v) \longmapsto (u', v'') \longmapsto (u', v'). \tag{52}$$

Nel primo passo costruiamo un prodotto di trasvezioni simplettiche che mandi u in u'. Supponiamo dapprima $B(u, u') \neq 0$. Allora $u \neq u'$ e, posto $w = u - u'$, si ha $w \neq 0$ e

$$\tau_{w,c}u = u + cB(u, w)w = u - cB(u, u')(u - u').$$

Per $c = B(u, u')^{-1}$ otteniamo dunque $\tau_{w,c}u = u'$. Supponiamo adesso $B(u, u') = 0$. Ci riconduciamo al caso precedente osservando che esiste u'' tale che $B(u, u'') \neq 0$ e $B(u', u'') \neq 0$. Esiste infatti una funzione lineare f su V tale che $f(u) \neq 0$ e $f(u') \neq 0$. Per la non degenerazione di B si ha $f(x) = B(x, u'')$ per un opportuno $u'' \in V$ (lemma di p. 347), e dunque $B(u, u'') \neq 0$, $B(u', u'') \neq 0$. Con una trasvezione simplettica si passa da u a u'' e con un'altra da u'' a u'. Il prodotto delle due porta quindi u in u'. È compiuto il primo passo della (52). Per il secondo, cambiando notazione, dobbiamo mostrare che, se (u, v) e (u, v') sono coppie iperboliche, esiste un prodotto di trasvezioni simplettiche che fissi u e mandi v in v'. Cominciamo ancora dal caso $B(v, v') \neq 0$. Si passa da v a v' con una trasvezione $\tau_{w,c}$ dove $w = v - v'$. Essa fissa anche u, perché $B(u, v) = 1 = B(u, v')$ e quindi $B(u, w) = 0$. Abbiamo concluso se $B(v, v') \neq 0$. Se invece $B(v, v') = 0$, interponiamo fra (u, v) e (u, v') la coppia $(u, u + v)$, anch'essa iperbolica perché $B(u, u + v) = B(u, u) + B(u, v) = 1$. Poiché $B(v, u + v) = B(v, u) = -1$ e $B(u + v, v') = B(u, v') = 1$, ci troviamo nel primo caso sia per le coppie (u, v) e $(u, u + v)$ sia per le coppie $(u, u + v)$ e (u, v'). Possiamo quindi passare da (u, v) a (u, v') con un prodotto di trasvezioni simplettiche. □

Dal lemma e dal fatto che $\det \tau_{u,c} = 1$ per ogni trasvezione simplettica segue subito $\det \eta = 1$ per ogni trasformazione simplettica. La generazione mediante trasvezioni permette anche di dimostrare il lemma seguente.

Lemma 2. *Il centro di* $\mathrm{Sp}_n(F)$ *è costituito dalle trasformazioni* 1 *e* -1.

Dimostrazione. Sia γ nel centro di $\mathrm{Sp}_n(F)$. Allora γ commuta con ogni trasvezione $\tau_{u,c}$. Se $c \neq 0$, i punti fissi di $\tau_{u,c}$ formano $(Fu)^\perp$. Poiché γ commuta con $\tau_{u,c}$, manda ogni suo punto fisso in un punto fisso. Dunque γ lascia stabile $(Fu)^\perp$. Essendo Fu il radicale di $(Fu)^\perp$ e γ simplettica, γ lascia stabile ogni Fu, $u \neq 0$. Come abbiamo mostrato a p. 385, ciò implica che γ sia una moltiplicazione scalare. Per $a \in F$ la trasformazione $a1$ è simplettica se e solo se $a = \pm 1$; segue $\gamma = \pm 1$. □

Studiamo ora l'azione di $\mathrm{Sp}_n(F)$ sullo spazio proiettivo $P_{n-1}(F)$ dei sottospazi unidimensionali di V. Ci occorre il risultato seguente.

Lemma 3. *Il gruppo* $\mathrm{Sp}_n(F)$ *agisce primitivamente sullo spazio proiettivo* $P_{n-1}(F)$.

Dimostrazione. Sia S una parte di una partizione di $P_{n-1}(F)$ stabile per $\mathrm{Sp}_n(F)$, con $|S| > 1$. Supponiamo dapprima che S contenga punti Fx, Fy tali che $B(x,y) \neq 0$; possiamo supporre $B(x,y) = 1$. Sia Fz un punto qualsiasi di $P_{n-1}(F)$. Se $B(x,z) \neq 0$, possiamo supporre anche $B(x,z) = 1$. Per l'analogo del teorema di prolungamento di Witt esiste $\eta \in \mathrm{Sp}_n(F)$ tale che $x \mapsto x$, $y \mapsto z$. Allora $\eta S = S$ perché $Fx \in S$, e $Fz \in S$ poiché $Fy \in S$ e $\eta y = z$. Supponiamo ora $B(x,z) = 0$ e $Fz \neq Fx$. Esiste $w \in V$ con $B(x,w) = 1 = B(z,w)$. Per quanto appena dimostrato $Fw \in S$. Esiste inoltre $\zeta \in \mathrm{Sp}_n(F)$ tale che $w \mapsto w$, $x \mapsto z$. Poiché $Fw \in S$, si ha $\zeta S = S$, e quindi $Fz \in S$ perché $Fx \in S$. Dunque $S = P_{n-1}(F)$ quando S contiene una coppia Fx, Fy con $B(x,y) \neq 0$.

Siano ora Fx, Fy due punti distinti di S tali che $B(x,y) = 0$. Esiste $u \in V$ con $B(x,u) = 1$, $B(y,u) = 0$. Sia $U = (Fx + Fu)^\perp$ e sia G il sottogruppo di $\mathrm{Sp}_n(F)$ delle trasformazioni uguali all'identità su $Fx + Fu$. Esse mandano lo spazio non degenere U in sé, e le restrizioni $\eta|_U$, $\eta \in G$, costituiscono il gruppo simplettico di U. Sia z un vettore non nullo di U. Poiché $y \in U$, il teorema di prolungamento per il gruppo simplettico di U dà $\eta \in G$ con $\eta y = z$. Si ha $\eta S = S$, poiché $Fx \in S$, e quindi $Fz \in S$ perché $Fy \in S$. Perciò S contiene ogni Fz con $z \neq 0$ in U. Poiché U contiene una coppia iperbolica, ci siamo ricondotti al primo caso. □

Prima di proseguire, esaminiamo il caso bidimensionale. Si ha una base simplettica (u,v) e, perché una trasformazione lineare η appartenga al gruppo simplettico, basta la condizione $B(\eta u, \eta v) = B(u,v) = 1$. Scrivendo $\eta u = au + bv$, $\eta v = cu + dv$, si ha $B(\eta u, \eta v) = ad - bc$. La condizione equivale pertanto a $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(F)$. Ne segue $\mathrm{Sp}_2(F) = \mathrm{SL}_2(F)$, e sappiamo che quest'ultimo gruppo modulo il centro è semplice salvo che $|F| = 2$ o 3. Quindi $\mathrm{PSp}_2(F) = \mathrm{Sp}_2(F)/\{1,-1\}$ è semplice, eccetto quando $|F| = 2$ o 3. Nel lemma seguente possiamo ormai supporre $n \geq 4$.

Lemma 4. *Il gruppo* $\mathrm{Sp}_n(F)$ *coincide con il proprio sottogruppo dei commutatori in tutti i casi, salvo* $n = 2$, $|F| = 2$ *o* 3, *e* $n = 4$, $|F| = 2$.

Dimostrazione. Supponiamo dapprima $|F| > 3$. Mostreremo che ogni trasvezione $\tau_{z,c} \neq 1$ è un commutatore. Esiste $d \in F$, $d \neq 0$, tale che $d^2 \neq 1$. Poniamo $b = (1-d^2)^{-1}c$, $a = -d^2 b$. Allora $a + b = c$, quindi $\tau_{z,c} = \tau_{z,a}\tau_{z,b}$. Sia η una trasformazione simplettica tale che $\eta z = dz$. Si ha

$$\eta \tau_{z,b}^{-1} \eta^{-1} = \eta \tau_{z,-b} \eta^{-1} = \tau_{\eta z,-b} = \tau_{dz,-b} = \tau_{z,-bd^2} = \tau_{z,a}.$$

Dunque $\tau_{z,c} = \tau_{z,a}\tau_{z,b} = \eta \tau_{z,b}^{-1} \eta^{-1} \tau_{z,b}$.

Consideriamo ora i due casi $|F| = 2$ e $|F| = 3$. In entrambi basta esibire una trasvezione diversa dall'identità appartenente al gruppo dei commutatori. Infatti, se $\tau_{z,c}$ è una tale trasvezione, il sottogruppo $H_z = \{\tau_{z,c} \mid c \in F\}$, ciclico di ordine due o tre, è contenuto nel gruppo dei commutatori. Lo stesso vale dunque per $H_{\eta z} = \eta H_z \eta^{-1}$. Vi appartiene quindi ogni H_x e, poiché gli H_x generano $\mathrm{Sp}_n(F)$, si ha $\mathrm{Sp}_n(F) = (\mathrm{Sp}_n(F), \mathrm{Sp}_n(F))$. In entrambi i casi partiamo da una base simplettica $(u_1, v_1, \ldots, u_r, v_r)$ e introduciamo trasformazioni lineari η la cui natura simplettica è chiara perché $(\eta u_1, \eta v_1, \ldots, \eta u_r, \eta v_r)$ è ancora una base simplettica. Gli esercizi che seguono chiariranno la ragione delle scelte. Trattiamo separatamente i due casi.

I. $|F| = 3$, $n \geq 4$. Siano σ ed η le trasformazioni lineari definite da

$$\begin{array}{llll} \eta u_1 = u_1 + u_2, & \eta v_1 = v_2, & \eta u_2 = u_1, & \eta v_2 = v_1 - v_2, \\ \sigma u_1 = u_1 - v_1 + v_2, & \sigma v_1 = v_1, & \sigma u_2 = u_2 + v_1, & \sigma v_2 = v_2, \\ \eta u_i = u_i, & \eta v_i = v_i, & \sigma u_i = u_i, & \sigma v_i = v_i \quad (i > 2). \end{array}$$

Sono simplettiche e si ha

$$\begin{array}{llll} \eta^{-1}u_1 = u_2, & \eta^{-1}v_1 = v_1 + v_2, & \eta^{-1}u_2 = u_1 - u_2, & \eta^{-1}v_2 = v_1, \\ \sigma^{-1}u_1 = u_1 + v_1 - v_2, & \sigma^{-1}v_1 = v_1, & \sigma^{-1}u_2 = u_2 - v_1, & \sigma^{-1}v_2 = v_2, \\ \eta^{-1}u_i = u_i, & \eta^{-1}v_i = v_i, & \sigma^{-1}u_i = u_i, & \sigma^{-1}v_i = v_i \quad (i > 2). \end{array}$$

Inoltre, posto $C = \eta\sigma\eta^{-1}\sigma^{-1}$, si ha $Cu_1 = u_1 + v_1$, $Cu_i = u_i$ per $i > 1$, e $Cv_j = v_j$ per $1 \leq j \leq r$. Dunque $\eta\sigma\eta^{-1}\sigma^{-1} = \tau_{v_1,1}$.

II. $|F| = 2$, $n \geq 6$. In questo caso definiamo η e σ mediante

$$\begin{array}{llll} \eta u_1 = u_1 + u_3, & \eta v_1 = v_3, & \eta u_2 = u_1, & \eta v_2 = v_1 + v_3, \\ \eta u_3 = u_2, & \eta v_3 = v_2, & \eta u_i = u_i, & \eta v_i = v_i \quad (i > 3), \\ \sigma u_1 = u_1 + v_2, & \sigma v_1 = v_1, & \sigma u_2 = u_2 + v_1 + v_2 + v_3, & \sigma v_2 = v_2, \\ \sigma u_3 = u_3 + v_2 + v_3, & \sigma v_3 = v_3, & \sigma u_i = u_i, & \sigma v_i = v_i \quad (i > 3). \end{array}$$

Sono simplettiche e soddisfano

$$\begin{array}{llll} \eta^{-1}u_1 = u_2, & \eta^{-1}v_1 = v_1 + v_2, & \eta^{-1}u_2 = u_3, & \eta^{-1}v_2 = v_3, \\ \eta^{-1}u_3 = u_1 - u_2, & \eta^{-1}v_3 = v_1, & \eta^{-1}u_i = u_i, & \eta^{-1}v_i = v_i \quad (i > 3), \end{array} \qquad \sigma^{-1} = \sigma.$$

Otteniamo di nuovo, per $C = \eta\sigma\eta^{-1}\sigma^{-1}$, $Cu_1 = u_1 + v_1$, $Cu_i = u_i$ per $i > 1$, e $Cv_j = v_j$ per $1 \leq j \leq r$. Dunque $\eta\sigma\eta^{-1}\sigma^{-1} = \tau_{v_1,1}$. Questo conclude la dimostrazione del lemma. □

Abbiamo così dimostrato che, in tutti i casi in esame, $\mathrm{Sp}_n(F)$ coincide con il proprio gruppo dei commutatori, che agisce primitivamente su $P_{n-1}(F)$ con nucleo $\{1, -1\}$ e che il sottogruppo H_x, $x \neq 0$, è normale abeliano in $\mathrm{Stab}(Fx)$. Come abbiamo visto nella dimostrazione precedente, i coniugati di H_x generano $\mathrm{Sp}_n(F)$. Dal criterio di semplicità segue allora il teorema seguente.

Teorema 6.16. *Il gruppo* $\mathrm{PSp}_n(F) = \mathrm{Sp}_n(F)/\{1, -1\}$ *è semplice salvo i casi* $n = 2$, $|F| = 2$ *o* 3, *e* $n = 4$, $|F| = 2$.

Esercizi

1. Sia (u_i, v_i) una base simplettica di V e siano U, U' i sottospazi generati rispettivamente dagli u_i e dai v_i. Sia K l'insieme degli $\eta \in \mathrm{Sp}_n(F)$ che lasciano stabili U e U'. Mostrare che una trasformazione lineare η appartiene a K se e solo se la sua matrice rispetto alla base

   $$(u_1, \ldots, u_r, v_1, \ldots, v_r) \tag{53}$$

   ha la forma

   $$\begin{pmatrix} A & 0 \\ 0 & ({}^tA)^{-1} \end{pmatrix}, \qquad A \in \mathrm{GL}_r(F). \tag{54}$$

   Osservare che K è un sottogruppo di $\mathrm{Sp}_n(F)$.

2. Con le notazioni dell'esercizio 1, sia L il sottogruppo di $\mathrm{Sp}_n(F)$ degli σ che fissano ogni $v \in U'$. Mostrare che $\sigma \in L$ se e solo se la sua matrice rispetto a $(u_1, \ldots, u_r, v_1, \ldots, v_r)$ ha la forma

   $$\begin{pmatrix} 1 & S \\ 0 & 1 \end{pmatrix}, \tag{55}$$

   con ${}^tS = S$. Mostrare che $\sigma \mapsto S$ è un monomorfismo di L nel gruppo additivo delle matrici simmetriche di ordine r. Mostrare che per $S = e_{ii}$, $1 \leq i \leq r$, la corrispondente σ è una trasvezione.

3. Siano $\sigma \in L$ ed $\eta \in K$, come negli esercizi 1 e 2. Verificare che $\eta\sigma\eta^{-1} \in L$. Verificare che, se le matrici di η e σ sono rispettivamente (54) e (55), la matrice del commutatore $\eta\sigma\eta^{-1}\sigma^{-1}$ è

$$\begin{pmatrix} 1 & S_1 \\ 0 & 1 \end{pmatrix}, \tag{56}$$

dove

$$S_1 = AS({}^tA) - S. \tag{57}$$

4. Applicare gli esercizi 1–3 alla verifica delle affermazioni dell'ultima parte della dimostrazione del Lemma 4 di p. 394.

5. Dimostrare che per $A \in M_n(F)$ la trasformazione lineare di $M_n(F)$ data da $X \mapsto AX({}^tA) - X$ è invertibile se e solo se nessuna coppia di autovalori di A è costituita da elementi inversi. (Osservare che la trasformazione lascia stabili lo spazio delle matrici simmetriche e quello delle matrici antisimmetriche.)

6. Dare un esempio di una trasformazione simplettica senza punti fissi diversi da zero.

7. Sia (u_i, v_i) una base simplettica ordinata come nella (53) e sia

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad A_{ij} \in M_r(F).$$

Mostrare che A è la matrice di una trasformazione simplettica se e solo se i blocchi A_{ij} soddisfano

$$\begin{aligned} {}^tA_{11}A_{22} - {}^tA_{21}A_{12} &= 1_r = {}^tA_{22}A_{11} - {}^tA_{12}A_{21}, \\ {}^tA_{11}A_{21} - {}^tA_{21}A_{11} &= 0_r = {}^tA_{22}A_{12} - {}^tA_{12}A_{22}. \end{aligned}$$

8. Dimostrare che il polinomio caratteristico di una trasformazione simplettica è della forma $g(x)x^r g(x^{-1})$, dove $g(x)$ è un polinomio di grado n $(= 2r)$.[11]

9. Mostrare che ogni elemento di $\mathrm{Sp}_n(F)$ è prodotto di al più $2n$ trasvezioni simplettiche.

6.10 Ordini dei gruppi ortogonali e simplettici su un campo finito

Contiamo dapprima gli elementi dei gruppi ortogonali su un campo finito F. Useremo alcune formule per il numero dei punti di una quadrica in uno spazio vettoriale su F. Nel § 6.3 abbiamo visto che esistono due classi di equivalenza di forme quadratiche non degeneri e che il discriminante permette di distinguerle. Sia d un elemento di F^* che non sia un quadrato. Per $n = 2r$ pari possiamo scegliere come rappresentanti le forme quadratiche associate alle matrici

$$\mathrm{diag}\{1, -1, 1, -1, \ldots, 1, -1\}, \tag{58}$$

$$\mathrm{diag}\{1, -1, \ldots, 1, -1, 1, -d\}. \tag{59}$$

Per $n = 2r + 1$ dispari $(r \geq 0)$ possiamo prendere come rappresentanti

$$\mathrm{diag}\{1, -1, \ldots, 1, -1, -1\}, \tag{60}$$

[11] Così nel testo a stampa. Il grado indicato è incompatibile con la formula. La proprietà generale sul campo di base è la reciprocità del polinomio caratteristico: $p(x) = x^{2r}p(x^{-1})$.

$$\text{diag}\{1,-1,\dots,1,-1,-d\}. \tag{61}$$

Nei casi (58), (59), (60) e (61) i discriminanti sono rispettivamente $(-1)^r$, $(-1)^r d$, $(-1)^{r+1}$ e $(-1)^{r+1}d$, a meno di quadrati. Poiché ci interessano i gruppi ortogonali associati e $\mathrm{O}(V,Q) = \mathrm{O}(V,dQ)$, possiamo moltiplicare per d l'ultima forma quadratica. Il discriminante viene così moltiplicato per d^{2r+1} e si ottiene una forma equivalente a quella associata alla (60). Basta dunque considerare i tre casi (58), (59) e (60). Il nostro conteggio si baserà su alcune formule per il numero dei punti di una quadrica in V su F.

Lemma. *Sia $|F| = q$. Il numero delle soluzioni di*

$$x_1^2 - y_1^2 + \cdots + x_r^2 - y_r^2 = b \tag{62}$$

è

$$N(2r,b) = \begin{cases} q^{2r-1} + q^r - q^{r-1} & \text{se } b = 0, \\ q^{2r-1} - q^{r-1} & \text{se } b \neq 0. \end{cases} \tag{63}$$

Il numero delle soluzioni di

$$x_1^2 - y_1^2 + \cdots + x_{r-1}^2 - y_{r-1}^2 + x_r^2 - dy_r^2 = b \tag{64}$$

è

$$N(2r,d,b) = \begin{cases} q^{2r-1} - q^r + q^{r-1} & \text{se } b = 0, \\ q^{2r-1} + q^{r-1} & \text{se } b \neq 0. \end{cases} \tag{65}$$

Il numero delle soluzioni di

$$x_1^2 - y_1^2 + \cdots + x_r^2 - y_r^2 - x_{r+1}^2 = b \qquad (r \geq 0) \tag{66}$$

è

$$N(2r+1,b) = \begin{cases} q^{2r} & \text{se } b = 0, \\ q^{2r} - q^r & \text{se } -b \text{ non è un quadrato}, \\ q^{2r} + q^r & \text{se } -b \text{ è un quadrato non nullo}. \end{cases} \tag{67}$$

Dimostrazione. Consideriamo dapprima i due casi per $n = 2r = 2$. L'equazione $x_1^2 - y_1^2 = b$ equivale a $uv = b$, con $u = x_1 - y_1$ e $v = x_1 + y_1$; essa ha $2q-1$ soluzioni per $b = 0$ e $q-1$ soluzioni per $b \neq 0$. Ciò concorda con la (63) per $r = 1$. Dimostriamo la (63) per induzione su r. Scriviamo $b = a + c$, con q scelte per a. La (62) equivale allora alle due equazioni

$$x_1^2 - y_1^2 + \cdots + x_{r-1}^2 - y_{r-1}^2 = a, \qquad x_r^2 - y_r^2 = c.$$

Se $b = 0$, il caso $a = 0 = c$ contribuisce $(2q-1)(q^{2r-3} + q^{r-1} - q^{r-2})$ soluzioni, mentre ciascun caso $a \neq 0$, $c = -a$ ne contribuisce $(q-1)(q^{2r-3} - q^{r-2})$. In totale si ottengono

$$(2q-1)(q^{2r-3} + q^{r-1} - q^{r-2}) + (q-1)^2(q^{2r-3} - q^{r-2})$$

soluzioni. Semplificando si trova il valore di $N(2r,0)$ dato dalla (63).[12] Analogamente si ottengono la seconda parte della (63) e le due parti della (65). Per la (66) scriviamo $x_1^2 - y_1^2 + \cdots + x_r^2 - y_r^2 = b + x_{r+1}^2$. Se $b = 0$, la scelta $x_{r+1} = 0$ dà $q^{2r-1} + q^r - q^{r-1}$ soluzioni, e le $q-1$ scelte $x_{r+1} \neq 0$

[12] Nel secondo addendo il testo a stampa, p. 399, reca q^{r-1} in luogo di q^{r-2}; l'esponente qui riportato segue dalla (63) applicata a $r-1$.

ne danno $(q-1)(q^{2r-1}-q^{r-1})$. Complessivamente le soluzioni sono q^{2r}, come nella (67). Allo stesso modo si ottengono gli altri casi della (67). □

Possiamo ora stabilire le formule per gli ordini dei gruppi ortogonali. Per un campo di q elementi indichiamo con $O_n(q)$ i gruppi associati alle matrici (58) e (60), con n pari nel primo caso e dispari nel secondo. Se d non è un quadrato in F, il gruppo ortogonale associato alla (59) si indica con $O_n(q,d)$. I rispettivi gruppi delle rotazioni si indicano con $O_n^+(q)$ e $O_n^+(q,d)$. Si ha il seguente risultato.

Teorema 6.17. *Gli ordini di $O_n(q)$ e $O_n(q,d)$ sono dati dalle formule*

$$|O_n(q)| = 2(q^{2r-1}-q^{r-1})(q^{2r-2}-1)q^{2r-3}\cdots(q^2-1)q, \quad n=2r; \tag{68}$$

$$|O_n(q,d)| = 2(q^{2r-1}+q^{r-1})(q^{2r-2}-1)q^{2r-3}\cdots(q^2-1)q, \quad n=2r; \tag{69}$$

$$|O_n(q)| = 2(q^{n-1}-1)q^{n-2}(q^{n-3}-1)q^{n-4}\cdots(q^2-1)q, \quad n=2r+1. \tag{70}$$

Dimostrazione. Per $n=1$ i gruppi ortogonali consistono di 1 e -1, dunque hanno ordine 2. Procediamo per induzione su n e supponiamo $n \geq 2$. Scegliamo $x \in V$ con $Q(x)=1$ e consideriamo l'orbita Gx, dove G è il gruppo ortogonale in esame. Per il teorema di Witt, Gx è l'insieme dei vettori y per i quali $Q(y)=1$. Per calcolare $|G|$ usiamo la formula

$$|G| = |Gx|\,|\operatorname{Stab} x|$$

((40), p. 76). Supponiamo anzitutto $G=O_n(q)$, $n=2r$. Allora $\operatorname{Stab} x$ è isomorfo al gruppo ortogonale di $(Fx)^\perp$ rispetto alla restrizione di Q. Questo sottogruppo è evidentemente isomorfo a $O_{n-1}(q)$. Inoltre il numero dei y con $Q(y)=1$ è il numero delle soluzioni della (62) con $b=1$, cioè $q^{2r-1}-q^{r-1}$ per la (63). Per induzione possiamo supporre la (70) valida per $n-1=2r-1$. Moltiplicando per $q^{2r-1}-q^{r-1}$ si ottiene la (68). Allo stesso modo si ricava la (69) moltiplicando per $q^{2r-1}+q^{r-1}$, che per la (65) è il numero dei y con $Q(y)=1$. Il caso rimanente $n=2r+1$ si tratta analogamente, scendendo al caso $O_{n-1}(q)$, $n-1=2r$. □

Gli ordini dei rispettivi gruppi delle rotazioni si ottengono sopprimendo il fattore 2 nelle formule. Si può mostrare che $O_n^+/(O_n,O_n)$ ha ordine due. Riguardo all'appartenenza di -1 a $\Omega_n(q)=(O_n(q),O_n(q))$ o a $\Omega_n(q,d)=(O_n(q,d),O_n(q,d))$ per n pari, nel testo si afferma che -1 appartiene al primo e non al secondo.[13] (Questi risultati sono stabiliti nel volume II.) Se ne ricavano le formule per gli ordini dei gruppi $P\Omega_n(q)$ e $P\Omega_n(q,d)$, la cui semplicità abbiamo dimostrato nel § 6.8.

Consideriamo ora i gruppi simplettici $\mathrm{Sp}_n(q)=\mathrm{Sp}_n(F)$, dove F è un campo di q elementi. Qui ammettiamo anche la caratteristica due, nel qual caso q è una potenza di due. Quante sono le coppie iperboliche di vettori (x,y) in V? Come primo vettore si può scegliere un qualsiasi vettore non nullo di V, con q^n-1 possibilità. Inoltre, data una coppia iperbolica (x,y), ogni altra coppia iperbolica con primo vettore x è della forma (x,y') con $y'=y+z$, $z\in(Fx)^\perp$. Questo spazio ha dimensione $n-1$ e contiene q^{n-1} vettori, perciò vi sono q^{n-1} scelte per z. Abbiamo dunque $(q^n-1)q^{n-1}$ coppie iperboliche (x,y). Per l'analogo del teorema di Witt, una trasformazione simplettica può portare una qualsiasi di esse in un'altra. Inoltre lo stabilizzatore di (x,y), ossia l'insieme degli $\sigma\in\mathrm{Sp}_n(q)$ tali che $(\sigma x,\sigma y)=(x,y)$, è un sottogruppo isomorfo a $\mathrm{Sp}_{n-2}(q)$. Come nella dimostrazione del teorema precedente si ha quindi

$$|\mathrm{Sp}_n(q)| = q^{n-1}(q^n-1)|\mathrm{Sp}_{n-2}(q)|.$$

[13] L'affermazione a p. 400 richiede che $(-1)^r$ sia un quadrato in F, con $n=2r$. Infatti, per il criterio di p. 390 e i discriminanti delle (58) e (59), se $(-1)^r$ non è un quadrato le due appartenenze si scambiano.

Ne segue subito il teorema seguente.

Teorema 6.18. *L'ordine di* $\mathrm{Sp}_n(q)$ *per un campo* F *di* q *elementi è*

$$|\mathrm{Sp}_n(q)| = q^{n-1}(q^n-1)q^{n-3}(q^{n-2}-1)\cdots q(q^2-1). \tag{71}$$

Gli ordini dei corrispondenti gruppi proiettivi $\mathrm{PSp}_n(q)$ sono ancora dati dalla (71) per $q = 2^t$ e sono $\frac{1}{2}|\mathrm{Sp}_n(q)|$ per q dispari. Per avere gruppi semplici occorre $n \geq 6$ se $q = 2$ e $n \geq 4$ se $q = 3$. Gli ordini per $q = 2$, $n = 6$ e per $q = 3$, $n = 4$ sono rispettivamente $2^9 \cdot 3^4 \cdot 5 \cdot 7$ e $2^6 \cdot 3^4 \cdot 5$.[14]

Esercizi

1. Enunciare il lemma come risultato sul numero dei vettori x tali che $Q(x) = 0$ oppure $Q(x) = 1$, dove Q è una qualsiasi forma quadratica non degenere su uno spazio vettoriale di dimensione n su un campo di q elementi.
2. Determinare il numero delle coppie iperboliche in V nelle ipotesi dell'esercizio 1.
3. Sia p la caratteristica di F, cosicché q è una potenza di p. Determinare un p-sottogruppo di Sylow per i gruppi ortogonali del Teorema 6.17 e per $\mathrm{Sp}_n(q)$ del Teorema 6.18.
4. Considerare le famiglie di gruppi semplici seguenti: (a) i gruppi alterni ($n \geq 5$), (b) i gruppi unimodulari proiettivi, (c) i gruppi $P\Omega_n(q)$ e $P\Omega_n(q,d)$, (d) i gruppi $\mathrm{PSp}_n(q)$. Determinare tutti i gruppi di queste famiglie di ordine inferiore a un milione.

6.11 Postilla sulle forme hermitiane e sulla geometria unitaria

Sia K un'estensione quadratica separabile del campo F. Esiste allora un automorfismo $a \mapsto \bar{a}$ di K su F che ha F come campo fisso e ha ordine due ($\overline{\bar{a}} = a$). Un esempio da tenere presente è $F = \mathbb{R}$, $K = \mathbb{C}$. Una matrice $h = (h_{ij})$ di ordine n a elementi in K si dice *hermitiana* se $h_{ij} = \overline{h_{ji}}$, ossia, in notazione matriciale, ${}^t\bar{h} = h$, dove $\bar{h} = (\overline{h_{ij}})$. Sia ora V uno spazio vettoriale di dimensione n su K con base $(e_1, e_2, \ldots, e_n)$. Per $x = \sum_i a_i e_i$, $y = \sum_i b_i e_i$, definiamo

$$H(x,y) = \sum_{i,j} h_{ij} a_i \overline{b_j}. \tag{72}$$

Quali proprietà ha l'applicazione $H\colon (x,y) \mapsto H(x,y)$? Anzitutto è biadditiva:

$$\begin{aligned} H(x+x',y) &= H(x,y) + H(x',y), \\ H(x,y+y') &= H(x,y) + H(x,y'). \end{aligned} \tag{73}$$

Inoltre soddisfa

$$H(ax,y) = aH(x,y), \qquad H(x,ay) = \bar{a}H(x,y). \tag{74}$$

La prima uguaglianza della (73), insieme alla prima della (74), afferma che per y fissato $x \mapsto H(x,y)$ è una funzione lineare su V. Le seconde uguaglianze della (73) e della (74) affermano che per x fissato la funzione $y \mapsto H(x,y)$ è *antilineare*, o *lineare coniugata*, in y. Osserviamo inoltre che $H(y,x) = \sum_{i,j} h_{ij} b_i \bar{a}_j = \sum_{i,j} \overline{h_{ji}} b_i \bar{a}_j$ e $\overline{H(x,y)} = \sum_{i,j} \overline{h_{ij}} \bar{a}_i b_j$. Scambiando gli indici di somma otteniamo

$$H(y,x) = \overline{H(x,y)}. \tag{75}$$

[14] Per $\mathrm{PSp}_4(3)$ l'originale, p. 401, riporta $2^7 \cdot 3^5 \cdot 5$. Dalla (71) si ricava invece $3^3(3^4-1)\,3(3^2-1)/2 = 25920 = 2^6 \cdot 3^4 \cdot 5$.

Un'applicazione $H\colon V \times V \to K$ che soddisfi le (73)–(75) si chiama *forma hermitiana* sullo spazio vettoriale V/K. La costruzione data a partire dalle matrici hermitiane fornisce tutte le forme hermitiane. Infatti, data una forma hermitiana H, poniamo $h_{ij} = H(e_i, e_j)$. Si ha $\overline{h_{ji}} = \overline{H(e_j, e_i)} = H(e_i, e_j) = h_{ij}$, quindi $h = (h_{ij})$ è hermitiana. Inoltre dalle (73) e (74) segue

$$H\left(\sum_i a_i e_i, \sum_j b_j e_j\right) = \sum_{i,j} H(e_i, e_j) a_i \bar{b}_j = \sum_{i,j} h_{ij} a_i \bar{b}_j,$$

il che mostra che H è la forma associata alla matrice hermitiana h.

Con qualche cautela si possono sviluppare questi concetti anche per i corpi non necessariamente commutativi. Anzitutto, l'algebra lineare si generalizza agli spazi vettoriali (o moduli) su corpi. La teoria della dipendenza lineare e l'invarianza della dimensione si estendono senza fare uso di determinanti. Sia Δ un corpo munito di un'*involuzione*, ossia di un antiautomorfismo $a \mapsto \bar{a}$ tale che $\overline{\bar{a}} = a$ (si veda il § 2.8, p. 112). Un buon esempio da tenere presente è il corpo $\mathbb{H}$ dei quaternioni di Hamilton. Sia $h = (h_{ij})$ una matrice hermitiana di ordine n, $\overline{h_{ij}} = h_{ji}$, e sia V uno spazio vettoriale su Δ con base $(e_1, e_2, \dots, e_n)$. Per $x = \sum_i a_i e_i$, $y = \sum_i b_i e_i$ poniamo

$$H(x, y) = \sum_{i,j} a_i h_{ij} \overline{b_j}. \tag{72′}$$

Valgono allora le (73), le relazioni

$$H(ax, y) = aH(x, y), \qquad H(x, ay) = H(x, y)\bar{a}, \tag{74′}$$

e la (75). Queste proprietà definiscono una *forma hermitiana* su V/Δ. È chiaro che si ha una biiezione fra l'insieme delle matrici hermitiane e quello delle forme hermitiane.

Il caso di un corpo Δ comprende quello considerato inizialmente, in cui K è un'estensione quadratica separabile di F. Poiché non richiediamo che $a \mapsto \bar{a}$ sia diversa dall'identità, ammettiamo anche il caso $\Delta = F$, $\bar{a} = a$ per ogni a. La nozione di forma hermitiana si riduce allora a quella di forma bilineare simmetrica. Una parte considerevole della teoria delle forme bilineari simmetriche, compresa la teoria strutturale dei gruppi corrispondenti, detti *gruppi unitari*, si estende alle forme hermitiane. Indicheremo alcuni risultati elementari e rinviamo a *La Géométrie des Groupes Classiques* di Dieudonné per la teoria dei gruppi. Dimostriamo dapprima l'esistenza di una base ortogonale.

Teorema 6.19. *Sia H una forma hermitiana su uno spazio vettoriale di dimensione finita V su un corpo Δ. A meno che H sia alternante e Δ sia un campo di caratteristica due, V ha una base $(u_1, \dots, u_r, z_1, \dots, z_{n-r})$ tale che*

$$\begin{aligned} H(u_i, u_j) &= \delta_{ij} b_j, \qquad b_i = \bar{b}_i \neq 0, \\ H(z_k, z_l) &= H(u_i, z_k) = 0. \end{aligned} \tag{76}$$

Dimostrazione. La dimostrazione del teorema di diagonalizzazione si estende purché si mostri che, per $H \neq 0$, esiste u con $H(u, u) \neq 0$.[15] Supponiamo dunque $H(u, u) = 0$ per ogni u. Ne segue $H(u, v) + H(v, u) = 0$ per ogni u, v. Poiché $H(v, u) = \overline{H(u, v)}$, si ha $H(u, v) + \overline{H(u, v)} = 0$ per ogni u, v. Se $H \neq 0$ possiamo scegliere u, v tali che $H(u, v) = 1$. Per ogni $a \in \Delta$ si ha $H(au, v) + \overline{H(au, v)} = 0$, cioè $aH(u, v) + \overline{H(u, v)}\bar{a} = 0$. Pertanto $a + \bar{a} = 0$. Poiché $\bar{1} = 1$, la caratteristica è due e $\bar{a} = a$ per ogni $a \in \Delta$. L'identità è un antiautomorfismo solo se Δ è commutativo; dunque Δ è un campo. Essendo $\bar{a} = a$, H è bilineare, ed essendo $H(u, u) = 0$

[15] Il rinvio dell'originale al Teorema 6.3 è da intendersi al Teorema 6.5, sulla diagonalizzazione.

si tratta di una forma alternante su uno spazio vettoriale su un campo di caratteristica due. Questo caso è stato escluso, e quindi il teorema è dimostrato. □

Due importanti casi particolari sono (1) $\Delta = \mathbb{C}$, con la consueta coniugazione $a \mapsto \bar{a}$, e (2) $\Delta = \mathbb{H}$, ancora con la coniugazione consueta. In entrambi si possono scegliere i b_i uguali a ± 1. In generale, sostituendo u_i con $c_i u_i$, si sostituisce b_i con $c_i b_i \bar{c}_i$. Nei due casi considerati, da $\bar{b}_i = b_i$ segue che b_i è reale. Possiamo allora scegliere c_i in modo che $c_i \bar{c}_i = |b_i|^{-1}$, sostituendo così b_i con ± 1. La dimostrazione del teorema di Sylvester si estende e mostra che il numero dei $b_i > 0$ è indipendente dalla scelta della base ortogonale. La differenza fra il numero dei b_i positivi e quello dei b_i negativi si chiama in entrambi i casi *segnatura*.

Si potrebbe poi considerare il teorema di Witt, che resta valido per le forme hermitiane sotto opportune ipotesi. Per questo risultato rinviamo al volume II delle nostre *Lectures in Abstract Algebra*[16] oppure a *La Géométrie des Groupes Classiques* di Dieudonné.

Esercizi

1. Sia K un'estensione quadratica di un campo F di caratteristica diversa da due, sia $a \mapsto \bar{a}$ l'automorfismo di K su F diverso dall'identità e sia $H(x, y)$ una forma hermitiana sullo spazio vettoriale V/K di dimensione n. Si ponga $B(x, y) = H(x, y) + H(y, x)$. Mostrare che B è una forma bilineare simmetrica su V/F, di dimensione $2n$, e soddisfa

   $$B(ax, ay) = N(a)B(x, y), \qquad N(a) = a\bar{a}. \tag{77}$$

   Sia $K = F(i)$, $i^2 = b \in F$, e sia B una forma bilineare simmetrica su V/F che soddisfa la (77). Si definisca

   $$H(x, y) = B(x, y) - b^{-1} i B(x, iy). \tag{78}$$

   Mostrare che H è hermitiana. Verificare che le due applicazioni $H \mapsto B$ e $B \mapsto H$ così definite sono inverse.[17]

2. Con le notazioni dell'esercizio 1, mostrare che due forme hermitiane H_1, H_2 sono equivalenti (con la stessa definizione usata per le forme bilineari) se e solo se sono equivalenti le forme bilineari simmetriche associate B_1, B_2 definite nell'esercizio 1.

3. Sia $\mathbb{H}$ l'algebra dei quaternioni di Hamilton con base $(1, i, j, k)$ su $\mathbb{R}$, dove $i^2 = j^2 = k^2 = -1$, $ij = k = -ji$, $jk = i = -kj$, $ki = j = -ik$. Sia V uno spazio vettoriale di dimensione n su $\mathbb{H}$ e sia H una forma hermitiana su $V/\mathbb{H}$. Si ponga $B(x, y) = H(x, y) + H(y, x)$. Mostrare che B è simmetrica sullo spazio vettoriale reale $V/\mathbb{R}$ di dimensione $4n$ e soddisfa la (77). Viceversa, data una forma bilineare simmetrica B su $V/\mathbb{R}$ che soddisfi la (77), si definisca

   $$H(x, y) = B(x, y) - iB(x, iy) - jB(x, jy) - kB(x, ky). \tag{79}$$

   Mostrare che H è hermitiana su $V/\mathbb{H}$ e che le applicazioni $B \mapsto H$ e $H \mapsto B$ sono inverse.[18]

4. Con le notazioni dell'esercizio 3, mostrare che due forme hermitiane H_1, H_2 su $V/\mathbb{H}$ sono equivalenti se e solo se sono equivalenti le corrispondenti forme bilineari simmetriche B_1, B_2 su $V/\mathbb{R}$.

[16] Springer-Verlag, New York, Heidelberg, Berlino. Pubblicato per la prima volta nel 1953 dalla D. Van Nostrand Company.

[17] Con la definizione di B data all'inizio dell'esercizio, il secondo membro della (78) va diviso per due per ottenere effettivamente l'applicazione inversa.

[18] La (79) è trascritta come a p. 404. Per la convenzione di linearità nel primo argomento adottata nella (74′), la ricostruzione è $\frac{1}{2}\big(B(x, y) + iB(x, iy) + jB(x, jy) + kB(x, ky)\big)$.

Capitolo 7
Algebre su un campo

Il concetto di algebra associativa nasce combinando quelli di anello e di spazio vettoriale, con certe relazioni fra le due strutture. In questo capitolo daremo un'introduzione alle algebre associative e ad alcune classi di algebre non associative: le algebre di Lie, di Jordan e alternative. Tutte e tre sono strettamente legate, in un modo o nell'altro, alle algebre associative. Le prime due nascono modificando in modo semplice il prodotto di un'algebra associativa; le algebre alternative si ottengono invece generalizzando di poco la legge associativa. Per le applicazioni a vasti campi della matematica e della fisica, le classi importanti di algebre sono proprio queste quattro.

Le algebre associative si incontrano spesso in algebra. Un esempio di primo piano è l'algebra delle trasformazioni lineari di uno spazio vettoriale di dimensione finita. Essa ha un ruolo universale, analogo a quello del gruppo simmetrico in teoria dei gruppi: ogni algebra associativa di dimensione finita è isomorfa a un'algebra di trasformazioni lineari di uno spazio vettoriale di dimensione finita. È naturale considerare gli omomorfismi di algebre associative in algebre di trasformazioni lineari o, equivalentemente, in algebre di matrici. Di particolare interesse sono le rappresentazioni regolari. Da queste si ricavano le applicazioni traccia e norma di un'algebra associativa nel suo campo di base, che generalizzano le nozioni introdotte per i campi nel § 4.15.

L'origine delle algebre alternative risale alla scoperta, dovuta indipendentemente a J. T. Graves e ad A. Cayley, dell'algebra $\mathbb{O}$ degli ottonioni: un'algebra di dimensione otto su $\mathbb{R}$ che contiene l'algebra $\mathbb{H}$ dei quaternioni di Hamilton. Essa conserva molte proprietà di $\mathbb{H}$, per esempio è un'algebra con divisione. La moltiplicazione, però, non è associativa. Valgono invece le leggi $(xx)y = x(xy)$ e $y(xx) = (yx)x$, più deboli dell'associatività. Gli ottonioni permettono di introdurre coordinate in certe geometrie «eccezionali», più precisamente in certi piani proiettivi non desarguesiani.

Le algebre di Lie prendono il nome dal grande matematico norvegese della fine dell'Ottocento, Sophus Lie. Si ottengono dalle algebre associative sostituendo al prodotto associativo il prodotto di Lie, o commutatore additivo, $[xy] = xy - yx$. Si considerano allora le sottoalgebre rispetto alla composizione $[xy]$, che possono anche non essere sottoalgebre dell'algebra associativa di partenza. Le algebre di Lie sono gli oggetti fondamentali della teoria dei gruppi continui di Lie. La grande conquista di Lie fu di ricondurre lo studio delle proprietà locali dei gruppi continui allo studio delle algebre di Lie associate.

Le algebre di Jordan sono di origine relativamente recente. Le introdusse nel 1931 un fisico, P. Jordan, in vista di applicazioni alla meccanica quantistica. Esse nascono dalla ricerca di leggi semplici per il prodotto di Jordan (o «anticommutatore») $x \cdot y = \frac{1}{2}(xy + yx)$ in un'algebra associativa su un campo di caratteristica diversa da 2. Le algebre di Jordan trovano applicazioni nell'analisi, nella geometria e nello studio dei gruppi di Lie.

Per la teoria associativa studieremo le rappresentazioni matriciali regolari e le proprietà della traccia e della norma. Introdurremo inoltre l'algebra esterna $E(V)$ di uno spazio vettoriale e la applicheremo per dedurre in modo rapido e diretto alcune delle proprietà principali dei determinanti. Dimostreremo infine i teoremi di Frobenius e di Wedderburn sulle algebre associative con divisione. Le algebre alternative compariranno in relazione a un problema sulle forme quadratiche studiato per la prima volta da A. Hurwitz. Tratteremo assai brevemente le algebre di Lie e di Jordan, poco oltre le definizioni fondamentali. Nel volume II generalizzeremo il concetto di algebra associativa su un campo a quello di algebra associativa su un anello commutativo e studieremo la teoria della struttura degli anelli e delle algebre associative.

7.1 Definizione ed esempi di algebre associative

Sebbene non abbiamo ancora definito formalmente il concetto di algebra associativa, ne abbiamo già incontrato parecchi esempi. In teoria dei campi abbiamo studiato un campo E relativamente a un sottocampo F. Abbiamo allora considerato E come spazio vettoriale su F, dove il prodotto au, con $a \in F$, $u \in E$, è quello definito in E. Abbiamo pure il prodotto uv di due elementi qualsiasi di E, che, insieme all'addizione, allo zero e all'unità, dà la struttura di anello di E. Le due strutture di spazio vettoriale e di anello hanno lo stesso gruppo additivo ed è $a(uv) = (au)v = u(av)$ per $a \in F$ e $u, v \in E$.

Una situazione simile si presenta per l'anello dei polinomi $F[x]$ in un'indeterminata x sul campo F. Oltre alla struttura di anello vi è quella di spazio vettoriale su F, in cui $af(x)$, per $a \in F$ e $f(x) \in F[x]$, è il prodotto nell'anello. Anche qui l'addizione e lo zero coincidono e si ha $a(f(x)g(x)) = (af(x))g(x) = f(x)(ag(x))$.

Un altro esempio è l'insieme $M_n(F)$ delle matrici $n \times n$ a elementi in F. Oltre alla struttura di anello si ha uno spazio vettoriale su F: se $M = (m_{ij})$ e $a \in F$, allora $aM = (am_{ij})$. Questo prodotto coincide con il prodotto di M per la matrice $a1 = \mathrm{diag}\{a, a, \dots, a\}$. Poiché $a1$ appartiene al centro dell'anello, cioè commuta con ogni matrice, si ha $a(MN) = (a1)(MN) = ((a1)M)N = (aM)N$ e $a(MN) = (a1)(MN) = M(a1)N = M(aN)$. Nel primo esempio, quello di un campo su un sottocampo, lo spazio vettoriale può avere o meno dimensione finita. Nel secondo, $F[x]$, la dimensione è certamente infinita, poiché $1, x, x^2, \dots$ sono linearmente indipendenti su F. Invece $M_n(F)$, con la struttura appena definita, ha dimensione finita. Una base particolarmente utile è costituita dalle n^2 unità matriciali $\{e_{ij} \mid i, j = 1, \dots, n\}$, dove e_{ij} ha un 1 nella posizione (i, j) e zeri altrove (§ 2.3). La tavola di moltiplicazione è

$$e_{ij}e_{kl} = \delta_{jk}e_{il}, \tag{1}$$

e l'unità si esprime nella base come

$$1 = e_{11} + e_{22} + \cdots + e_{nn}. \tag{2}$$

Possiamo ora dare la definizione formale.

Definizione 1. *Un'algebra (associativa) su un campo F è una coppia costituita da un anello $(A, +, \cdot, 0, 1)$ e da uno spazio vettoriale A su F, con lo stesso insieme sottostante, la stessa addizione e lo stesso zero, tale che*

$$a(xy) = (ax)y = x(ay) \qquad (a \in F,\ x, y \in A). \tag{3}$$

Se A ha dimensione finita su F, diremo che l'algebra ha dimensione finita (o che ha una base finita). Indicheremo di solito l'algebra con la lettera, per esempio A, che designa l'insieme sottostante.

Gli esempi precedenti sono evidentemente algebre. Eccone un altro importante. Sia $G = \{s_1 = 1, s_2, \dots, s_n\}$ un gruppo finito e sia F un campo. Con $F[G]$ indichiamo lo spazio vettoriale su F di base G. I suoi elementi sono dunque le somme $\sum_1^n \alpha_i s_i$, con $\alpha_i \in F$. Per s_i distinti, una tale somma è nulla se e solo se tutti gli α_i sono nulli. L'addizione e la moltiplicazione per scalari sono quelle ovvie. Definiamo il prodotto in $F[G]$ ponendo

$$\left(\sum_i \alpha_i s_i\right)\left(\sum_j \beta_j s_j\right) = \sum_{i,j} (\alpha_i \beta_j)(s_i s_j),$$

dove $s_i s_j$ è il prodotto nel gruppo G (esercizio 8, p. 127 dell'originale). Dall'associatività in G segue subito l'associatività di questo prodotto. Le leggi distributive in F danno le stesse leggi

in $F[G]$ e $s_1 = 1$ è l'unità moltiplicativa. È chiara infine la (3), che lega il prodotto in $F[G]$ alla moltiplicazione per elementi di F. Abbiamo dunque un'algebra, detta *algebra di gruppo* di G su F.

Sia ora A una qualsiasi algebra sul campo F. Per $a \in F$ consideriamo l'elemento $a1 \in A$. Per la (3), $(a1)x = a(1x) = ax$ e $x(a1) = a(x1) = ax$. Il prodotto per lo scalare a coincide quindi con il prodotto nell'anello per $a1$, e $a1$ appartiene al centro di A. Si ha così un'applicazione $a \mapsto a1$ da F in A. Poiché $1 \mapsto 1$, $(a+b)1 = a1 + b1$ e $(ab)1 = (ab)1^2 = (a1)(b1)$, è un omomorfismo di anelli; infatti $(ax)(by) = (b(ax))y = ((ba)x)y = (ba)(xy) = (ab)(xy)$. Se $A \neq 0$, si ha $1 \neq 0$ in A e l'omomorfismo è iniettivo, poiché F è un campo. Viceversa, se il centro di un anello R contiene un sottoanello F che è un campo, possiamo considerare R come algebra su F prendendo per ax, con $a \in F$ e $x \in R$, il prodotto nell'anello. La (3) è allora immediata. Tutti gli esempi iniziali si ottengono così. Un'algebra su F è dunque, in sostanza, un anello con un sottocampo distinto nel suo centro.[1] La definizione un po' più astratta che abbiamo adottato presenta tuttavia alcuni vantaggi: per esempio, rende più naturale la nozione di omomorfismo che stiamo per introdurre.

Il lettore ha probabilmente esperienza sufficiente per formulare da sé i concetti fondamentali connessi a quello di algebra.[2] Eccone l'elenco: sottoalgebre, ideali, algebre quoziente, omomorfismi e nucleo di un omomorfismo. Passiamoli in rassegna, con qualche breve osservazione.

1. *Sottoalgebre.* Un sottoinsieme B di un'algebra A è una sottoalgebra se è un sottoanello di A e un sottospazio dello spazio vettoriale A. L'intersezione di sottoalgebre è una sottoalgebra. Se $S \subseteq A$, la sottoalgebra $F[S]$ generata da S è l'intersezione di tutte le sottoalgebre contenenti S. Si vede subito che $F[S]$ consiste delle combinazioni F-lineari di 1 e dei monomi $s_{i_1} \cdots s_{i_r}$, con $s_{i_j} \in S$: gli elementi hanno dunque la forma $\alpha_0 1 + \sum \alpha_{i_1 \ldots i_r} s_{i_1} \cdots s_{i_r}$, a coefficienti in F.

2. *Ideali.* Un sottoinsieme I di A è un ideale dell'algebra A se è un ideale di A come anello e un sottospazio di A come spazio vettoriale su F.

3. *Algebre quoziente.* Se I è un ideale dell'algebra A, abbiamo l'anello quoziente A/I e lo spazio vettoriale A/I. Insieme costituiscono l'*algebra quoziente* (o *differenza*) di A rispetto a I.

4. *Omomorfismo.* Un'applicazione di un'algebra A in un'algebra B sullo stesso campo è un omomorfismo di algebre se è insieme un omomorfismo di anelli e un'applicazione lineare. Monomorfismi, epimorfismi, endomorfismi e automorfismi di algebre si definiscono nel modo consueto. Se I è un ideale, si ha l'epimorfismo canonico $\nu : a \mapsto a + I$ di A su A/I. Se S genera A e $\eta_1, \eta_2 : A \to B$ sono omomorfismi di algebre tali che $\eta_1(s) = \eta_2(s)$ per ogni $s \in S$, allora $\eta_1 = \eta_2$.

5. *Nucleo di un omomorfismo.* Se $\eta : A \to B$ è un omomorfismo di algebre, allora $K = \eta^{-1}(0)$, costituito dagli elementi k con $\eta(k) = 0$, è un ideale di A, detto *nucleo* di η. Se I è un ideale contenuto in K, si ha l'omomorfismo indotto $\bar{\eta} : A/I \to B$ tale che $\bar{\eta}(a + I) = \eta(a)$, cioè $\eta = \bar{\eta}\nu$, dove ν è l'omomorfismo canonico di A su A/I.

[1] Un anello R può non avere alcun sottocampo di questo tipo. Per esempio $\mathbb{Z}/(6)$ non è un'algebra su alcun campo.

[2] Suggeriamo al lettore di interrompere qui la lettura e di formulare per proprio conto le nozioni che ritiene fondamentali, per confrontarle poi con quelle del testo.

Esercizi

1. Per $S \subseteq A$ sia $C_A(S)$ l'insieme dei $c \in A$ tali che $cs = sc$ per ogni $s \in S$. Verificare che $C_A(S)$ è una sottoalgebra. In particolare, anche il centro $C = C_A(A)$ è una sottoalgebra.

2. Sia $\mathbb{H}$ l'anello con divisione dei quaternioni reali del § 2.4, p. 98 dell'originale. È un'algebra su $\mathbb{R}$ di base $(1, i, j, k)$, con $i^2 = j^2 = k^2 = -1$, $ij = k = -ji$, $jk = i = -kj$, $ki = j = -ik$. Il sottoinsieme $G = \{\pm 1, \pm i, \pm j, \pm k\}$ è un sottogruppo del gruppo moltiplicativo, detto *gruppo dei quaternioni*. Scrivere $1' = -1$, $i' = -i$, $j' = -j$, $k' = -k$ in G e considerare l'algebra di gruppo $\mathbb{R}[G]$ di $G = \{1, 1', i, i', j, j', k, k'\}$. Dimostrare che esiste un omomorfismo di $\mathbb{R}[G]$ in $\mathbb{H}$ che manda $1'$ in -1, i in i, i' in $-i$, j in j, j' in $-j$, k in k e k' in $-k$. Determinarne il nucleo.

3. Sia $A = F[a]$ un'algebra generata da un elemento a. Dimostrare che $A \cong F[x]/(f(x))$, dove $f(x) = 0$ oppure $f(x)$ è monico. Nel primo caso $F[a] \cong F[x]$ ha dimensione infinita. Se f è monico di grado n, dimostrare che $(1, a, \dots, a^{n-1})$ è una base di $F[a]$. In tal caso a è detto *algebrico* e $f(x)$ il suo *polinomio minimo*.

4. Sia $A = F[a]$ come nell'esercizio 3, con a algebrico di polinomio minimo $f(x)$. Sia $f = q_1 \cdots q_r$ la fattorizzazione in fattori $q_i = p_i^{k_i}$, con p_i monici e primi, distinti fra loro. Posto $s_i = \prod_{j \neq i} q_j$, mostrare che esistono polinomi t_i con $\sum_{i=1}^r s_i(x)t_i(x) = 1$. Porre $e_i = s_i(a)t_i(a)$. Dimostrare che

   $$e_i^2 = e_i, \quad e_i e_j = 0 \ (i \neq j), \qquad e_1 + \cdots + e_r = 1, \tag{4}$$

   e quindi

   $$F[a] = F[a]e_1 \oplus F[a]e_2 \oplus \cdots \oplus F[a]e_r, \tag{5}$$

   nel senso che ogni elemento si scrive in modo unico $b_1 + \cdots + b_r$ con $b_i \in F[a]e_i$. Posto $a_i = ae_i$, mostrare che $F[a]e_i$ è un'algebra di unità e_i e che in questa algebra a_i è algebrico di polinomio minimo $q_i(x)$.

5. Nelle ipotesi dell'esercizio 4 sia $f(x) = (x - b_1) \cdots (x - b_r)$ in $F[x]$, con i b_i distinti. Sia $e_i(x)$ il polinomio interpolatore di Lagrange

   $$e_i(x) = \frac{f(x)}{(x - b_i)f'(b_i)}, \qquad 1 \leq i \leq r. \tag{6}$$

   Mostrare che per $e_i = e_i(a)$ valgono (4) e (5), con $F[a]e_i = Fe_i$ di dimensione uno.

6. Per l'algebra di gruppo $F[G]$ di $G = \{s_1 = 1, s_2, \dots, s_n\}$ porre $T(a) = \sum \alpha_i$ se $a = \sum \alpha_i s_i$. Dimostrare che $a \mapsto T(a)$ è un omomorfismo di $F[G]$ in F e determinarne il nucleo.

7. Con le notazioni dell'esercizio 6 dimostrare che esiste un omomorfismo di $F[G]$ in $F[G \times G]$ che manda ogni $s_i \in G$ in $s_i \times s_i$.

8. Sempre nelle ipotesi dell'esercizio 6, siano G di ordine p^m, con p primo, e F di caratteristica p. Posto $K = \ker T$, dimostrare che ogni elemento di K è nilpotente.

9. Mostrare che $A = \sum_{i=1}^{n-1} e_{i,i+1}$ e $B = \sum_{i=1}^{n-1} e_{i+1,i}$ generano l'algebra $M_n(F)$ delle matrici $n \times n$ su F.

10. Se $c_1, \dots, c_n$ sono elementi distinti di F, mostrare che $C = \sum_i c_i e_{ii}$ e $D = \sum_{i=1}^{n-1} e_{i,i+1} + e_{n,1}$ generano $M_n(F)$.

11. Dimostrare che, se $C \in M_n(F)$ e il polinomio caratteristico di C ha n radici distinte in un campo di spezzamento, esiste $D \in M_n(F)$ tale che C e D generino $M_n(F)$.

7.2 Algebre esterne. Applicazione ai determinanti

Definiremo alcune algebre che hanno importanti applicazioni geometriche e permettono di ricavare con chiarezza le principali proprietà dei determinanti. Queste algebre, oggi dette *algebre esterne*, furono introdotte nel 1844 da H. G. Grassmann.[3] Esse nascono dal problema seguente: dato uno spazio vettoriale V di dimensione finita su F, lo si vuole ampliare a un'algebra A generata da V, nella quale $v^2 = 0$ per ogni $v \in V$. Si vuole farlo nel modo più generale possibile, imponendo soltanto le condizioni enunciate e le loro conseguenze.

Supponiamo di avere un'algebra A contenente un sottospazio V che la genera e i cui elementi hanno quadrato nullo. Come nel § 7.1, ogni elemento di A è combinazione F-lineare di 1 e di monomi $v_1 \cdots v_k$, $k \geq 1$, con $v_i \in V$. Se $(u_1, \dots, u_n)$ è una base di V, ciascun v è combinazione lineare degli u_i e ciascun monomio nei v è combinazione lineare di monomi negli u_i. Quindi anche gli u_i generano A. Un monomio $u_{i_1} \cdots u_{i_r}$ si dirà *standard* se $i_1 < i_2 < \cdots < i_r$. Dimostreremo che ogni elemento di A è combinazione lineare di 1 e di monomi standard. Basta provarlo per i monomi negli u_i. Per questi proveremo anzi che un monomio in cui uno stesso u_i compare più volte è nullo e che, se $i_1 < \cdots < i_r$, vale

$$u_{i_{1'}} \cdots u_{i_{r'}} = (\operatorname{sg}\sigma) u_{i_1} \cdots u_{i_r}, \tag{7}$$

dove $\sigma = \left(\begin{smallmatrix} 1 & 2 & \cdots & r \\ 1' & 2' & \cdots & r' \end{smallmatrix}\right)$ e $\operatorname{sg}\sigma$ vale 1 o -1 secondo che σ sia pari o dispari. Poiché ogni elemento di V ha quadrato nullo, $0 = (u+v)^2 - u^2 - v^2 = uv + vu$. Ne segue $uv = -vu$ e in particolare

$$u_i^2 = 0, \quad u_i u_j = -u_j u_i, \qquad 1 \leq i, j \leq n. \tag{8}$$

La seconda relazione permette di scambiare due fattori consecutivi cambiando segno. Con tali scambi si può portare un'occorrenza di u_i accanto a un'altra e, per la prima relazione, un monomio che contiene due volte u_i è nullo. Nel prodotto a sinistra della (7), se $i_{j'} > i_{(j+1)'}$, scambiando i corrispondenti fattori cambia il segno e la permutazione differisce dalla precedente per una trasposizione. Un numero finito di simili scambi porta gli indici in ordine crescente. Il numero degli scambi è quello delle trasposizioni in una decomposizione di σ, e la sua parità dà la (7).

Ogni elemento di A è dunque combinazione lineare degli elementi

$$1, \quad u_{i_1} \cdots u_{i_r}, \qquad i_1 < i_2 < \cdots < i_r. \tag{9}$$

Essi non sono più numerosi dei sottoinsiemi $\{i_1, \dots, i_r\}$ di $N = \{1, \dots, n\}$, compreso il vuoto.[4] Otteniamo un fatto forse inatteso: A ha dimensione finita e $\dim A \leq |\mathcal{P}(N)| = 2^n$. Calcoliamo il prodotto di due monomi della (9). Per $S = \{i_1 < \cdots < i_r\}$ poniamo $u_S = u_{i_1} \cdots u_{i_r}$ e $u_\varnothing = 1$. Per $s, t \in N$ definiamo

$$\varepsilon_{s,t} = \begin{cases} 1 & s < t, \\ 0 & s = t, \\ -1 & s > t, \end{cases} \tag{10}$$

e per $S, T \subseteq N$ poniamo

$$\varepsilon_{S,T} = \begin{cases} \displaystyle\prod_{s \in S,\ t \in T} \varepsilon_{s,t} & S, T \neq \varnothing, \\ 1 & S = \varnothing \text{ o } T = \varnothing. \end{cases} \tag{11}$$

[3] Si veda H. G. Grassmann, *Ausdehnungslehre*. La prima edizione uscì nel 1844. Una seconda, ampliata e migliorata, apparve nel 1862.

[4] Alcuni elementi della (9) potrebbero coincidere e, anche se distinti, potrebbero essere linearmente dipendenti.

Se T_1, T_2 sono non vuoti e disgiunti, allora $\varepsilon_{S,T_1\cup T_2} = \varepsilon_{S,T_1}\varepsilon_{S,T_2}$ e $\varepsilon_{T_1\cup T_2,S} = \varepsilon_{T_1,S}\varepsilon_{T_2,S}$. Ne segue facilmente che

$$u_S u_T = \varepsilon_{S,T} u_{S\cup T}. \tag{12}$$

Costruiamo ora $E(V)$. Prendiamo lo spazio vettoriale di dimensione 2^n con base l'insieme $\mathcal{P}(N)$ dei sottoinsiemi di N. I suoi elementi sono $\sum_{S\in\mathcal{P}(N)} a_S S$, con $a_S \in F$. Identifichiamo S con $1S \in E(V)$. Definiamo il prodotto degli elementi della base mediante

$$ST = \varepsilon_{S,T}(S \cup T) \tag{13}$$

ed estendiamolo per linearità a tutto $E(V)$:

$$\left(\sum_S a_S S\right)\left(\sum_T b_T T\right) = \sum_{S,T} \varepsilon_{S,T} a_S b_T (S \cup T). \tag{14}$$

Dalla (14) seguono le leggi distributive e $a(XY) = (aX)Y = X(aY)$ per $a \in F$. Poiché $\varepsilon_{\varnothing,S} = \varepsilon_{S,\varnothing} = 1$, le (13) e (14) mostrano che $\varnothing$ è l'unità moltiplicativa; d'ora in poi la scriveremo 1. Verifichiamo $(RS)T = R(ST)$ per $R, S, T \in \mathcal{P}(N)$. Se uno dei tre è 1, è evidente. Se i tre insiemi non sono a due a due disgiunti, entrambi i membri sono nulli per la (13). Quando sono disgiunti, il coefficiente di $R \cup S \cup T$ in $(RS)T$ è $\varepsilon_{R,S}\varepsilon_{R\cup S,T} = \varepsilon_{R,S}\varepsilon_{R,T}\varepsilon_{S,T}$; in $R(ST)$ è $\varepsilon_{S,T}\varepsilon_{R,S\cup T} = \varepsilon_{S,T}\varepsilon_{R,S}\varepsilon_{R,T}$. L'uguaglianza prova l'associatività in $E(V)$, che è quindi un'algebra.

Identifichiamo u_i con l'elemento $\{i\}$ della base di $E(V)$. Ciò immerge V in $E(V)$ come insieme delle somme $\sum a_i u_i$. La (13) dà $u_i^2 = 0$, $u_i u_j = -u_j u_i$ e $u_{i_1}\cdots u_{i_r} = \{i_1, \ldots, i_r\}$ per gli indici crescenti. Se $v = \sum a_i u_i$, allora $v^2 = \sum a_i^2 u_i^2 + \sum_{i<j} a_i a_j (u_i u_j + u_j u_i) = 0$. Dunque V genera $E(V)$ come algebra e ogni suo elemento ha quadrato nullo. Inoltre $\dim E(V) = 2^n$, poiché l'unità e i monomi standard costituiscono una base. Chiameremo $E(V)$ *algebra esterna di* V. Più della costruzione specifica conta la proprietà seguente, che caratterizza il risultato finale (esercizio 1).

Teorema 7.1. *Sia $L : V \to A$ un'applicazione lineare in un'algebra A tale che $(Lv)^2 = 0$ per ogni $v \in V$. Allora L si estende in uno e un solo modo a un omomorfismo $\eta(L) : E(V) \to A$.*

Dimostrazione. Scriviamo $\bar v = Lv$. Da $\bar v^2 = 0$ segue come prima $\bar u \bar v = -\bar v \bar u$. Ripetendo il ragionamento che conduce alla (12), se $\bar u_S = \bar u_{i_1}\cdots\bar u_{i_r}$ per $S = \{i_1 < \cdots < i_r\}$, si ha $\bar u_S \bar u_T = \varepsilon_{S,T}\bar u_{S\cup T}$. Sia $\eta(L)$ l'applicazione lineare definita sulla base da $\eta(L)\varnothing = 1$ e $\eta(L)S = \bar u_S$. Dato che $\eta(L)u_i = \bar u_i = Lu_i$, essa estende L. Inoltre

$$\eta(L)(ST) = \varepsilon_{S,T}\eta(L)(S\cup T) = \varepsilon_{S,T}\bar u_{S\cup T} = \bar u_S \bar u_T = (\eta(L)S)(\eta(L)T).$$

Per linearità, $\eta(L)(XY) = (\eta(L)X)(\eta(L)Y)$ per ogni $X = \sum a_S S$ e $Y = \sum b_T T$. Poiché $\eta(L)1 = 1$, è un omomorfismo di algebre. L'unicità è chiara. □

Corollario 1. *Se U è un sottospazio di V, la sottoalgebra di $E(V)$ generata da U è isomorfa a $E(U)$.*

Dimostrazione. Per qualsiasi base $(u_1, \ldots, u_n)$ di V, l'unità e i monomi standard generano linearmente $E(V)$, come per l'algebra A considerata all'inizio. Poiché sono $2^n = \dim E(V)$, sono indipendenti. Scegliamo la base in modo che $(u_1, \ldots, u_m)$ sia base di U. L'unità e i monomi standard nei primi m vettori sono indipendenti e appartengono alla sottoalgebra B generata da U, dunque $\dim B \geq 2^m$. D'altra parte ogni $v \in U$ ha quadrato nullo, perciò il teorema dà un omomorfismo di $E(U)$ in $E(V)$ che coincide con l'inclusione su U. La sua immagine è B. Poiché $\dim E(U) = 2^m \leq \dim B$, le dimensioni coincidono e l'omomorfismo è un isomorfismo su B. □

Corollario 2. *Se L è una trasformazione lineare di V e $\eta(L)$ l'endomorfismo di $E(V)$ determinato da L, allora*

$$\eta(1) = 1, \qquad \eta(L_1 L_2) = \eta(L_1)\eta(L_2). \tag{15}$$

Se L è biiettiva, $\eta(L)$ è un automorfismo.

Dimostrazione. L'automorfismo identico di $E(V)$ induce l'identità su V, perciò $\eta(1) = 1$. I due endomorfismi nella seconda uguaglianza inducono entrambi L_1L_2 su V, dunque coincidono. Se L è biiettiva, $LL^{-1} = 1 = L^{-1}L$ implica $\eta(L)\eta(L^{-1}) = 1 = \eta(L^{-1})\eta(L)$. □

Prima delle applicazioni osserviamo che $E(V)$ ammette la decomposizione diretta

$$E(V) = F1 \oplus V \oplus V^2 \oplus V^3 \oplus \cdots \oplus V^n, \tag{16}$$

dove V^r è generato dai prodotti $v_1 \cdots v_r$ di vettori di V. Infatti un tale prodotto è combinazione di monomi di grado r negli elementi della base, ciascuno nullo o, a meno del segno, standard. I monomi standard di grado r sono dunque una base di V^r. Ciò dà la (16) e, contando tali monomi,

$$\dim V^r = \binom{n}{r}. \tag{17}$$

In particolare $\dim V^n = 1$, con base $u_1 \cdots u_n$. Se $L : V \to V$ è lineare, dalla definizione di V^r segue

$$\eta(L)V^r \subseteq V^r. \tag{18}$$

In particolare $\eta(L)(u_1 \cdots u_n) = \Delta u_1 \cdots u_n$ con $\Delta \in F$. Scriviamo

$$Lu_i = \sum_j l_{ij}u_j, \qquad 1 \le i \le n, \tag{19}$$

così che $\Lambda = (l_{ij})$ è la matrice di L. Allora

$$\begin{aligned}\Delta u_1 \cdots u_n = \eta(L)u_1 \cdots \eta(L)u_n &= \sum_{j_1,\cdots,j_n} l_{1j_1} \cdots l_{nj_n} u_{j_1} \cdots u_{j_n} \\ &= \sum_{\sigma \in S_n} (\operatorname{sg}\sigma) l_{1,\sigma(1)} \cdots l_{n,\sigma(n)} u_1 \cdots u_n\end{aligned}$$

per la (7). Per definizione di determinante otteniamo dunque

$$\eta(L)(u_1 \cdots u_n) = (\det \Lambda)u_1 \cdots u_n. \tag{20}$$

Se Λ_i è la matrice di L_i secondo la (19), quella di L_1L_2 è $\Lambda_2\Lambda_1$. Da una parte $\eta(L_1L_2)$ moltiplica $u_1 \cdots u_n$ per $\det(\Lambda_2\Lambda_1)$, dall'altra $\eta(L_1)\eta(L_2)$ lo moltiplica per $\det\Lambda_2 \det\Lambda_1$. La (15) dà quindi la proprietà moltiplicativa per le matrici di $M_n(F)$:

$$\det(\Lambda_1\Lambda_2) = (\det\Lambda_1)(\det\Lambda_2). \tag{21}$$

Applichiamo lo stesso metodo alla formula di Laplace per lo sviluppo secondo i minori di un insieme di righe e i corrispondenti cofattori. Fissiamo $S = \{i_1 < \cdots < i_k\} \subseteq N$. Per $T = \{j_1 < \cdots < j_k\}$ sia $\Lambda_{S,T}$ il minore delle righe di indici in S e delle colonne di indici in T. Dalla (7) si ricava

$$\eta(L)(u_{i_1} \cdots u_{i_k}) = \sum_{|T|=k} \Lambda_{S,T} u_{j_1} \cdots u_{j_k}. \tag{22}$$

Sia $S' = \{i_{k+1} < \cdots < i_n\}$ il complementare di S. Allora

$$\eta(L)(u_{i_{k+1}} \cdots u_{i_n}) = \sum_{|T'|=n-k} \Lambda_{S',T'} u_{j_{k+1}} \cdots u_{j_n}, \tag{23}$$

con $T' = \{j_{k+1} < \cdots < j_n\}$. Il prodotto ST' è nullo salvo quando $T' = S'$. In tal caso

$$u_{i_1} \cdots u_{i_k} u_{i_{k+1}} \cdots u_{i_n} = \varepsilon_{S,S'} u_1 \cdots u_n, \tag{24}$$

dove

$$\varepsilon_{S,S'} = (-1)^{\lambda_{S,S'}}, \qquad \lambda_{S,S'} = (i_1 - 1) + (i_2 - 2) + \cdots + (i_k - k). \tag{25}$$

Infatti i numeri di N minori di i_1 sono $i_1 - 1$, tutti in S'; quelli minori di i_2 sono $i_2 - 1$, tutti tranne uno in S', e così via. Applicando $\eta(L)$ al prodotto in (24) e usando (22) e (23) si ottiene

$$(-1)^{\sum_q (i_q - q)} \det \Lambda = \sum_T \Lambda_{S,T} \Lambda_{S',T'} (-1)^{\sum_q (j_q - q)},$$

dove T' è ora il complementare di T. Definiamo il cofattore $\Lambda'_{S,T}$ come il minore complementare $\Lambda_{S',T'}$ moltiplicato per $(-1)^{\sum_q (i_q + j_q)}$. Si ha così lo sviluppo di Laplace

$$\det \Lambda = \sum_T \Lambda_{S,T} \Lambda'_{S,T} \tag{26}$$

secondo i minori delle righe $i_1, \ldots, i_k$. Per $k = 1$ si ottiene lo sviluppo secondo gli elementi di una riga e i relativi cofattori.

Ordiniamo lessicograficamente i sottoinsiemi S di cardinalità k. La (22) mostra che la matrice della trasformazione indotta su V^k da $\eta(L)$ è $C_k(\Lambda) = (\Lambda_{S,T})$. Ha $\binom{n}{k}$ righe e altrettante colonne ed è detta *k-esima matrice composta* di Λ. Dalla (15) segue

$$C_k(\Lambda_1) C_k(\Lambda_2) = C_k(\Lambda_1 \Lambda_2), \tag{27}$$

che generalizza la proprietà moltiplicativa dei determinanti.

I risultati sui determinanti di matrici su un campo valgono, come è noto, su ogni anello commutativo. Vediamo come passare al caso generale partendo da un solo campo $\mathbb{Q}(x_1, \ldots, x_m)$ con un numero sufficiente di indeterminate. Per il teorema di moltiplicazione scegliamo $m = 2n^2$ e chiamiamo le indeterminate x_{ij}, y_{ij}, $1 \le i, j \le n$. Per $X = (x_{ij})$, $Y = (y_{ij})$ vale $\det(XY) = \det X \det Y$ sul campo $\mathbb{Q}(x_{ij}, y_{ij})$, quindi già su $\mathbb{Z}[x_{ij}, y_{ij}]$. Per ogni anello commutativo R esiste un omomorfismo di quest'anello di polinomi in R che manda le $2n^2$ indeterminate in elementi prescritti di R.[5] Date $\Lambda_1, \Lambda_2 \in M_n(R)$, assegniamo a x_{ij} l'elemento (i, j) di Λ_1 e a y_{ij} quello di Λ_2. L'omomorfismo trasforma l'identità precedente nella (21) per matrici su R. Analogamente per gli altri teoremi.

Fra le funzioni polinomiali su $M_n(F)$ (§ 2.12) la proprietà moltiplicativa permette di caratterizzare il determinante. Ci occorre il risultato seguente.

Teorema 7.2. *Sia $F[x_{ij}]$ l'anello dei polinomi nelle n^2 indeterminate x_{ij} sul campo F e sia $X = (x_{ij})$. Allora* $\det X$ *è irriducibile in $F[x_{ij}]$.*

Dimostrazione. Procediamo per induzione su n, ricordando che $F[x_{ij}]$ è fattoriale (p. 154 dell'originale). Scriviamo $\det X = \sum_i x_{1i} X_{1i}$, con X_{ij} cofattore di x_{ij}. Sia D l'anello dei polinomi nelle indeterminate diverse da x_{11}. Allora $F[x_{ij}] = D[x_{11}]$ e $\det X = x_{11} X_{11} + Y$ con $Y \in D$. Per induzione X_{11} è irriducibile nelle variabili con $i, j > 1$, dunque anche in D. Poiché il determinante ha grado uno in x_{11}, ogni sua fattorizzazione ha la forma $(Px_{11} + Q)R$ con $P, Q, R \in D$. Si ha $PR = X_{11}$, perciò, passando ad associati, possiamo supporre $P = 1$ oppure $P = X_{11}$. Nel secondo caso $R = 1$ e la fattorizzazione è banale. Se il determinante

[5] L'omomorfismo $\mathbb{Z} \to R$, $m \mapsto m1$, si estende all'anello di polinomi assegnando valori arbitrari alle indeterminate.

è riducibile, $P = 1$, $R = X_{11}$ e $X_{11} \mid \det X$. Analogamente $X_{ii} \mid \det X$ per ogni i. Questi cofattori sono irriducibili e non associati a due a due, perciò $X_{11}X_{22}\cdots X_{nn} \mid \det X$. Il primo membro ha grado totale $n(n-1)$, il secondo n, impossibile se $n > 2$. Per $n = 2$ ne seguirebbe $x_{11}x_{22} \mid (x_{11}x_{22} - x_{12}x_{21})$, altrettanto impossibile. □

Teorema 7.3. *Sia F un campo infinito e sia $Q(x_{ij})$ un polinomio omogeneo di grado q nelle indeterminate x_{ij}, $1 \le i, j \le n$. Supponiamo che la corrispondente funzione polinomiale $A = (a_{ij}) \mapsto Q(a_{ij}) = Q(A)$ su $M_n(F)$ soddisfi (i) $Q(1) = 1$ e (ii) $Q(AB) = Q(A)Q(B)$. Allora $Q(x_{ij})$ è una potenza di* $\det X$.

Dimostrazione. Per $A \in M_n(F)$ consideriamo la matrice aggiunta classica $\operatorname{adj} A$ (p. 96 dell'originale). Si ha $A \operatorname{adj} A = (\det A)1$, quindi l'omogeneità dà

$$Q(A)Q(\operatorname{adj} A) = Q((\det A)1) = (\det A)^q Q(1) = (\det A)^q.$$

Gli elementi di $\operatorname{adj} X$ sono polinomi di grado $n-1$. Il polinomio $P = Q(X)Q(\operatorname{adj} X) - (\det X)^q$ si annulla per ogni specializzazione degli x_{ij} in F. Poiché F è infinito, il teorema 2.19 implica $P = 0$. Pertanto $Q(X) \mid (\det X)^q$. Dall'irriducibilità di $\det X$ e da $\det 1 = Q(1) = 1$ segue $Q(X) = (\det X)^m$ per un opportuno m (con $1 \le m \le q$ se $q > 0$). □

Il determinante è così caratterizzato come il polinomio di minimo grado positivo con queste proprietà. Ne ricaviamo altri risultati.

Corollario 1. *Per $A \in M_n(R)$, con R anello commutativo, vale* $\det A = \det({}^tA)$, *dove ${}^tA$ è la trasposta di A.*

Dimostrazione. Per il metodo delle indeterminate basta considerare $\mathbb{Q}(x_{ij})$. Su un campo infinito F, poniamo $Q(X) = \det({}^tX)$. Si ha $Q(1) = 1$ e

$$Q(AB) = \det({}^tB\,{}^tA) = \det({}^tB)\det({}^tA) = Q(B)Q(A) = Q(A)Q(B).$$

Per il teorema 7.3 Q è una potenza di $\det X$ e, poiché i gradi coincidono, $Q = \det X$. Il risultato vale dunque per ogni R. □

Dallo sviluppo di Laplace secondo le righe si ricava quello secondo le colonne. Dal teorema 7.3 e da un confronto fra i gradi segue anche il risultato di cui lasciamo al lettore la dimostrazione.

Corollario 2. *Se $C_r(A)$ è l'r-esima matrice composta di A, allora* $\det C_r(A) = (\det A)^{\binom{n-1}{r-1}}$.

L'algebra esterna, e precisamente il sottospazio V^r, permette di introdurre coordinate nell'insieme $\Gamma_r(V)$ dei sottospazi di dimensione r di V. Per il corollario 1 del teorema 7.1, $E(U)$ si identifica con la sottoalgebra di $E(V)$ generata da U. Se $(v_1, \dots, v_r)$ è una base di U, questi vettori generano tale sottoalgebra. Abbiamo $E(U) = F \oplus U \oplus U^2 \oplus \cdots \oplus U^r$ e $U^r = Fv_1 \cdots v_r = E(U) \cap V^r$. Il prodotto $v_1 \cdots v_r$ è determinato dal sottospazio U a meno di un fattore non nullo in F: un'altra base dà $v_1' \cdots v_r' = \rho v_1 \cdots v_r$ con $\rho \neq 0$.

Diremo *decomponibile* un elemento di $E(V)$ della forma $v_1 \cdots v_r$ con fattori linearmente indipendenti in V. Un sottospazio U di dimensione r determina una retta Fu con $u = v_1 \cdots v_r$ decomponibile. Sottospazi distinti danno rette distinte. Infatti, se $U' \neq U$, scegliamo una base $(v_1, \dots, v_n)$ di V in modo che $(v_1, \dots, v_r)$ sia base di U, $(v_1, \dots, v_s)$ sia base di $U \cap U'$ con $0 \le s < r$, e $(v_1, \dots, v_s, v_{r+1}, \dots, v_{2r-s})$ sia base di U'. La base corrispondente di V^r mostra che $v_1 \cdots v_r$ e $v_1 \cdots v_s v_{r+1} \cdots v_{2r-s}$ sono indipendenti.

Abbiamo quindi una corrispondenza biunivoca tra $\Gamma_r(V)$ e le rette Fu generate da un elemento decomponibile $u \in V^r$. Alla base $(v_1, \dots, v_r)$ di U corrisponde la retta generata da $u = v_1 \cdots v_r$; inversamente, a tale u corrisponde $U = \sum Fv_i$. Fissiamo una base $(u_1, \dots, u_n)$ di V e ordiniamo lessicograficamente i $\binom{n}{r}$ monomi standard di grado r, base di V^r. Se

$$v_1 \cdots v_r = \sum_{i_1 < \cdots < i_r} \Delta_{i_1 \dots i_r} u_{i_1} \cdots u_{i_r},$$

le coordinate $(\Delta_{i_1 \dots i_r})$ sono determinate da $U = \sum Fv_i$ a meno di un fattore comune non nullo. Si chiamano *coordinate di Plücker* di U rispetto alla base scelta di V.

Esercizi

1. Verificare che il teorema 7.1 caratterizza $E(V)$: se $E'(V)$ è una seconda algebra con la proprietà enunciata, esiste un isomorfismo $E(V) \to E'(V)$ che è l'identità su V.

2. Completare il corollario 2 del teorema 7.1 mostrando che $\eta(L)$ è un automorfismo soltanto se L è biiettiva.

3. Siano x_{ij}, $i \le j$, indeterminate su F. Sia $X = (x_{ij})$, con $x_{ji} = x_{ij}$, la matrice simmetrica «generica». Dimostrare che $\det X$ è irriducibile in $F[x_{ij}]$.

4. Siano x_{ij}, $1 \le i < j \le n$, indeterminate su F e sia X la matrice avente questi elementi sopra la diagonale, $x_{ii} = 0$, $x_{ji} = -x_{ij}$. Mostrare che lo pfaffiano $\operatorname{Pf} X$ è irriducibile in $F[x_{ij}]$ (§ 6.2, p. 352 dell'originale).

5. Applicare lo sviluppo di Laplace per dimostrare che, se M è triangolare inferiore a blocchi, con blocchi diagonali quadrati $M_1, \dots, M_r$, allora $\det M = \prod_i \det M_i$.

Gli esercizi 6–10 delineano una seconda dimostrazione del teorema 7.3, in cui non si suppone Q omogeneo. Questa dimostrazione mi è stata comunicata da George Seligman.

6. Sia Q come nel teorema 7.3, senza l'ipotesi di omogeneità. Restringere l'applicazione al gruppo D delle matrici diagonali $\operatorname{diag}\{a_1, \dots, a_n\}$ con $\prod a_i \neq 0$. La restrizione è data dal polinomio $Q(x_{11}, \dots, x_{nn}, 0, \dots, 0)$ ottenuto annullando le indeterminate fuori diagonale. Usare il teorema sull'indipendenza lineare dei caratteri distinti di un gruppo (p. 291 dell'originale) per provare che tale polinomio è $x_{11}^{m_1} \cdots x_{nn}^{m_n}$ con $m_i \ge 0$.

7. Mostrare che gli m_i sono uguali usando $P \operatorname{diag}\{a_1, \dots, a_n\} P^{-1} = \operatorname{diag}\{a_2, \dots, a_n, a_1\}$, dove $P = e_{12} + e_{23} + \cdots + e_{n-1,n} + e_{n,1}$.

8. Per un'estensione K/F, usare Q per definire un'applicazione $M_n(K) \to K$. Mostrare che $Q(AB) = Q(A)Q(B)$ vale anche in $M_n(K)$.

9. Dedurre che esiste $m \in \{0, 1, 2, \dots\}$ tale che $Q(A) = (\det A)^m$ per ogni A invertibile a radici caratteristiche distinte.

10. Sia $X = (x_{ij})$ e sia $\Delta \in F[x_{ij}]$ il discriminante del polinomio caratteristico $\det(\lambda 1 - X)$. Mostrare che $\Delta \det X((\det X)^m - Q(X))$ si annulla su tutte le matrici $X = A$ in $M_n(F)$. Dedurne $Q(X) = (\det X)^m$.

11. Sia $\mathbb{Z}[x_{ij}]$ l'anello dei polinomi nelle n^2 indeterminate x_{ij} e sia $X = (x_{ij})$. Scrivere $f(\lambda) = \det(\lambda 1 - X) = \lambda^n - p_1 \lambda^{n-1} + \cdots + (-1)^n p_n$, con $p_i \in \mathbb{Z}[x_{ij}]$, e $s_i = \operatorname{tr} X^i$. Diagonalizzare X in un'opportuna estensione del campo delle frazioni di $\mathbb{Z}[x_{ij}]$ e usare le identità di Newton (p. 140 dell'originale) per mostrare che $n!\, p_i \in \sum_{j=1}^{n} s_j \mathbb{Z}[s_1, \dots, s_n]$.

12. Usare l'esercizio 11 e il teorema di Hamilton–Cayley per mostrare che, se R è un anello commutativo e $A \in M_n(R)$ soddisfa $\operatorname{tr} A = \operatorname{tr} A^2 = \cdots = \operatorname{tr} A^n = 0$, allora $n! A^n = 0$.

13. Per $A_1, \dots, A_r \in M_n(R)$, con R commutativo, definire il polinomio standard $[A_1, \dots, A_r] = \sum_{\pi \in S_r} (\operatorname{sg} \pi) A_{\pi(1)} \cdots A_{\pi(r)}$. Mostrare che la sua traccia è nulla se r è pari.

14. Sia $E^+(V) = F1 + V^2 + V^4 + \cdots$ (si veda (16)). Mostrare che $E^+(V)$ è una sottoalgebra commutativa di $E(V)$.

15. *Teorema di Amitsur–Levitzki.* Dimostrare che $[A_1, \ldots, A_{2n}] = 0$ per $A_i \in M_n(R)$, R commutativo. (Traccia di una dimostrazione di S. Rosset: per multilinearità basta provare l'identità per le unità matriciali e_{ij}. È quindi sufficiente prendere $R = \mathbb{Z}$ o $\mathbb{Q}$. Scegliere $\mathbb{Q}$ e lo spazio V di base $(u_1, \ldots, u_{2n})$. Per $A = \sum_{i=1}^{2n} A_i u_i \in M_n(E(V))$, l'identità equivale ad $A^{2n} = 0$ e, in caratteristica zero, a $n!A^{2n} = 0$. Si ha $A^2 \in M_n(E^+(V))$ e $\operatorname{tr} A^{2k} = 0$ per $k = 1, 2, \ldots$ per l'esercizio 13. L'esercizio 12 dà quindi $n!(A^2)^n = 0$.)

16. Mostrare che $[A_1, \ldots, A_k] = 0$ per $k \geq 2n$ e qualsiasi scelta degli $A_i \in M_n(R)$, mentre, se $k < 2n$, esistono A_i con $[A_1, \ldots, A_k] \neq 0$. (Per la seconda parte prendere $A_1 = e_{11}$, $A_2 = e_{12}$, $A_3 = e_{22}$, $A_4 = e_{23}$, ecc.)

7.3 Rappresentazioni matriciali regolari delle algebre associative. Norme e tracce

Sia V uno spazio vettoriale sul campo F e sia $\operatorname{End}_F V = \operatorname{Hom}_F(V, V)$ l'insieme delle trasformazioni lineari di V in sé. Abbiamo visto nel § 3.3 che $\operatorname{End}_F V$ ammette una struttura di anello in cui la somma, il prodotto, lo zero e l'unità sono definiti da $(L + M)x = Lx + Mx$, $(LM)x = L(Mx)$, $0x = 0$ e $1x = x$, per $L, M \in \operatorname{End}_F V$, $x \in V$. Poiché F è commutativo, le moltiplicazioni per «scalari», ossia per elementi di F, sono trasformazioni lineari. Per un tale operatore $x \mapsto ax$, con $a \in F$, si ha infatti $a(x + y) = ax + ay$ e, per $b \in F$, $a(bx) = b(ax)$. Dunque $x \mapsto ax$ appartiene a $\operatorname{End}_F V$. Poiché $a(Lx) = L(ax)$ per ogni $L \in \operatorname{End}_F V$, l'applicazione $a_V : x \mapsto ax$ appartiene al centro di $\operatorname{End}_F V$. È immediato che $a \mapsto a_V$ è un monomorfismo di F in $\operatorname{End}_F V$, la cui immagine $F_V = \{a_V \mid a \in F\}$ è un sottoanello del centro di $\operatorname{End}_F V$.[6] Possiamo perciò dotare $\operatorname{End}_F V$ di una struttura di algebra in cui la struttura di anello è quella consueta e la struttura di spazio vettoriale è data dalla somma usuale e da $aL = a_V L = L a_V$ (cfr. § 7.1). D'ora in poi considereremo $\operatorname{End}_F V$ come algebra in questo senso e la chiameremo *algebra (associativa) delle trasformazioni lineari* dello spazio vettoriale V su F.

Supponiamo ora che V abbia dimensione finita su F, con base $(u_1, \ldots, u_n)$. Per $L \in \operatorname{End}_F V$ otteniamo la matrice $\Lambda = (l_{ij})$ di L rispetto alla base ordinata scrivendo $Lu_i = \sum_{j=1}^{n} l_{ij} u_j$ per $1 \leq i \leq n$. Passando alla base $(v_1, \ldots, v_n)$, con $v_i = \sum_j c_{ij} u_j$ e $\Gamma = (c_{ij})$ invertibile, la matrice di L nella nuova base è $\Gamma \Lambda \Gamma^{-1}$. Ricordiamo anche la definizione del polinomio caratteristico di Λ (p. 196 dell'originale):

$$f(\lambda) = \det(\lambda 1 - \Lambda). \tag{28}$$

Poiché

$$\lambda 1 - \Lambda = \begin{pmatrix} \lambda - l_{11} & -l_{12} & \cdots & -l_{1n} \\ -l_{21} & \lambda - l_{22} & \cdots & -l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -l_{n1} & -l_{n2} & \cdots & \lambda - l_{nn} \end{pmatrix}, \tag{29}$$

si ha

$$f(\lambda) = \lambda^n - \left(\sum_i l_{ii} \right) \lambda^{n-1} + \cdots + (-1)^n \det \Lambda. \tag{30}$$

[6] In realtà F_V è il centro di $\operatorname{End}_F V$. Si veda il corollario 2 del teorema 3.17, p. 208 dell'originale.

L'elemento $\sum_i l_{ii}$ è detto *traccia*, $\operatorname{tr}\Lambda$, della matrice Λ. Evidentemente

$$\operatorname{tr}(\Lambda_1+\Lambda_2)=\operatorname{tr}\Lambda_1+\operatorname{tr}\Lambda_2,\qquad \operatorname{tr}(a\Lambda)=a\operatorname{tr}\Lambda. \tag{31}$$

La traccia è dunque una funzione lineare su $M_n(F)$. Inoltre

$$\det(\Lambda_1\Lambda_2)=\det\Lambda_1\det\Lambda_2,\qquad \det(a\Lambda)=a^n\det\Lambda. \tag{32}$$

Se M è simile a Λ, ossia $M=\Gamma\Lambda\Gamma^{-1}$, le due matrici hanno lo stesso polinomio caratteristico, quindi la stessa traccia e lo stesso determinante. Questi oggetti dipendono perciò soltanto dalla trasformazione lineare L. Possiamo definire il polinomio caratteristico, la traccia e il determinante di L mediante una sua matrice qualsiasi.

Ricordiamo che $L\mapsto\Lambda$ da $\operatorname{End}_F V$ in $M_n(F)$, per una base assegnata, è un anti-isomorfismo di anelli (p. 111 dell'originale). Poiché la matrice di aL, $a\in F$, è $a\Lambda$, si tratta anzi di un anti-isomorfismo di algebre. È più comodo ragionare in termini di isomorfismi. Nel nostro caso basta passare dall'applicazione $L\mapsto\Lambda$ all'applicazione $L\mapsto{}^t\Lambda$. Poiché Λ e la sua trasposta hanno lo stesso polinomio caratteristico, quello di L si può calcolare indifferentemente dall'una o dall'altra. Al cambiare della base, ${}^t\Lambda$ viene sostituita da ${}^tM=\Delta^{-1}\,{}^t\Lambda\,\Delta$, dove $\Delta={}^t\Gamma$.

Sia ora A un'algebra associativa sul campo di base F. Mostriamo, con lo stesso metodo usato per il teorema di Cayley e per il corrispondente risultato sugli anelli, che A è isomorfa a un'algebra di trasformazioni lineari.

Teorema 7.4. *Ogni algebra associativa A è isomorfa a una sottoalgebra dell'algebra* $\operatorname{End}_F A$ *delle trasformazioni lineari dello spazio vettoriale A su F.*

Dimostrazione. Come per gli anelli (teorema 3.2, p. 162 dell'originale), un monomorfismo di A in $\operatorname{End}_F A$ è dato da $u\mapsto u_L$, dove u_L è la moltiplicazione a sinistra $x\mapsto ux$ in A. Dagli assiomi di algebra, $u_L(ax)=u(ax)=a(ux)=au_Lx$ per $a\in F$, dunque $u_L\in\operatorname{End}_F A$. Inoltre $u\mapsto u_L$ è un monomorfismo di anelli per il teorema 3.2. Poiché $(au)_Lx=a(ux)$ e $a(u_Lx)=a(ux)$, è un monomorfismo di algebre. La sua immagine A_L è una sottoalgebra di $\operatorname{End}_F A$ isomorfa ad A. □

Un omomorfismo di un'algebra A su F in un'algebra $\operatorname{End}_F V$ di trasformazioni lineari di uno spazio vettoriale V su F si chiama *rappresentazione* di A. La particolare rappresentazione $u\mapsto u_L$ usata nella dimostrazione è la *rappresentazione regolare* di A. Data una rappresentazione di A su uno spazio vettoriale V di dimensione finita, possiamo comporla con l'isomorfismo $L\mapsto{}^t\Lambda$ di $\operatorname{End}_F V$ su $M_n(F)$ determinato da una base $(u_1,\dots,u_n)$ e ottenere un omomorfismo $u\mapsto\rho(u)$ da A in $M_n(F)$. Lo si chiama *rappresentazione matriciale* di A. Un cambiamento di base dà una rappresentazione *equivalente* (o *simile*), $u\mapsto\Delta^{-1}\rho(u)\Delta$. Le rappresentazioni matriciali di un'algebra di dimensione finita associate alla rappresentazione regolare si chiamano *rappresentazioni matriciali regolari* di A.

Sia $u\mapsto\rho(u)$ una rappresentazione matriciale regolare di A, di dimensione finita su F. Definiamo le funzioni *traccia* T e *norma* N su A mediante

$$T(u)=\operatorname{tr}\rho(u),\qquad N(u)=\det\rho(u),\qquad u\in A. \tag{33}$$

Poiché matrici simili hanno uguale traccia e determinante, queste funzioni non cambiano passando da una rappresentazione matriciale regolare a un'altra. Dal fatto che ρ è un omomorfismo di algebre e dalle proprietà della traccia e del determinante di matrici seguono

$$T(u+v)=T(u)+T(v),\qquad T(au)=aT(u),\qquad a\in F, \tag{34}$$

$$N(uv)=N(u)N(v),\qquad N(au)=a^nN(u), \tag{35}$$

dove $n=[A:F]$ è la dimensione. Nella sezione seguente vedremo che, quando A è un'estensione di Galois di F, queste definizioni danno le stesse funzioni del § 4.15.

Esempi

1. Sia $\mathbb{H}$ l'algebra dei quaternioni di Hamilton su $\mathbb{R}$, con base $(1, i, j, k)$ e tavola di moltiplicazione

$$i^2 = j^2 = k^2 = -1, \quad ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j.$$

Determiniamo la rappresentazione matriciale regolare di $\mathbb{H}$ data da questa base. Per $u = a_0 + a_1 i + a_2 j + a_3 k$ si ha

$$\begin{aligned} u1 &= a_0 + a_1 i + a_2 j + a_3 k, \\ ui &= -a_1 + a_0 i + a_3 j - a_2 k, \\ uj &= -a_2 - a_3 i + a_0 j + a_1 k, \\ uk &= -a_3 + a_2 i - a_1 j + a_0 k. \end{aligned}$$

Trasponendo la matrice dei coefficienti dei secondi membri si ottiene la rappresentazione

$$u \longmapsto \begin{pmatrix} a_0 & -a_1 & -a_2 & -a_3 \\ a_1 & a_0 & -a_3 & a_2 \\ a_2 & a_3 & a_0 & -a_1 \\ a_3 & -a_2 & a_1 & a_0 \end{pmatrix}. \tag{36}$$

2. Sia $A = F[u]$ con u algebrico di polinomio minimo $f(\lambda)$. Allora $A \cong F[\lambda]/(f(\lambda))$. Supponiamo $f(\lambda) = \lambda^n - a_{n-1}\lambda^{n-1} - \cdots - a_0$. La base è $(1, u, \ldots, u^{n-1})$ e si ha

$$u1 = u, \quad uu = u^2, \quad \ldots, \quad uu^{n-2} = u^{n-1}, \qquad uu^{n-1} = u^n = a_0 + a_1 u + \cdots + a_{n-1}u^{n-1}.$$

Dunque per la rappresentazione matriciale regolare associata alla base $(1, u, \ldots, u^{n-1})$ si ottiene

$$\rho(u) = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \ddots & \vdots & a_2 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & a_{n-1} \end{pmatrix}. \tag{37}$$

3. Come caso particolare dell'esempio precedente prendiamo $f(\lambda) = \lambda^n - 1$. Allora, per $x = x_0 + x_1 u + x_2 u^2 + \cdots + x_{n-1}u^{n-1}$, si calcola facilmente

$$\rho(x) = \begin{pmatrix} x_0 & x_{n-1} & x_{n-2} & \cdots & x_1 \\ x_1 & x_0 & x_{n-1} & \cdots & x_2 \\ x_2 & x_1 & x_0 & \cdots & x_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n-1} & x_{n-2} & x_{n-3} & \cdots & x_0 \end{pmatrix}. \tag{38}$$

Si ha $N(x) = \det \rho(x)$. Un determinante di questa forma è detto *determinante circolante*. L'esercizio 6 fornisce una formula per calcolarlo.

Esercizi

1. Sia A l'algebra con base $(e_1, \dots, e_n)$ in cui $e_i^2 = e_i$ e $e_i e_j = 0$ per $i \neq j$. Provare che per $x = \sum_{i=1}^n x_i e_i$ la matrice $\rho(x)$ relativa a questa base è $\operatorname{diag}(x_1, \dots, x_n)$.

2. Per $u = a_0 + a_1 i + a_2 j + a_3 k \in \mathbb{H}$ verificare che $T(u) = 4a_0$ e $N(u) = (a_0^2 + a_1^2 + a_2^2 + a_3^2)^2$.

3. Determinare $\rho(x)$ per $x = x_{11}e_{11} + x_{12}e_{12} + x_{21}e_{21} + x_{22}e_{22}$ nella base $(e_{11}, e_{21}, e_{12}, e_{22})$ di $M_2(F)$.

4. Sia $A = M_n(F)$. Per $X = (x_{ij}) \in M_n(F)$, provare che $T(X) = n \operatorname{tr} X$ e $N(X) = (\det X)^n$.

5. Sia A un'estensione di dimensione finita del campo F. Per $u \in A$ sia $m(\lambda)$ il polinomio minimo e sia $f(\lambda)$ il polinomio caratteristico di $\rho(u)$ in una rappresentazione matriciale regolare. Provare che $f(\lambda) = m(\lambda)^{[A:F(u)]}$. Se $m(\lambda) = \prod(\lambda - u_i)$ in un campo di spezzamento, mostrare che $N(u) = (\prod u_i)^{[A:F(u)]}$.

6. Supponiamo che F contenga n radici n-esime distinte dell'unità $\zeta_1, \dots, \zeta_n$. Per la $\rho(x)$ della (38), mostrare che $\det \rho(x) = \prod_{i=1}^n (x_0 + x_1\zeta_i + \cdots + x_{n-1}\zeta_i^{n-1})$.

7. Per $\Lambda_1, \Lambda_2 \in M_n(F)$ verificare che

$$\operatorname{tr}(\Lambda_1\Lambda_2) = \operatorname{tr}(\Lambda_2\Lambda_1). \tag{39}$$

   Dedurne che la traccia su un'algebra soddisfa

$$T(uv) = T(vu). \tag{40}$$

7.4 Cambiamento del campo di base. Transitività della traccia e della norma

Sia A un'algebra sul campo F e sia K un sottocampo di F con $[F : K] < \infty$. Come nel caso in cui A è un campo (teorema 4.2), si vede facilmente che $[A : K] = [A : F][F : K]$ e che, se $(u_1, \dots, u_n)$ è una base di A su F e $(v_1, \dots, v_r)$ una base di F su K, gli nr elementi $v_j u_i$ costituiscono una base di A su K. Possiamo considerare A come algebra sia su K sia su F, e F è un'algebra su K. Abbiamo dunque le applicazioni norma e traccia di A in F, di A in K e di F in K, che indicheremo rispettivamente con $N_{A/F}, T_{A/F}$, $N_{A/K}, T_{A/K}$ e $N_{F/K}, T_{F/K}$. Dimostreremo le seguenti formule di transitività per $u \in A$:

$$T_{A/K}(u) = T_{F/K}(T_{A/F}(u)), \tag{41}$$

$$N_{A/K}(u) = N_{F/K}(N_{A/F}(u)). \tag{42}$$

La prima è facile: basta esaminare le matrici in basi opportune. Scegliamo le basi u_i e v_j come sopra. La base di A su K è allora, nell'ordine,

$$(v_1u_1, \dots, v_ru_1;\ v_1u_2, \dots, v_ru_2;\ \dots;\ v_1u_n, \dots, v_ru_n). \tag{43}$$

Siano ρ la rappresentazione matriciale regolare di A nella base $(u_1, \dots, u_n)$ e μ quella di F su K nella base $(v_1, \dots, v_r)$. Per $u \in A$, $v \in F$ scriviamo $\rho(u) = (\rho_{ij}(u))$ e $\mu(v) = (\mu_{kl}(v))$. Tenendo presente la trasposizione che interviene nella definizione, si ha

$$uu_i = \sum_{j=1}^n \rho_{ji}(u)u_j, \qquad 1 \leq i \leq n, \tag{44}$$

$$vv_k = \sum_{l=1}^r \mu_{lk}(v)v_l, \qquad 1 \leq k \leq r. \tag{45}$$

Ne segue

$$
\begin{aligned}
u(v_k u_i) = v_k(uu_i) = v_k \sum_j \rho_{ji}(u) u_j = \sum_j (\rho_{ji}(u) v_k) u_j \\
= \sum_{j=1}^{n} \sum_{l=1}^{r} \mu_{lk}(\rho_{ji}(u)) v_l u_j.
\end{aligned} \tag{46}
$$

La rappresentazione matriciale regolare di A su K nella base (43) è dunque

$$
u \longmapsto \begin{pmatrix} \mu(\rho_{11}(u)) & \cdots & \mu(\rho_{1n}(u)) \\ \vdots & & \vdots \\ \mu(\rho_{n1}(u)) & \cdots & \mu(\rho_{nn}(u)) \end{pmatrix}. \tag{47}
$$

In altre parole, si prende una rappresentazione matriciale regolare di A/F e se ne sostituisce ogni elemento, che appartiene a F, con la matrice che lo rappresenta in una rappresentazione regolare di F/K. Dalla (46) risulta

$$
T_{A/K}(u) = \sum_i \operatorname{tr} \mu(\rho_{ii}(u)) = \sum_i T_{F/K}(\rho_{ii}(u)) = T_{F/K}\left(\sum_i \rho_{ii}(u) \right) = T_{F/K}(T_{A/F}(u)).
$$

Questo dimostra la (41).

Per la (42) dimostreremo una proprietà generale di transitività dei determinanti. Sia M una matrice $nr \times nr$ a elementi nel campo K, divisa in n righe e n colonne di blocchi A_{ij} di ordine r, e supponiamo che questi blocchi commutino fra loro. È proprio il caso della (47), poiché $\rho_{ij}(u) \in F$ e $v \mapsto \mu(v)$ è un omomorfismo. Le A_{ij} appartengono a un sottoanello commutativo B di $M_r(K)$. Possiamo quindi considerare M come elemento di $M_n(B)$ e calcolarne il determinante in B:

$$
\det_B M = \sum_{\sigma \in S_n} (\operatorname{sg} \sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}. \tag{48}
$$

Questo elemento appartiene al sottoanello generato dalle A_{ij} ed è indipendente dalla scelta di B. Essendo una matrice di $M_r(K)$, ha a sua volta un determinante in K. Vogliamo provare che[7]

$$
\det(\det_B M) = \det M. \tag{49}
$$

Supponiamo dapprima $\det A_{11} \neq 0$. Poiché A_{11}^{-1} commuta con ogni A_{ij}, aggiungendola a B possiamo supporre $A_{11}^{-1} \in B$. Eseguiamo in $M_n(B)$ il prodotto

$$
\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} 1 & -A_{11}^{-1}A_{12} & \cdots & -A_{11}^{-1}A_{1n} \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} A'_{11} & 0 & \cdots & 0 \\ A'_{21} & A'_{22} & \cdots & A'_{2n} \\ \vdots & \vdots & & \vdots \\ A'_{n1} & A'_{n2} & \cdots & A'_{nn} \end{pmatrix}
$$

con $A'_{11} = A_{11}$. Chiamando M' l'ultima matrice, otteniamo $\det_B M = \det_B M'$, poiché il fattore triangolare ha determinante 1 su B. Anche $\det M = \det M'$ su K. Basta quindi verificare la (49) per M'. Per $N' = (A'_{ij})_{2 \leq i,j \leq n}$ si ha $\det_B M' = A'_{11} \det_B N'$, donde $\det(\det_B M') = \det A'_{11} \det(\det_B N')$. Inoltre $\det M' = \det A'_{11} \det N'$ (esercizio 5 del § 7.2). Per induzione su n, $\det(\det_B N') = \det N'$, da cui l'uguaglianza desiderata.

[7] Questo risultato sembra dovuto a M. H. Ingraham, *Bulletin of the American Mathematical Society*, 43 (1937), pp. 579–580.

Se A_{11} non è invertibile, estendiamo il campo a $K(\lambda)$ con λ indeterminata. Sia $M(\lambda)$ la matrice ottenuta sostituendo A_{11} con $A_{11} - \lambda 1$. Sia $B(\lambda)$ un sottoanello commutativo di $M_r(K(\lambda))$ che contenga i blocchi di $M(\lambda)$. Poiché $\det(A_{11} - \lambda 1) = (-1)^r\lambda^r + \cdots \neq 0$, il caso precedente dà $\det(\det_{B(\lambda)} M(\lambda)) = \det M(\lambda)$. Questa identità in $K[\lambda]$ dà la (49) per $\lambda = 0$.

Applichiamo il risultato alle norme. Sia M la matrice a destra della (47), così che $N_{A/K}(u) = \det M$. Prendiamo per B l'anello commutativo delle matrici $\mu(v)$, $v \in F$. Poiché μ è un omomorfismo,

$$\det_B M = \mu(\det(\rho_{ij}(u))) = \mu(N_{A/F}(u)).$$

Dato che $\det \mu(v) = N_{F/K}(v)$, la (49) dà $\det M = \det \mu(N_{A/F}(u)) = N_{F/K}(N_{A/F}(u))$, ossia la (42).

Consideriamo ora il caso in cui $A = E$ è un'estensione di dimensione finita di F. Scriviamo il polinomio minimo di $u \in E$ su F nella forma

$$m_u(\lambda) = \lambda^m - a_1\lambda^{m-1} + a_2\lambda^{m-2} + \cdots + (-1)^m a_m. \tag{50}$$

Nella rappresentazione regolare di $F(u) = F[u]$ su F definita dalla base $(1, u, \ldots, u^{m-1})$, a u corrisponde la matrice

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & (-1)^{m-1}a_m \\ 1 & 0 & \cdots & 0 & (-1)^{m-2}a_{m-1} \\ 0 & 1 & \cdots & 0 & (-1)^{m-3}a_{m-2} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_1 \end{pmatrix}. \tag{51}$$

La traccia e il determinante valgono rispettivamente a_1 e a_m. Dunque $T_{F(u)/F}(u) = a_1$ e $N_{F(u)/F}(u) = a_m$. Inoltre $m = [F(u) : F]$ e $r = [E : F(u)] = [E : F]/m$. Essendo $u \in F(u)$, si ha $T_{E/F(u)}(u) = ru$ e $N_{E/F(u)}(u) = u^r$. La transitività dà $T_{E/F}(u) = T_{F(u)/F}(ru) = ra_1$ e analogamente $N_{E/F}(u) = a_m^r$. Così

$$T_{E/F}(u) = [E : F(u)]a_1, \tag{52}$$

$$N_{E/F}(u) = a_m^{[E:F(u)]}. \tag{53}$$

Se $m_u(\lambda) = \prod_{i=1}^m (\lambda - u_i)$ in un campo di spezzamento, si può sostituire nelle formule precedenti $a_1 = \sum_i u_i$, $a_m = \prod_i u_i$.

Supponiamo infine E/F di Galois, con gruppo $G = \{\eta_1 = 1, \eta_2, \ldots, \eta_n\}$. Il polinomio m_u si spezza in $E[\lambda]$ e l'insieme delle radici è l'orbita di $u = u_1$ sotto G. La successione $(\eta_1(u), \eta_2(u), \ldots, \eta_n(u))$ contiene r copie dell'orbita. Quindi $\sum_i \eta_i(u) = [E : F(u)]a_1$ e $\prod_i \eta_i(u) = a_m^{[E:F(u)]}$, ossia

$$T_{E/F}(u) = \sum_{i=1}^n \eta_i(u), \tag{54}$$

$$N_{E/F}(u) = \prod_{i=1}^n \eta_i(u). \tag{55}$$

Sono appunto le definizioni date nel § 4.15 per la traccia e la norma di un elemento di un'estensione di Galois.

7.5 Algebre non associative. Algebre di Lie e di Jordan

Un modo per cercare di creare nuova matematica a partire da una teoria esistente, specialmente se esposta in forma assiomatica, è generalizzarla sopprimendo o indebolendo alcune ipotesi. Facendo questo gioco con il concetto di algebra associativa si arriva naturalmente all'algebra non associativa: basta sopprimere la legge associativa della moltiplicazione. Se si giungesse a questo punto senza contatto con altre realtà matematiche, quasi certamente si abbandonerebbe presto il progetto. Sulle algebre non associative in generale si può dire ben poco di interessante. Si sono rivelate interessanti alcune classi particolari, portate all'attenzione degli algebristi dalle applicazioni, effettive o sperate, ad altri campi.

Esamineremo anzitutto i due esempi principali, le algebre di Lie e quelle di Jordan, cominciando dalle prime. Sophus Lie le introdusse col nome di «gruppi infinitesimi» nei suoi studi sui gruppi continui, più precisamente sui gruppi continui soggetti a opportune restrizioni che oggi chiamiamo gruppi di Lie. Non daremo qui definizioni precise: basti dire che un gruppo continuo è un concetto composto che coinvolge un gruppo e uno spazio topologico.[8] Ne è un esempio il gruppo $\mathrm{GL}_n(\mathbb{R})$ delle matrici reali invertibili $n \times n$. Oltre alla struttura di gruppo abbiamo quella di spazio topologico, che proviene dall'immersione di $\mathrm{GL}_n(\mathbb{R})$ in $M_n(\mathbb{R})$ e dal fatto che quest'ultimo si può riguardare come spazio euclideo di dimensione n^2. Il legame tra struttura algebrica e topologica consiste nella continuità della moltiplicazione di gruppo e dell'applicazione $X \mapsto X^{-1}$. Un altro esempio è il gruppo ortogonale reale $O_n(\mathbb{R})$. Un altro ancora è il gruppo di Lorentz, fondamentale nella teoria della relatività.

Il grande risultato di Sophus Lie fu la riduzione di problemi locali sui gruppi di Lie a problemi sulle algebre di Lie. A ogni gruppo di Lie è associata un'algebra di Lie. Per $\mathrm{GL}_n(\mathbb{R})$ essa è l'algebra di Lie $M_n(\mathbb{R})^-$ di tutte le matrici reali $n \times n$. La struttura di algebra di Lie su $M_n(\mathbb{R})^-$ è data dallo spazio vettoriale $M_n(\mathbb{R})$ con l'operazione di Lie, o commutatore (additivo),

$$[X, Y] = XY - YX. \tag{56}$$

L'algebra di Lie di $O_n(\mathbb{R})$ è l'insieme $\mathrm{Sk}_n(\mathbb{R})$ delle matrici antisimmetriche $n \times n$. Come $O_n(\mathbb{R})$ è un sottogruppo di $\mathrm{GL}_n(\mathbb{R})$, così $\mathrm{Sk}_n(\mathbb{R})$ è una sottoalgebra di $M_n(\mathbb{R})^-$, cioè un sottospazio di $M_n(\mathbb{R})$ chiuso rispetto al commutatore.

Gli esempi sono casi particolari di una costruzione generale che ricava algebre di Lie da algebre associative. Sia A un'algebra associativa su un campo qualsiasi F. Otteniamo una nuova struttura sostituendo al prodotto associativo il prodotto di Lie, o commutatore,

$$[x, y] = xy - yx. \tag{57}$$

Si ottiene così l'algebra di Lie A^-. È naturale considerare anche le sue sottoalgebre. Un'importante classe di esempi si presenta quando l'algebra associativa A ha un'involuzione j, cioè un antiautomorfismo $x \mapsto \bar{x}$ tale che $j^2 = 1$. È il caso di $A = M_n(F)$ con $j = t$, la trasposizione $X \mapsto {}^tX$. Indichiamo con $\mathrm{Sk}(A, j)$ l'insieme degli elementi antisimmetrici rispetto a j, cioè tali che $\bar{s} = -s$. Se $s_1, s_2 \in \mathrm{Sk}(A, j)$ e $a_1, a_2 \in F$, allora $s = a_1 s_1 + a_2 s_2 \in \mathrm{Sk}(A, j)$ poiché

$$\bar{s} = \overline{a_1 s_1 + a_2 s_2} = a_1 \bar{s}_1 + a_2 \bar{s}_2 = a_1(-s_1) + a_2(-s_2) = -(a_1 s_1 + a_2 s_2) = -s.$$

Inoltre

$$\begin{aligned} \overline{[s_1, s_2]} &= \overline{s_1 s_2 - s_2 s_1} = \bar{s}_2 \bar{s}_1 - \bar{s}_1 \bar{s}_2 \\ &= (-s_2)(-s_1) - (-s_1)(-s_2) = s_2 s_1 - s_1 s_2 = -[s_1, s_2]. \end{aligned}$$

[8] Una buona introduzione alla teoria di Lie si trova in P. M. Cohn, *Lie Groups*, Cambridge Tracts in Mathematics, n. 46, 1957, oppure in L. Pontrjagin, *Topological Groups*, Princeton University Press, 1939.

Dunque $\mathrm{Sk}(A,j)$ è una sottoalgebra di A^-.

Quali proprietà del prodotto di Lie in un'algebra associativa possiamo scoprire facilmente? Anzitutto, per $x, x_1, x_2, y, y_1, y_2 \in A$ e $a \in F$ è immediato che

$$[x_1+x_2,y]=[x_1,y]+[x_2,y], \qquad [x,y_1+y_2]=[x,y_1]+[x,y_2], \tag{58}$$
$$a[x,y]=[ax,y]=[x,ay].$$

Omettiamo la verifica, che è banale. Inoltre

$$[x,x]=0, \tag{59}$$

poiché $[x,x]=x^2-x^2$. Il prodotto $[xy]$, abbreviazione di $[x,y]$, è associativo? In termini del prodotto associativo di A si ha

$$[[x,y],z]=(xy-yx)z-z(xy-yx)=xyz-yxz-zxy+zyx,$$
$$[x,[y,z]]=x(yz-zy)-(yz-zy)x=xyz-xzy-yzx+zyx.$$

L'associatività equivale quindi a $yxz+zxy=xzy+yzx$, ovvero $y(xz-zx)-(xz-zx)y=0$, ossia $[y,[x,z]]=0$. Essa vale solo se quest'ultima uguaglianza è verificata per ogni $x,y,z\in A$. Già il primo esempio che viene in mente, $A=M_2(F)$, mostra che ciò in generale non accade. Prendendo $x=e_{12}$, $y=e_{12}$, $z=e_{21}$, si ottiene $[y,[x,z]]=-2e_{12}$ (non nullo in caratteristica diversa da due). Riesaminando i calcoli, troviamo un risultato positivo sui commutatori iterati: $[[x,y],z]-[x,[y,z]]=[[x,z],y]$. Poiché $[x,x]=0$, si ha $0=[x+y,x+y]=[x,x]+[x,y]+[y,x]+[y,y]$, per cui $[x,y]=-[y,x]$. Usando questa proprietà e la relazione precedente otteniamo l'*identità di Jacobi*

$$[[x,y],z]+[[y,z],x]+[[z,x],y]=0. \tag{60}$$

In altri termini, permutando ciclicamente i tre elementi di un prodotto $[[x,y],z]$ e sommando, si ottiene zero.

Useremo fra poco queste proprietà per definire le algebre di Lie astratte. Prima consideriamo la seconda classe di algebre non associative di questo paragrafo: le algebre di Jordan. Partiamo da un'algebra associativa A su un campo F di caratteristica diversa da 2. Introduciamo il *prodotto di Jordan* (chiamato dai fisici anticommutatore)

$$x\cdot y=\tfrac{1}{2}(xy+yx). \tag{61}$$

Sostituendo al prodotto associativo xy il prodotto di Jordan $x\cdot y$ otteniamo l'algebra di Jordan A^+. Anche le sottoalgebre di A^+ sono algebre di Jordan. Per esempio, se j è un'involuzione di A, l'insieme $\mathrm{Sym}(A,j)$ degli elementi simmetrici rispetto a j ne è una sottoalgebra. Come per $\mathrm{Sk}(A,j)$ si verifica che è un sottospazio e, per $h_1,h_2\in\mathrm{Sym}(A,j)$,

$$\overline{h_1h_2+h_2h_1}=\bar h_2\bar h_1+\bar h_1\bar h_2=h_2h_1+h_1h_2.$$

Il prodotto di Jordan è commutativo:

$$x\cdot y=y\cdot x. \tag{62}$$

Vi sono perciò solo tre prodotti distinti di tre elementi, $(x\cdot y)\cdot z$, $(y\cdot z)\cdot x$ e $(z\cdot x)\cdot y$. Inoltre il calcolo diretto dà $4(x\cdot y)\cdot z=xyz+yxz+zxy+zyx$. Quindi

$$4((x\cdot y)\cdot z-x\cdot(y\cdot z))=[y,[x,z]], \tag{63}$$

e, come per il commutatore, il prodotto di Jordan non è in generale associativo. La (63) e l'identità di Jacobi per i commutatori danno

$$(x \cdot y) \cdot z - x \cdot (y \cdot z) + (y \cdot z) \cdot x - y \cdot (z \cdot x) + (z \cdot x) \cdot y - z \cdot (x \cdot y) = 0.$$

Si tratta però di una conseguenza immediata della commutatività. Cerchiamo identità che non ne discendano in questo modo. Osserviamo dapprima che $x^{\cdot 2} = x \cdot x = \frac{1}{2}(x^2 + x^2) = x^2$. Se definiamo $x^{\cdot k} = x^{\cdot(k-1)} \cdot x$, $x^{\cdot 1} = x$, per induzione $x^{\cdot k} = x^k$. Ne segue

$$x^{\cdot k} \cdot x^{\cdot l} = x^{\cdot(k+l)}, \tag{64}$$

proprietà detta *associatività delle potenze*. Calcoliamo poi

$$2x^{\cdot 2} \cdot y = x^2 y + y x^2, \qquad 4x \cdot (x \cdot y) = x^2 y + 2xyx + y x^2.$$

Prendendo $x = e_{12} + e_{21}$, $y = e_{11}$ in $A = M_2(F)$, risulta $x^2 y + y x^2 = 2e_{11}$ e $xyx = e_{22}$, per cui $x^{\cdot 2} \cdot y$ e $x \cdot (x \cdot y)$ sono linearmente indipendenti. Non vi è dunque una relazione della forma $ax^{\cdot 2} \cdot y = bx \cdot (x \cdot y)$, con $a, b \in F$ non nulli, valida in ogni A^+. D'altra parte,

$$4(x^{\cdot 2} \cdot y) \cdot x = x^3 y + xyx^2 + x^2 yx + y x^3,$$
$$4x^{\cdot 2} \cdot (y \cdot x) = x^2 yx + x^3 y + y x^3 + xyx^2,$$

sicché in ogni A^+ vale l'*identità di Jordan*

$$x^{\cdot 2} \cdot (y \cdot x) = (x^{\cdot 2} \cdot y) \cdot x. \tag{65}$$

Riprendiamo ora da capo, dando le definizioni formali di algebra non associativa (ossia non necessariamente associativa), di algebra di Lie e di algebra di Jordan.

Definizione 7.2. *Un'algebra non associativa A su un campo F è uno spazio vettoriale munito di un prodotto binario $(x, y) \mapsto xy$ bilineare nel senso seguente:*

$$(x_1 + x_2)y = x_1 y + x_2 y, \qquad x(y_1 + y_2) = xy_1 + xy_2, \qquad a(xy) = (ax)y = x(ay)$$

per $x, x_i, y, y_i \in A$ e $a \in F$.

A differenza di quanto facciamo per le algebre associative non supponiamo l'esistenza di un'unità. Il motivo principale è che un'unità non può esistere nel caso particolare più importante, quello delle algebre di Lie non nulle. La definizione è la seguente.

Definizione 7.3. *Un'algebra di Lie è un'algebra non associativa il cui prodotto, denotato con $[xy]$ o $[x, y]$, soddisfa le leggi*

$$[x, x] = 0, \qquad [[x, y], z] + [[y, z], x] + [[z, x], y] = 0.$$

Una conseguenza immediata della prima è l'*anticommutatività*, $[x, y] = -[y, x]$. Viceversa l'anticommutatività implica $2[x, x] = 0$. Quindi, se la caratteristica del campo di base è diversa da 2, l'anticommutatività equivale a $[x, x] = 0$ e può sostituirla nella definizione.

Il risultato ottenuto sopra, nelle (58)–(60), è che a ogni algebra associativa A corrisponde un'algebra di Lie A^- sullo stesso spazio vettoriale, con prodotto $[x, y] = xy - yx$. Inoltre, se A ha un'involuzione j, gli elementi antisimmetrici formano una sottoalgebra $\mathrm{Sk}(A, j)$, nel senso evidente, dell'algebra di Lie A^-. Naturalmente questa in generale non è una sottoalgebra di A.

Vediamo un'altra origine delle algebre di Lie: le algebre delle derivazioni. Sia A un'algebra non associativa qualsiasi. Una *derivazione* di A è un'applicazione lineare $D\colon A \to A$ tale che

$$D(xy) = (Dx)y + x(Dy), \qquad x, y \in A.$$

Se D_1, D_2 hanno questa proprietà, è chiaro che la loro somma la possiede. Inoltre, per $a \in F$,

$$\begin{aligned}(aD)(xy) &= a(D(xy)) = a[(Dx)y + x(Dy)] \\ &= (a(Dx))y + x(a(Dy)) = ((aD)x)y + x((aD)y).\end{aligned}$$

L'insieme $\operatorname{Der} A$ delle derivazioni è quindi un sottospazio dello spazio vettoriale $\operatorname{End}_F A$ delle trasformazioni lineari di A su F. Se D_1, D_2 sono derivazioni, la composta è una trasformazione lineare e

$$\begin{aligned}D_1D_2(xy) &= D_1((D_2x)y + x(D_2y)) \\ &= (D_1D_2x)y + (D_1x)(D_2y) + (D_2x)(D_1y) + x(D_1D_2y).\end{aligned}$$

La composta può dunque non essere una derivazione. Tuttavia i due termini centrali hanno somma simmetrica in D_1, D_2; scambiando questi ultimi e sottraendo, tale somma si annulla. Pertanto

$$[D_1, D_2](xy) = ([D_1, D_2]x)y + x([D_1, D_2]y).$$

Si ha dunque $[D_1, D_2] \in \operatorname{Der} A$. Abbiamo così mostrato che $\operatorname{Der} A$ è una sottoalgebra dell'algebra di Lie $(\operatorname{End}_F A)^-$ delle trasformazioni lineari dello spazio vettoriale A. La chiamiamo *algebra delle derivazioni* di A.

Definizione 7.4. *Un'algebra di Jordan è un'algebra non associativa su un campo di caratteristica diversa da 2 il cui prodotto, denotato con $x \cdot y$, soddisfa*

$$x \cdot y = y \cdot x, \qquad (x^2 \cdot y) \cdot x = x^2 \cdot (y \cdot x), \qquad x^2 = x \cdot x. \tag{66}$$

Nella discussione preliminare abbiamo mostrato che ogni algebra associativa A su un campo di caratteristica diversa da 2 determina un'algebra di Jordan A^+ con lo stesso spazio vettoriale e il prodotto $x \cdot y = \frac{1}{2}(xy + yx)$. Se A ha un'involuzione j, gli elementi simmetrici rispetto a j formano una sottoalgebra $\operatorname{Sym}(A, j)$ di A^+.

Mostreremo ora che anche l'associatività delle potenze, osservata per A^+ nella (64), consegue dalla definizione e vale dunque in ogni algebra di Jordan. In un'algebra non associativa definiamo l'*associatore* di x, y, z ponendo

$$[x, y, z] = (xy)z - x(yz). \tag{67}$$

È additivo in ciascun argomento: $[x_1+x_2, y, z] = [x_1, y, z]+[x_2, y, z]$, e analogamente per gli altri. Inoltre $a[x, y, z] = [ax, y, z] = [x, ay, z] = [x, y, az]$ per $a \in F$. Queste proprietà si esprimono dicendo che l'associatore è una funzione *trilineare* dei suoi argomenti. L'ultima condizione che definisce un'algebra di Jordan equivale a

$$[x^2, y, x] = 0. \tag{68}$$

Questa condizione è cubica in x. Ne dedurremo un'identità multilineare per *linearizzazione* o *polarizzazione*, procedimento che risale all'antichità. Si può procedere in più modi. Il più diretto, benché forse non il più breve, consiste nel calcolare

$$\begin{aligned}0 &= [(x_1 + x_2)^2, y, x_1 + x_2] = [x_1^2 + 2x_1 \cdot x_2 + x_2^2, y, x_1 + x_2] \\ &= [x_1^2, y, x_1] + [x_1^2, y, x_2] + 2[x_1 \cdot x_2, y, x_1] + 2[x_1 \cdot x_2, y, x_2] \\ &\quad + [x_2^2, y, x_1] + [x_2^2, y, x_2] \\ &= [x_1^2, y, x_2] + 2[x_1 \cdot x_2, y, x_1] + 2[x_1 \cdot x_2, y, x_2] + [x_2^2, y, x_1].\end{aligned}$$

Sostituendo $x_2 + x_3$ a x_2, nell'ultima espressione otteniamo

$$\begin{aligned} 0 = {} & [x_1^2, y, x_2] + [x_1^2, y, x_3] + 2[x_1 \cdot x_2, y, x_1] + 2[x_1 \cdot x_3, y, x_1] \\ & + 2[x_1 \cdot x_2, y, x_2] + 2[x_1 \cdot x_3, y, x_2] + 2[x_1 \cdot x_2, y, x_3] + 2[x_1 \cdot x_3, y, x_3] \\ & + [x_2^2, y, x_1] + 2[x_2 \cdot x_3, y, x_1] + [x_3^2, y, x_1]. \end{aligned}$$

Sottraiamo la relazione precedente e quella ottenuta sostituendovi x_3 a x_2. Rimane l'identità multilineare cercata:

$$2[x_1 \cdot x_2, y, x_3] + 2[x_2 \cdot x_3, y, x_1] + 2[x_3 \cdot x_1, y, x_2] = 0.$$

Poiché la caratteristica è diversa da 2, possiamo cancellare 2 e ottenere

$$[x_1 \cdot x_2, y, x_3] + [x_2 \cdot x_3, y, x_1] + [x_3 \cdot x_1, y, x_2] = 0. \tag{69}$$

In un'algebra non associativa A denotiamo con x_R l'applicazione lineare $y \mapsto yx$, moltiplicazione a destra per x, e con x_L l'applicazione $y \mapsto xy$, moltiplicazione a sinistra per x. La legge commutativa si esprime con $x_L = x_R$ per ogni x. Usandola, l'identità di Jordan (68) diventa

$$(x^2)_L x_L = x_L (x^2)_L, \qquad \text{ossia } [(x^2)_L, x_L] = 0, \qquad x^2 = x \cdot x. \tag{68$'$}$$

La (69) equivale a

$$[(x_1 \cdot x_2)_L, x_{3L}] + [(x_2 \cdot x_3)_L, x_{1L}] + [(x_3 \cdot x_1)_L, x_{2L}] = 0. \tag{69$'$}$$

Possiamo ricavare un'altra identità di operatori equivalente, scegliendo uno degli x_i come elemento su cui agire. Esplicitamente,

$$\begin{aligned} & ((x_1 \cdot x_2) \cdot y) \cdot x_3 + ((x_2 \cdot x_3) \cdot y) \cdot x_1 + ((x_3 \cdot x_1) \cdot y) \cdot x_2 \\ & = (x_1 \cdot x_2) \cdot (y \cdot x_3) + (x_2 \cdot x_3) \cdot (y \cdot x_1) + (x_3 \cdot x_1) \cdot (y \cdot x_2). \end{aligned}$$

Scambiando y e x_2 otteniamo

$$\begin{aligned} & x_{1L} x_{2L} x_{3L} + x_{3L} x_{2L} x_{1L} + ((x_1 \cdot x_3) \cdot x_2)_L \\ & = x_{1L} (x_2 \cdot x_3)_L + x_{2L} (x_3 \cdot x_1)_L + x_{3L} (x_1 \cdot x_2)_L. \end{aligned} \tag{70}$$

Definiamo ora le potenze x^k (o $x^{\cdot k}$) per mezzo di $x^1 = x$, $x^k = x \cdot x^{k-1}$. La (70) dà la ricorrenza

$$(x^{k+2})_L = (x^k)_L (x^2)_L + 2x_L (x^{k+1})_L - (x_L)^2 (x^k)_L - (x^k)_L (x_L)^2. \tag{71}$$

Poiché x_L e $(x^2)_L$ commutano, generano un'algebra commutativa X di trasformazioni lineari. Per la ricorrenza tutti gli $(x^k)_L$, $k \geq 1$, appartengono a X. Dunque

$$[(x^k)_L, (x^l)_L] = 0, \qquad k, l \geq 1, \tag{72}$$

che equivale a

$$[x^k, y, x^l] = 0. \tag{72$'$}$$

Ne segue l'associatività delle potenze, $x^k \cdot x^l = x^{k+l}$. Infatti essa vale per definizione per $k = 1$ e ogni l. Supponendola valida per k fissato e ogni l, si ha

$$x^{k+1} \cdot x^l = (x \cdot x^k) \cdot x^l = x \cdot (x^k \cdot x^l) = x \cdot x^{k+l} = x^{k+l+1}.$$

Esercizi

1. Verificare che in ogni algebra non associativa vale l'identità fra associatori

$$v[x,y,z] + [v,x,y]z = [vx,y,z] - [v,xy,z] + [v,x,yz].$$

2. In un'algebra non associativa A si chiama *nucleo* $N(A)$ l'insieme degli elementi v che si associano con ogni coppia di elementi, ossia tali che ogni associatore in cui compare v è nullo. Usare l'esercizio 1 per mostrare che $N(A)$ è una sottoalgebra associativa di A.

3. Il *centro* $C(A)$ di un'algebra non associativa è l'insieme dei $c \in N(A)$ tali che $cx = xc$ per ogni $x \in A$. Mostrare che è una sottoalgebra associativa commutativa di A.

4. Sia D una derivazione di $F[\lambda]$ su F, con λ indeterminata, e sia $D\lambda = f(\lambda)$. Mostrare che $f(\lambda)$ determina D e che $D \mapsto f(\lambda)$ è un isomorfismo di spazi vettoriali di $\operatorname{Der} F[\lambda]$ su $F[\lambda]$. Se $D \mapsto f(\lambda)$ ed $E \mapsto g(\lambda)$, mostrare che $[D,E] \mapsto f(\lambda)g'(\lambda) - f'(\lambda)g(\lambda)$, dove l'apice indica la derivata formale (§ 4.4, p. 230).

5. Generalizzare l'esercizio 4 a $F[\lambda_1, \lambda_2, \ldots, \lambda_r]$, con le λ_i indeterminate.

6. Sia A un'algebra associativa su F che è un dominio commutativo. Mostrare che ogni derivazione di A si estende in un solo modo al suo campo delle frazioni.

7. Sia A un'estensione separabile di dimensione finita di F. Mostrare che $\operatorname{Der} A = 0$.

8. Determinare $\operatorname{Der} A$ se F ha caratteristica $p \neq 0$ e $A = F[\lambda]/(\lambda^p - a)$, $a \in F$.

9. Se A è un'algebra, l'insieme $M_n(A)$ delle matrici $n \times n$ a coefficienti in A è un'algebra con le usuali operazioni di spazio vettoriale e il prodotto matriciale usuale. Per un'applicazione D di A in sé, sia $\eta(D)$ l'applicazione di A in $M_2(A)$ definita da

$$x \longmapsto \begin{pmatrix} x & Dx \\ 0 & x \end{pmatrix}.$$

   Mostrare che $\eta(D)$ è un omomorfismo se e solo se D è una derivazione.

10. Mostrare che una trasformazione lineare D di un'algebra non associativa A è una derivazione se e solo se vale una delle due condizioni

$$[D, x_L] = (Dx)_L \quad (x \in A), \qquad [D, x_R] = (Dx)_R \quad (x \in A).$$

    Mostrare che, se A è associativa, $x_R - x_L$ è una derivazione per ogni $x \in A$; se A è di Lie, $x_L = -x_R$ è una derivazione; se A è di Jordan, $[x_L, y_L] = [x_R, y_R]$ è una derivazione per ogni $x, y \in A$.

11. Sia $B(u,v)$ una forma bilineare simmetrica su uno spazio vettoriale V su F, di caratteristica diversa da 2. Sia $A = F1 \oplus V$ la somma diretta di V e di uno spazio unidimensionale con base 1. Per $x = a1 + u$, $y = b1 + v$, $a, b \in F$, $u, v \in V$, definiamo il prodotto

$$(a1 + u) \cdot (b1 + v) = (ab + B(u,v))1 + (av + bu).$$

    Verificare che A con questo prodotto è un'algebra di Jordan con unità 1.

12. Sia $E(V)$ l'algebra esterna su V. Mostrare che $F1 + V$ è una sottoalgebra di $E(V)^+$. Mostrare che, quando $B = 0$ nell'esercizio 11, l'algebra di Jordan ottenuta è isomorfa a questa sottoalgebra.

7.6 Il problema di Hurwitz. Algebre di composizione

Nel 1898 A. Hurwitz considerò il seguente problema. Per quali valori di n esistono identità della forma

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) = \sum_{i=1}^{n} z_i^2, \tag{73}$$

dove gli z_i hanno la forma

$$z_i = \sum_{j,k=1}^{n} a_{ijk} x_j y_k, \tag{74}$$

con a_{ijk} numeri complessi? Quando Hurwitz propose e risolse il problema, si conoscevano già varie identità di questo tipo e vi erano stati parecchi tentativi infruttuosi di trovarne altre. Le identità note corrispondevano a $n = 1, 2, 4, 8$. La prima è banale: $x_1^2 y_1^2 = (x_1 y_1)^2$. Le due successive sono già non banali:

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1 y_1 - x_2 y_2)^2 + (x_1 y_2 + x_2 y_1)^2,$$
$$(x_1^2 + x_2^2 + x_3^2 + x_4^2)(y_1^2 + y_2^2 + y_3^2 + y_4^2) = z_1^2 + z_2^2 + z_3^2 + z_4^2,$$

dove

$$\begin{aligned} z_1 &= x_1 y_1 - x_2 y_2 - x_3 y_3 - x_4 y_4, \\ z_2 &= x_1 y_2 + x_2 y_1 + x_3 y_4 - x_4 y_3, \\ z_3 &= x_1 y_3 - x_2 y_4 + x_3 y_1 + x_4 y_2, \\ z_4 &= x_1 y_4 + x_2 y_3 - x_3 y_2 + x_4 y_1. \end{aligned}$$

Si possono verificare direttamente o, meglio, dedurre dalle proprietà della moltiplicazione dei numeri complessi e dei quaternioni (si veda l'esercizio 3 a p. 100 dell'originale). È piuttosto laborioso scrivere l'identità corrispondente per $n = 8$. Più avanti ne daremo una forma meno esplicita da cui, volendo, si potrebbe ricavare quella esplicita. Non si sa chi abbia scoperto per primo l'identità per $n = 2$. Quella per $n = 4$ sembra dovuta a Eulero e, secondo L. E. Dickson, quella per $n = 8$ fu trovata da C. F. Degen nel 1822. L'identità sulle somme di quadrati per $n = 4$ ha un ruolo importante nella dimostrazione di un bel teorema di Lagrange: ogni intero positivo si può esprimere come somma di quattro quadrati di interi.[9] Il teorema di Hurwitz, che dimostreremo e generalizzeremo in questo paragrafo, afferma che esistono identità della forma (73)–(74) soltanto per $n = 1, 2, 4, 8$.

Il problema di Hurwitz si può considerare da un punto di vista formale o funzionale. Nel primo caso gli x_i e gli y_i sono indeterminate, e (73) è una relazione nell'anello dei polinomi in tali indeterminate su $\mathbb{C}$. Nel secondo il punto di partenza è la funzione $(x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} x_i^2$ definita sullo spazio vettoriale di dimensione n delle n-uple di elementi di $\mathbb{C}$. La funzione è una forma quadratica e (73)–(74) è una relazione funzionale. La soluzione del problema da ciascuno dei due punti di vista implica la soluzione dall'altro. L'implicazione dal formale al funzionale è immediata. L'implicazione inversa segue nel modo consueto dal Teorema 2.19, poiché $\mathbb{C}$ è un campo infinito. Adotteremo il punto di vista funzionale. Consideriamo pertanto lo spazio vettoriale $\mathbb{C}^{(n)}$ delle n-uple $x = (x_1, \ldots, x_n)$ di numeri complessi, con la forma quadratica non degenere $x \mapsto \sum x_i^2$. Per $y = (y_1, \ldots, y_n)$ e $z = (z_1, \ldots, z_n)$ si ha la composizione binaria $(x, y) \mapsto z$, dove le coordinate z_i sono date da (74) per numeri complessi fissati a_{ijk}. Dalla forma

[9] Si veda G. H. Hardy ed E. M. Wright, *An Introduction to the Theory of Numbers*, 5ª ed., Oxford University Press, 1975, p. 302.

di (74) è chiaro che la composizione è bilineare. Indicando z con xy, otteniamo un'algebra non associativa. Se scriviamo $Q(x) = \sum x_i^2$, si ha $Q(x)Q(y) = Q(xy)$.

Precisiamo e generalizziamo ora il problema di Hurwitz. Supponiamo dato uno spazio vettoriale A di dimensione finita su un campo F di caratteristica diversa da 2, dotato di una forma quadratica non degenere Q.[10] Diremo che Q *ammette composizione* se è possibile definire in A un prodotto bilineare xy tale che

$$Q(x)Q(y) = Q(xy) \qquad (x, y \in A). \tag{75}$$

Lo spazio vettoriale e il prodotto xy definiscono un'algebra non associativa. Mostriamo che, modificando il prodotto, possiamo supporre che l'algebra abbia un'unità. Scegliamo v con $Q(v) \neq 0$ e poniamo $u = Q(v)^{-1}v^2$. Allora $Q(u) = 1$, quindi $Q(xu) = Q(x) = Q(ux)$ per ogni x. Le moltiplicazioni u_R e u_L sono trasformazioni ortogonali di A rispetto a Q e pertanto sono invertibili, con inverse ortogonali. Definiamo un nuovo prodotto in A mediante

$$x * y = (u_R^{-1}x)(u_L^{-1}y).$$

Abbiamo $Q(x * y) = Q(u_R^{-1}x)Q(u_L^{-1}y) = Q(x)Q(y)$. Inoltre $u_L^{-1}u^2 = u_L^{-1}(u_L u) = u$ e $u_R^{-1}u^2 = u_R^{-1}(u_R u) = u$, per cui

$$u^2 * x = (u_R^{-1}u^2)(u_L^{-1}x) = u(u_L^{-1}x) = x, \qquad x * u^2 = (u_R^{-1}x)(u_L^{-1}u^2) = (u_R^{-1}x)u = x.$$

Per la moltiplicazione $*$ l'elemento u^2 è dunque un'unità. Torniamo alla notazione xy per $x * y$ e supponiamo sin dall'inizio che A abbia un'unità denotata con 1.

Definizione 7.5. *Un'algebra di composizione è una coppia formata da un'algebra non associativa A con unità 1 e da una forma quadratica non degenere Q su A tale che $Q(xy) = Q(x)Q(y)$.*

Il nostro obiettivo è determinare tutte le algebre di composizione. Può sembrare un programma troppo ambizioso, ma vedremo che lo si realizza in modo sorprendentemente elementare.

Sia (A, Q) un'algebra di composizione. Osserviamo anzitutto che la legge (75) dà $Q(x)Q(1) = Q(x)$, da cui

$$Q(1) = 1. \tag{76}$$

Linearizziamo la legge di composizione rispetto a x sostituendo $x + z$ a x. Otteniamo $Q(x + z)Q(y) = Q((x + z)y) = Q(xy + zy)$. Poiché $Q(x + y) - Q(x) - Q(y) = B(x, y)$ è la forma bilineare simmetrica associata a Q, risulta

$$\begin{aligned} Q(x)Q(y) + B(x, z)Q(y) + Q(z)Q(y) &= Q(xy) + B(xy, zy) + Q(zy) \\ &= Q(x)Q(y) + B(xy, zy) + Q(z)Q(y). \end{aligned}$$

Pertanto

$$B(x, z)Q(y) = B(xy, zy). \tag{77}$$

Linearizzando analogamente rispetto a y otteniamo

$$Q(x)B(y, w) = B(xy, xw). \tag{78}$$

[10] Né la dimensione finita né la caratteristica diversa da 2 sono restrizioni essenziali, come si vedrà negli esercizi.

Linearizziamo poi (77) rispetto a y sostituendovi $y+w$ a y. Ne segue

$$B(x,z)B(y,w) = B(xy,zw) + B(zy,xw). \tag{79}$$

Poiché il primo membro resta invariato scambiando x con y e z con w, abbiamo anche

$$B(x,z)B(y,w) = B(yx,wz) + B(yz,wx). \tag{80}$$

Introduciamo $j = -S_1$, dove S_1 è la simmetria rispetto all'iperpiano ortogonale a 1:

$$j\colon x \longmapsto \frac{B(x,1)}{Q(1)}1 - x = B(x,1)1 - x.$$

Scriviamo $\bar{x} = j(x)$ e $T(x) = B(x,1)$. Allora

$$Q(\bar{x}) = Q(x), \qquad \bar{\bar{x}} = x. \tag{81}$$

Dimostriamo il seguente lemma.

Lemma 1. *Valgono le proprietà*

$$\bar{x}x = Q(x)1 = x\bar{x}, \tag{82}$$

$$\bar{x}(xy) = (\bar{x}x)y = Q(x)y, \tag{83}$$

$$(yx)\bar{x} = y(x\bar{x}) = Q(x)y, \tag{84}$$

$$\overline{xy} = \bar{y}\bar{x}. \tag{85}$$

Dimostrazione. Osserviamo anzitutto che

$$\begin{aligned} B(x,\bar{y}z) &= B(x, B(y,1)z - yz) = B(y,1)B(x,z) - B(x,yz) \\ &= B(yx,z) + B(yz,x) - B(x,yz) = B(yx,z), \end{aligned}$$

dove si è usata (79). Abbiamo dunque

$$B(yz,x) = B(z,\bar{y}x), \tag{86}$$

e analogamente

$$B(xy,z) = B(x,z\bar{y}). \tag{87}$$

Per (78) e (86), $Q(x)B(1,y) = B(x,xy) = B(\bar{x}x,y)$. Pertanto $B(Q(x)1,y) = B(\bar{x}x,y)$, ossia $B(Q(x)1 - \bar{x}x, y) = 0$ per ogni y. Poiché B è non degenere, $\bar{x}x = Q(x)1$. Sostituendo x con $\bar{x}$ si ottiene anche $x\bar{x} = \bar{\bar{x}}\bar{x} = Q(\bar{x})1 = Q(x)1$. È così provata (82). Inoltre

$$\begin{aligned} B(\bar{x}(xy),z) &= B((B(x,1)1 - x)(xy), z) = B(x,1)B(xy,z) - B(x(xy),z) \\ &= B(x(xy),z) + B(xy,xz) - B(x(xy),z) \\ &= B(xy,xz) = Q(x)B(y,z) = B(Q(x)y,z), \end{aligned}$$

ancora per (79). Per la non degenerazione di B segue $\bar{x}(xy) = Q(x)y$. Poiché $\bar{x}x = Q(x)1$, si ottiene (83). In modo analogo si prova (84). Per dimostrare (85) calcoliamo

$$\begin{aligned} B(\overline{xy},z) &= B(\overline{xy}\,1,z) = B(1,(xy)z), \\ B(\bar{y}\bar{x},z) &= B((B(y,1)1 - y)(B(x,1)1 - x), z) \\ &= B(B(x,1)B(y,1)1 - B(y,1)x - B(x,1)y + yx, z) \\ &= B(x,1)B(y,1)B(z,1) - B(y,1)B(x,z) - B(x,1)B(y,z) + B(yx,z) \\ &= B(xy,1)B(z,1) + B(x,y)B(z,1) - B(yz,x) - B(xy,z) - B(xz,y) \\ &= B((xy)z,1) + B(xy,z) + B(xz,y) + B(yz,x) \\ &\quad - B(yz,x) - B(xy,z) - B(xz,y) \\ &= B((xy)z,1). \end{aligned}$$

Si ha dunque $B(\overline{xy}, z) = B(\bar{y}\bar{x}, z)$ e ancora per la non degenerazione segue (85). □

L'applicazione $j\colon x \mapsto \bar{x}$ è lineare, $\bar{\bar{x}} = x$ e vale (85). Dunque j è un'involuzione di A. Inoltre per definizione $x + \bar{x} = T(x)1$. Le relazioni (83) e (84) danno le seguenti relazioni fra associatori:

$$[\bar{x}, x, y] = 0 = [y, x, \bar{x}]. \tag{88}$$

Poiché $[1, x, y] = 0 = [y, x, 1]$ e $x = T(x)1 - \bar{x}$, ne segue

$$[x, x, y] = 0 = [y, x, x]. \tag{89}$$

Dunque A è un'algebra alternativa nel senso della definizione seguente.

Definizione 7.6. *Un'algebra si dice alternativa se le identità (89) valgono per ogni coppia di suoi elementi x, y.*

Abbiamo mostrato che, se (A, Q) è un'algebra di composizione, A è alternativa con involuzione $j\colon x \mapsto \bar{x}$ tale che $\bar{x}x = Q(x)1$. Queste condizioni sono anche sufficienti. Per dimostrarlo dobbiamo prima ricavare alcune proprietà fondamentali delle algebre alternative.

Sia ora A una qualsiasi algebra alternativa. La linearizzazione delle leggi (89) dà $[x, z, y] + [z, x, y] = 0 = [y, x, z] + [y, z, x]$. Quindi l'associatore $[x, y, z]$ è una funzione alternante dei suoi argomenti: rimane invariato rispetto alle permutazioni pari, ossia $[x, y, z] = [y, z, x] = [z, x, y]$, e cambia segno rispetto a quelle dispari. Inoltre $[x, y, x] = -[y, x, x] = 0$. Abbiamo pertanto le leggi

$$x^2y = x(xy), \qquad (xy)x = x(yx), \qquad yx^2 = (yx)x. \tag{90}$$

Scriveremo xyx per $(xy)x = x(yx)$. Dimostriamo ora la seguente importante identità per le algebre alternative, dovuta a R. Moufang:

$$(ux)(yu) = u(xy)u. \tag{91}$$

Partiamo dalla relazione

$$(ux)y + x(yu) = u(xy) + (xy)u, \tag{92}$$

equivalente a $[u, x, y] = [x, y, u]$. Sostituiamo nella (92) prima ux a x, poi, nella relazione originale, yu a y, e sommiamo:

$$\begin{aligned}(u^2x)y + 2(ux)(yu) + x(yu^2) &= u((ux)y) + ((ux)y)u + u(x(yu)) + (x(yu))u \\ &= u[(ux)y + x(yu)] + [(ux)y + x(yu)]u \\ &= u[u(xy) + (xy)u] + [u(xy) + (xy)u]u \quad \text{per (92)} \\ &= u^2(xy) + 2u(xy)u + (xy)u^2.\end{aligned}$$

Sottraendo da questa la relazione che si ottiene da (92) sostituendovi u^2 a u, otteniamo l'identità di Moufang.[11]

Supponiamo ora che A sia un'algebra alternativa con 1 e involuzione $j\colon x \mapsto \bar{x}$ tale che $\bar{x}x = Q(x)1$, dove Q è una forma quadratica non degenere. Per linearizzazione abbiamo

$$\bar{x}y + \bar{y}x = Q(x, y)1, \tag{93}$$

[11] La dimostrazione data usa l'ipotesi che la caratteristica sia diversa da 2. Il risultato vale anche senza tale restrizione.

dove $Q(x,y) = B(x,y)$ indica la forma polare di Q. Ponendo $y = 1$ otteniamo $x + \bar{x} = T(x)1$, con $T(x) = Q(x,1)$. Le leggi alternative $[x,x,y] = 0 = [y,x,x]$ e $x + \bar{x} = T(x)1$ danno (88), sicché $\bar{x}(xy) = (\bar{x}x)y = Q(x)y$. Ora

$$\begin{aligned} Q(xy)1 = \overline{xy}(xy) &= (\bar{y}\bar{x})(xy) = [(T(y)1 - y)\bar{x}](xy) \\ &= (T(y)\bar{x} - y\bar{x})(xy) = T(y)\bar{x}(xy) - (y\bar{x})(xy) \\ &= Q(x)T(y)y - y(\bar{x}x)y \qquad \text{(Moufang)} \\ &= Q(x)[T(y)1 - y]y = Q(x)(\bar{y}y) = Q(x)Q(y)1. \end{aligned}$$

Pertanto $Q(xy) = Q(x)Q(y)$ e (A,Q) è un'algebra di composizione. Abbiamo ottenuto il primo risultato importante della nostra analisi.

Teorema 7.5. *Ogni algebra di composizione (A,Q) è alternativa e possiede un'involuzione $j\colon x \mapsto \bar{x}$ tale che $\bar{x}x = Q(x)1$. Viceversa, se A è un'algebra alternativa con unità e involuzione $j\colon x \mapsto \bar{x}$ tale che $\bar{x}x = Q(x)1$, con Q forma quadratica non degenere, allora (A,Q) è un'algebra di composizione.*

Diamo ora una costruzione di algebre di composizione. Sarà una generalizzazione quasi immediata della consueta costruzione dei numeri complessi come coppie di numeri reali. Per il momento tralasciamo la legge alternativa e supponiamo soltanto che A sia un'algebra non associativa con unità 1 e involuzione j tale che $\bar{x}x = Q(x)1$, con Q forma quadratica non degenere. Valgono allora (93) e $x + \bar{x} = T(x)1$, dove $T(x) = B(x,1)$. Sia c un elemento non nullo del campo base F. Da A, j, c costruiremo un'algebra D che soddisfa le stesse condizioni di A e ha dimensione $2 \dim A$. Sia $D = A^{(2)}$ lo spazio vettoriale delle coppie (x,y), $x,y \in A$, con l'usuale struttura di somma diretta. Introduciamo in D un prodotto binario mediante

$$(u,v)(x,y) = (ux + c\bar{y}v,\ yu + v\bar{x}). \tag{94}$$

Esso è evidentemente bilineare e definisce con la struttura di spazio vettoriale un'algebra su D. La (94) mostra che $(1,0)$ è un'unità in D e poniamo $1 = (1,0)$. Inoltre $(u,0)(x,0) = (ux,0)$, per cui $u \mapsto (u,0)$ è un monomorfismo di A in D. Possiamo quindi identificare A con la sottoalgebra di D formata dagli elementi $(u,0)$, $u \in A$. Estendiamo l'involuzione j di A all'applicazione lineare

$$j\colon (x,y) \longmapsto \overline{(x,y)} = (\bar{x}, -y). \tag{95}$$

Chiaramente $j^2 = 1$. Una verifica diretta, lasciata al lettore, mostra che j è un'involuzione di D. Inoltre

$$\overline{(x,y)}(x,y) = (\bar{x}, -y)(x,y) = ((Q(x) - cQ(y))1, 0) = (Q(x) - cQ(y))1.$$

Ora $(x,y) \mapsto Q(x) - cQ(y)$ è una forma quadratica su D. La forma bilineare simmetrica corrispondente è $((u,v),(x,y)) \mapsto B(u,x) - cB(v,y)$, che si vede facilmente essere non degenere. Dunque D e la sua involuzione soddisfano le stesse condizioni di A. Chiameremo D il *c-doppio* di A.

Lemma 2. *(1) Il c-doppio D è commutativo e associativo se e solo se A è commutativa e associativa e $j = 1$. (2) D è associativo se e solo se A è commutativa e associativa. (3) D è alternativo se e solo se A è associativa.*

Dimostrazione. Scriviamo $X = (x, y)$, $U = (u, v)$, $Z = (z, t)$ con $x, y, \ldots$ in A. Allora

$$[U, X] = ([u, x] + c(\bar{y}v - \bar{v}y),\ y(u - \bar{u}) + v(\bar{x} - x)). \tag{96}$$

$$\begin{aligned}[U, X, Z] = (&[u, x, z] + c\{\bar{t}(yu) - u(\bar{t}y) + \bar{t}(v\bar{x}) - (\bar{x}\bar{t})v + (\bar{y}v)z - (z\bar{y})v\},\\ &t(ux) - (tx)u + (yu)\bar{z} - (y\bar{z})u + (v\bar{x})\bar{z} - v(\bar{z}\bar{x}) + c\{t(\bar{y}v) - v(\bar{y}t)\}).\end{aligned} \tag{97}$$

Poiché A è una sottoalgebra di D mediante l'identificazione $x \mapsto (x, 0)$, la commutatività o l'associatività di D implica rispettivamente la commutatività o l'associatività di A. Per la (96) con $u = x = 0$ e $v = 1$, $[U, X] = 0$ implica $\bar{y} = y$, dunque $j = 1$. Viceversa, dalla (96) risulta che, se A è associativa e commutativa e $j = 1$, allora D è commutativo. Ponendo inoltre $v = x = z = 0$, $t = 1$, nella (97) otteniamo, come condizione necessaria per l'associatività di D, la relazione $yu = uy$ per $y, u \in A$. Quindi, se D è associativo, A è associativa e commutativa. Viceversa, la (97) mostra che se A è associativa e commutativa D è associativo. Questo prova (1) e (2).

Per (3), osserviamo che D è alternativo se e solo se $[\bar{X}, X, Z] = 0$ per ogni X, Z. Infatti, poiché $X + \bar{X} = T(X)1$, tale condizione equivale a $[X, X, Z] = 0$. Applicandovi l'involuzione si ottiene $[\bar{Z}, \bar{X}, \bar{X}] = 0$, dato che

$$\overline{[X, Y, Z]} = \overline{(XY)Z} - \overline{X(YZ)} = \bar{Z}(\bar{Y}\bar{X}) - (\bar{Z}\bar{Y})\bar{X} = -[\bar{Z}, \bar{Y}, \bar{X}].$$

Segue $[Z, X, X] = 0$ per ogni Z, X. Supponiamo ora che A sia alternativa, condizione necessaria perché D lo sia. Ponendo $U = \bar{X} = (\bar{x}, -y)$ nella (97) e usando $[\bar{x}, x, y] = 0$, $(\bar{y}y)z = Q(y)z = (z\bar{y})y$, ecc., si ottiene

$$[\bar{X}, X, Z] = (c[\bar{x}, \bar{t}, y],\ -[y, \bar{z}, \bar{x}]).$$

Quindi $[\bar{X}, X, Z] = 0$ per ogni X, Z se e solo se A è associativa. Ciò prova (3). □

Le algebre considerate nel Lemma 2 sono algebre di composizione se e solo se sono alternative (Teorema 7.5). Ne ricaviamo la seguente successione di esempi. Partiamo da $A = F$, che soddisfa banalmente le condizioni. Il suo doppio è un'algebra commutativa e associativa di dimensione due, per il Lemma 2 (1). Queste algebre di composizione si chiameranno *algebre quadratiche*. Fra di esse vi sono le estensioni quadratiche del campo F. Il doppio di un'algebra quadratica è associativo ma non commutativo, perché l'involuzione nell'algebra quadratica non è l'identità. I doppi delle algebre quadratiche si chiamano *algebre di quaternioni (generalizzate)* su F. Hanno dimensione quattro su F. Raddoppiando ancora, otteniamo algebre alternative di dimensione otto. Queste algebre di composizione si chiamano *algebre di ottonioni* o *algebre di Cayley*. Poiché le algebre di quaternioni non sono commutative, quelle di ottonioni non sono associative. Per le algebre di composizione il procedimento termina qui.

Dimostreremo ora che queste costruzioni danno tutte le algebre di composizione. Ci occorre il seguente lemma.

Lemma 3. *Sia (A, Q) un'algebra di composizione, e sia C una sottoalgebra propria contenente 1, stabile per l'involuzione j di A ($\bar{C} \subseteq C$) e non degenere come sottospazio di A rispetto a B. Allora C è contenuta in una sottoalgebra D di A che soddisfa le stesse condizioni di stabilità e non degenerazione ed è isomorfa a un doppio di C.*

Dimostrazione. Essendo C non degenere, $A = C \oplus C^{\perp}$, e possiamo scegliere $t \in C^{\perp}$ tale che $Q(t) = -c \neq 0$. Poiché $1 \in C$, $T(t) = B(1, t) = 0$, quindi $\bar{t} = -t$ e

$$\text{(i)} \qquad t^2 = -\bar{t}t = -Q(t)1 = c1.$$

Per $x \in C$ si ha $B(x,t) = 0$. Allora per (93), $\bar{x}t + \bar{t}x = 0$, cioè

$$\text{(ii)} \qquad tx = \bar{x}t \qquad (x \in C).$$

Se $x, y \in C$, allora $\bar{y}x \in C$, dunque $B(x, yt) = B(\bar{y}x, t) = 0$. Il sottospazio $Ct = \{yt \mid y \in C\}$ è pertanto contenuto in $C^\perp$ e $D = C + Ct = C \oplus Ct$. In A vale la relazione $\bar{x}(xy) = Q(x)y$, che per linearizzazione dà

$$\bar{x}(zy) + \bar{z}(xy) = B(x,z)y. \tag{98}$$

Ponendo $x, y \in C$, $z = t$, otteniamo $\bar{x}(ty) = t(xy)$. Per (ii) segue $\bar{x}(\bar{y}t) = (\bar{y}\bar{x})t$. Abbiamo dunque

$$\text{(iii)} \qquad x(yt) = (yx)t \qquad (x, y \in C).$$

Applicando l'involuzione e usando (ii) otteniamo anche

$$\text{(iv)} \qquad (yt)x = (y\bar{x})t.$$

Infine $(xt)(yt) = (t\bar{x})(yt) = t(\bar{x}y)t$ per l'identità di Moufang, e quest'ultimo prodotto è $(\bar{y}x)t^2 = c\bar{y}x$. Pertanto

$$\text{(v)} \qquad (xt)(yt) = c\bar{y}x \qquad (x, y \in C).$$

Per (i)–(v), se $u, v, x, y \in C$, allora

$$\text{(vi)} \qquad (u + vt)(x + yt) = (ux + c\bar{y}v) + (yu + v\bar{x})t.$$

Quindi $D = C + Ct$ è una sottoalgebra di A contenente C. Inoltre $\overline{u + vt} = \bar{u} + \bar{t}\bar{v} = \bar{u} - t\bar{v} = \bar{u} - vt$, per cui D è stabile per j, e $Q(xt) = Q(x)Q(t) = -cQ(x)$. Quest'ultima relazione, polarizzata, mostra che $x \mapsto xt$ è una biiezione lineare di C su Ct. Dunque C e Ct sono isomorfi come spazi vettoriali. Inoltre Ct è non degenere, e tale è D, somma diretta ortogonale di C e Ct. Dal confronto di (vi) e (94) si vede che $(x, y) \mapsto x + yt$ è un isomorfismo del c-doppio di C su D. Ciò conclude la dimostrazione. □

Possiamo ora dimostrare il risultato principale.

Teorema generalizzato di Hurwitz. *La lista completa delle algebre di composizione su un campo F di caratteristica diversa da 2 è la seguente: (I) $F1$, (II) algebre quadratiche, (III) algebre di quaternioni, (IV) algebre di ottonioni.*

Dimostrazione. Abbiamo visto che tutte le algebre elencate sono algebre di composizione, con la forma Q definita nella costruzione. Sia ora (A, Q) un'algebra di composizione. Se $A = F1$ siamo nel caso I. Altrimenti $F1$ è propriamente contenuta in A, e per il Lemma 3 A contiene una sottoalgebra quadratica non degenere e stabile per j. Se A coincide con essa, abbiamo II. Altrimenti A contiene una sottoalgebra di quaternioni stabile per j e non degenere. Se A coincide con questa siamo in III. Altrimenti A contiene una sottoalgebra di ottonioni stabile per j e non degenere. Allora A deve coincidere con quest'ultima, perché in caso contrario conterrebbe un doppio di un'algebra di ottonioni. Un tale doppio non è alternativo, il che è impossibile poiché A è alternativa. Siamo dunque nel caso IV. □

Ricaviamo ora esplicitamente le basi e le tavole di moltiplicazione che il procedimento di raddoppiamento fornisce in modo naturale. Con esse si possono scrivere le leggi di composizione delle forme quadratiche.

In primo luogo abbiamo F con base $i_0 = 1$ e moltiplicazione $i_0^2 = i_0$. Sia A_1 il suo c_1-doppio. Per A_1 scegliamo la base $i_0 = 1$, $i_1 = (0, 1)$. Tralasciando i prodotti che coinvolgono i_0, la moltiplicazione è descritta da

$$i_1^2 = c_1 1. \tag{99}$$

Formiamo poi il c_2-doppio A_2 di A_1 e scriviamo $i_2 = (0, 1)$ in esso. Abbiamo la base $(i_0, i_1, i_2, i_3 = i_1 i_2)$ poiché $A_2 = A_1 \oplus A_1 i_2$. La parte essenziale della tavola di moltiplicazione rispetto a questa base dell'algebra di quaternioni A_2 è

$$\begin{gathered} i_1^2 = c_1 1, \qquad i_2^2 = c_2 1, \qquad i_3^2 = -c_1 c_2 1, \\ i_1 i_2 = i_3 = -i_2 i_1, \qquad i_2 i_3 = -c_2 i_1 = -i_3 i_2, \qquad i_3 i_1 = -c_1 i_2 = -i_1 i_3. \end{gathered} \tag{100}$$

Tutte queste formule seguono dalla legge associativa e dalle relazioni $i_1^2 = c_1 1$, $i_2^2 = c_2 1$, $i_1 i_2 = -i_2 i_1$, ossia da (i) e (ii) sopra. Consideriamo infine l'algebra di ottonioni A_3, c_3-doppio di A_2, che ha dunque la base

$$i_0 = 1,\ i_1,\ i_2,\ i_3 = i_1 i_2,\ i_4,\ i_5 = i_1 i_4,\ i_6 = i_2 i_4,\ i_7 = (i_1 i_2) i_4. \tag{101}$$

Da (i)–(v) si deduce per tale base la tavola di moltiplicazione

(102)

	i_0	i_1	i_2	i_3	i_4	i_5	i_6	i_7
i_0	i_0	i_1	i_2	i_3	i_4	i_5	i_6	i_7
i_1	i_1	$c_1 i_0$	i_3	$c_1 i_2$	i_5	$c_1 i_4$	$-i_7$	$-c_1 i_6$
i_2	i_2	$-i_3$	$c_2 i_0$	$-c_2 i_1$	i_6	i_7	$c_2 i_4$	$c_2 i_5$
i_3	i_3	$-c_1 i_2$	$c_2 i_1$	$-c_1 c_2 i_0$	i_7	$c_1 i_6$	$-c_2 i_5$	$-c_1 c_2 i_4$
i_4	i_4	$-i_5$	$-i_6$	$-i_7$	$c_3 i_0$	$-c_3 i_1$	$-c_3 i_2$	$-c_3 i_3$
i_5	i_5	$-c_1 i_4$	$-i_7$	$-c_1 i_6$	$c_3 i_1$	$-c_1 c_3 i_0$	$c_3 i_3$	$c_1 c_3 i_2$
i_6	i_6	i_7	$-c_2 i_4$	$c_2 i_5$	$c_3 i_2$	$-c_3 i_3$	$-c_2 c_3 i_0$	$-c_2 c_3 i_1$
i_7	i_7	$c_1 i_6$	$-c_2 i_5$	$c_1 c_2 i_4$	$c_3 i_3$	$-c_1 c_3 i_2$	$c_2 c_3 i_1$	$c_1 c_2 c_3 i_0$

Se $x = x_0 i_0 + x_1 i_1 \in A_1$, allora $\bar{x} x = (x_0^2 - c_1 x_1^2) 1$, per cui

$$Q(x) = x_0^2 - c_1 x_1^2. \tag{103}$$

Per $y = y_0 i_0 + y_1 i_1$ si ha $xy = (x_0 y_0 + c_1 x_1 y_1) i_0 + (x_0 y_1 + x_1 y_0) i_1$, quindi la legge di composizione in questo caso è

$$(x_0^2 - c_1 x_1^2)(y_0^2 - c_1 y_1^2) = (x_0 y_0 + c_1 x_1 y_1)^2 - c_1 (x_0 y_1 + x_1 y_0)^2. \tag{104}$$

Analogamente, per $x = x_0 i_0 + x_1 i_1 + x_2 i_2 + x_3 i_3$ e $y = y_0 i_0 + y_1 i_1 + y_2 i_2 + y_3 i_3$ in A_2 si ottiene

$$Q(x) = x_0^2 - c_1 x_1^2 - c_2 x_2^2 + c_1 c_2 x_3^2, \tag{105}$$

con legge di composizione

$$\begin{aligned} &(x_0^2 - c_1 x_1^2 - c_2 x_2^2 + c_1 c_2 x_3^2)(y_0^2 - c_1 y_1^2 - c_2 y_2^2 + c_1 c_2 y_3^2) \\ &\quad = (x_0 y_0 + c_1 x_1 y_1 + c_2 x_2 y_2 - c_1 c_2 x_3 y_3)^2 \\ &\qquad - c_1 (x_0 y_1 + x_1 y_0 - c_2 x_2 y_3 + c_2 x_3 y_2)^2 \\ &\qquad - c_2 (x_0 y_2 + x_2 y_0 + c_1 x_1 y_3 - c_1 x_3 y_1)^2 \\ &\qquad + c_1 c_2 (x_0 y_3 + x_3 y_0 + x_1 y_2 - x_2 y_1)^2. \end{aligned} \tag{106}$$

Ponendo $c_i = -1$ si ritrovano le identità elencate all'inizio della discussione. Potremmo scrivere anche la forma quadratica Q fornita dall'algebra di ottonioni e la sua legge di composizione. Ce ne asteniamo per la lunghezza delle formule.

Se il campo base è $F = \mathbb{R}$ e $c_1 = -1$, l'algebra quadratica A_1 ha base (i_0, i_1) con i_0 unità e $i_1^2 = -1$. È evidentemente il campo $\mathbb{C}$ dei numeri complessi. Ponendo $c_2 = -1$ otteniamo l'algebra di quaternioni con base (i_0, i_1, i_2, i_3), dove $i_0 = 1$, $i_j^2 = -1$ per $1 \leq j \leq 3$, $i_1 i_2 = i_3 = -i_2 i_1$, $i_2 i_3 = i_1 = -i_3 i_2$, $i_3 i_1 = i_2 = -i_1 i_3$. È l'algebra $\mathbb{H}$ dei quaternioni di Hamilton. Ponendo $c_3 = -1$ otteniamo l'algebra classica degli ottonioni $\mathbb{O}$, scoperta indipendentemente da J. J. Graves prima del 1844 e da A. Cayley nel 1845. La sua definizione come doppio di un'algebra di quaternioni è dovuta a L. E. Dickson.[12] L'algebra di Cayley–Graves $\mathbb{O}$ è un'algebra di divisione nel senso che ogni $x \neq 0$ possiede un inverso x^{-1} tale che $xx^{-1} = 1 = x^{-1}x$. Se $x = \sum_{j=0}^{7} x_j i_j$, allora $\bar{x} = x_0 i_0 - \sum_{k=1}^{7} x_k i_k$ e $Q(x) = \sum x_i^2 \neq 0$. Possiamo prendere $x^{-1} = Q(x)^{-1}\bar{x}$. Più in generale, si vede che ogni algebra di composizione la cui forma quadratica è anisotropa è un'algebra di divisione.

Esercizi.

1. Mostrare che ogni x in un'algebra di composizione soddisfa l'equazione quadratica $x^2 - T(x)x + Q(x)1 = 0$.

2. Usare il teorema di Witt e il procedimento di raddoppiamento per dimostrare che, se (A, Q) e (A', Q') sono algebre di composizione con forme quadratiche equivalenti Q, Q' e se C, C' sono sottoalgebre non degeneri isomorfe rispettivamente di A, A', allora ogni isomorfismo di C su C' si estende a un isomorfismo di A su A'. Dedurre che (A, Q) e (A', Q') sono isomorfe se e solo se Q e Q' sono equivalenti.

3. Mostrare che se (A, Q) è un'algebra di composizione che non è di divisione, allora Q ha indice di Witt massimo, uguale a $n/2$ con $n = 2, 4, 8$. Una tale algebra di composizione si dice *scissa*. Mostrare che due algebre scisse della stessa dimensione sono isomorfe.

4. Mostrare che, su un campo finito F, ogni algebra di composizione di dimensione 4 oppure 8 è scissa. Vale lo stesso per $n = 2$?

5. Su un campo F di caratteristica qualsiasi si definisca *algebra quadratica* un'algebra $F[\lambda]/(\lambda^2 - \lambda + a)$ con $4a \neq 1$, insieme con la forma $Q(b + cu) = b^2 + bc + c^2 a$, dove $u = \lambda + (\lambda^2 - \lambda + a)$. Mostrare che essa possiede un'involuzione $u \mapsto \bar{u} = 1 - u$ per la quale $\bar{x}x = Q(x)1$ per $x = b + cu$. Mostrare che in caratteristica diversa da 2 si ottengono, a meno di isomorfismo, le algebre quadratiche definite nel testo.

6. Su un campo F di caratteristica qualsiasi si definisca un'algebra di quaternioni come doppio di un'algebra quadratica nel senso dell'esercizio 5, e un'algebra di ottonioni come doppio di un'algebra di quaternioni. Si definiscano le algebre di composizione su F come nel testo, intendendo per non degenerazione che l'unico z tale che $Q(z) = 0$ e $Q(x, z) = 0$ per ogni x è $z = 0$. Dimostrare il seguente teorema generalizzato di Hurwitz per ogni F. Le algebre di composizione su F sono: (I) F, (II) le algebre quadratiche, (III) le algebre di quaternioni, (IV) le algebre di ottonioni, (V) per $\operatorname{char} F = 2$, le estensioni di campi A di dimensione finita su F tali che $x^2 = Q(x) \in F$ per ogni $x \in A$.

[12] L. E. Dickson, *Linear Algebras*, Cambridge Tracts in Mathematics, n. 16, 1914, p. 15, oppure *Algebras and Their Arithmetics*, University of Chicago Press, 1923, p. 62.

7. Mostrare che, definendo le algebre di composizione come nella Definizione 7.5, il teorema generalizzato di Hurwitz rimane valido anche sopprimendo l'ipotesi implicita di dimensione finita.

8. Dimostrare le seguenti identità di Moufang per algebre alternative di caratteristica qualsiasi:
$$(uvu)x = u(v(ux)), \qquad x(uvu) = ((xu)v)u, \qquad u(xy)u = (ux)(yu).$$

9. Mostrare che la seconda identità precedente equivale all'identità fra associatori $[x, uv, u] = -[x, u, v]u$ e che per linearizzazione questa dà
$$[x, uv, w] + [x, wv, u] = -[x, u, v]w - [x, w, v]u.$$
Servirsene per provare il teorema di Artin: la sottoalgebra generata da due elementi qualsiasi di un'algebra alternativa è associativa. (Si osservi che ne segue l'associatività delle potenze nelle algebre alternative.)

7.7 I teoremi di Frobenius e di Wedderburn sulle algebre associative di divisione

Nel volume II studieremo la teoria di struttura degli anelli e delle algebre associative. Uno dei suoi risultati principali è la riduzione dello studio di classi abbastanza generali di anelli a quello degli anelli di divisione. Nel caso delle algebre di dimensione finita la riduzione porta alle algebre di divisione. Che cosa si può dire di esse? La risposta dipende sensibilmente dal campo base. In questo paragrafo esamineremo i tre casi più semplici, nei quali F è algebricamente chiuso, oppure è il campo $\mathbb{R}$ dei numeri reali, oppure è un campo finito.[13]

Sia A un'algebra associativa di dimensione finita su F che sia un'algebra di divisione, nel senso che ogni $x \neq 0$ di A ha un inverso in A. Se $m_x(\lambda)$ è il polinomio minimo di x e $m_x(\lambda) = m_1(\lambda)m_2(\lambda)$ in $F[\lambda]$, allora $m_1(x)m_2(x) = 0$ e quindi $m_1(x) = 0$ oppure $m_2(x) = 0$. Ne segue che il polinomio minimo di ogni $x \in A$ è irriducibile. Identificheremo F con la sottoalgebra $F1$ dei multipli $a1$, $a \in F$. Risulta allora evidente che $m_x(\lambda)$ è lineare se e solo se $x = a \in F$.

Determinare le algebre di divisione di dimensione finita su un campo algebricamente chiuso F è immediato. Vale infatti il seguente teorema.

Teorema 7.6. *Se F è algebricamente chiuso, l'unica algebra associativa di divisione di dimensione finita su F è F stesso.*

Dimostrazione. Sia A un'algebra di divisione di dimensione finita sul campo algebricamente chiuso F, e sia $x \in A$. Il polinomio minimo di x è irriducibile e F è algebricamente chiuso, dunque esso è lineare. Pertanto $F[x] = F$ e $x \in F$. Poiché ciò vale per ogni $x \in A$, si ha $A = F$. □

Consideriamo ora il caso $F = \mathbb{R}$, con A algebra di divisione di dimensione finita su $\mathbb{R}$. I polinomi monici irriducibili di $\mathbb{R}[\lambda]$ sono quelli lineari $\lambda - a$ e quelli quadratici $\lambda^2 - 2a\lambda + b$ con $a^2 < b$. Se $x \notin \mathbb{R}$, il suo polinomio minimo ha quest'ultima forma e $x = y + a$, dove il polinomio minimo di y è $\lambda^2 + (b - a^2)$. Quindi ogni elemento di A ha la forma $a + y$ con $a \in \mathbb{R}$ e $y = 0$ oppure $y^2 = b \in \mathbb{R}$, $b < 0$. Questa semplice osservazione ci consentirà di dimostrare il seguente teorema.

[13] Il caso $F = \mathbb{Q}$ si rivela sorprendentemente difficile e richiede profondi risultati di aritmetica.

Teorema di Frobenius. *Le uniche algebre associative di divisione di dimensione finita sui reali sono (1)* $\mathbb{R}$*, (2)* $\mathbb{C}$*, (3)* $\mathbb{H}$*.*

Dimostrazione. La nostra dimostrazione è una versione leggermente perfezionata di quella di Dickson.[14] Indichiamo con A' il sottoinsieme di A formato dagli elementi u il cui quadrato è un reale ≤ 0. Affermiamo che A' è un sottospazio di A. Poiché per $u \in A'$ e $a \in \mathbb{R}$ è evidente che $au \in A'$, basta mostrare che se $u, v \in A'$ sono linearmente indipendenti, allora $u + v \in A'$. Osserviamo anzitutto che non può sussistere una relazione della forma $u = av + b$ con $a, b \in \mathbb{R}$. Infatti $u^2 = c < 0$, $v^2 = d < 0$, essendo $u, v \neq 0$, e da $u = av + b$ seguirebbe $c = (av + b)^2 = a^2d + 2abv + b^2$. Poiché $v \notin \mathbb{R}$, avremmo $ab = 0$ e dunque $a = 0$ o $b = 0$. Nel primo caso u sarebbe reale, nel secondo sarebbe un multiplo di v. Entrambi i casi sono esclusi. Ne segue che $1, u, v$ sono linearmente indipendenti.

Consideriamo ora $u + v$ e $u - v$. Entrambi soddisfano equazioni quadratiche, quindi esistono $p, q, r, s \in \mathbb{R}$ tali che

$$(u + v)^2 = p(u + v) + q, \qquad (u - v)^2 = r(u - v) + s.$$

Poiché $(u \pm v)^2 = u^2 \pm (uv + vu) + v^2$, e $u^2 = c$, $v^2 = d$, ricaviamo

$$c + d + (uv + vu) = p(u + v) + q, \qquad c + d - (uv + vu) = r(u - v) + s.$$

Sommando otteniamo $(p + r)u + (p - r)v + (q + s - 2c - 2d) = 0$. Per l'indipendenza lineare di $u, v, 1$ segue $p = r = 0$. Allora $(u + v)^2 = q \in \mathbb{R}$ e, poiché $u + v \notin \mathbb{R}$, si ha $q < 0$. Dunque $u + v \in A'$ e A' è un sottospazio. Sopra abbiamo visto che ogni elemento di A si scrive $a + y$ con $a \in \mathbb{R}$ e $y \in A'$, per cui $A = \mathbb{R} \oplus A'$.

Per $u \in A'$ poniamo $u^2 = -Q(u)$, dove $Q(u) \in \mathbb{R}$ e $Q(u) \geq 0$. Inoltre $Q(u) = 0$ se e solo se $u = 0$. Per $a \in \mathbb{R}$ è chiaro che $Q(au) = a^2Q(u)$ e

$$B(u, v) = Q(u + v) - Q(u) - Q(v) = -(u + v)^2 + u^2 + v^2 = -(uv + vu).$$

Da quest'ultima formula si vede che B è bilineare e simmetrica. Pertanto Q è una forma quadratica con forma bilineare simmetrica associata B ed è definita positiva.

Possiamo ora concludere. Se $A = \mathbb{R}$ siamo nel primo caso. Supponiamo $\mathbb{R} \subsetneq A$. Allora $A' \neq 0$ e possiamo scegliere $i \in A'$ con $Q(i) = 1$. Si ha $i^2 = -1$ e $\mathbb{R}[i] = \mathbb{C} = \mathbb{R} + \mathbb{R}i$. Se $A = \mathbb{C}$ abbiamo la seconda possibilità. Supponiamo ora $\mathbb{C} \subsetneq A$. Allora $\mathbb{R}i \subsetneq A'$ e possiamo scegliere $j \perp \mathbb{R}i$ con $Q(j) = 1$. Abbiamo $j^2 = -1$ e $ij + ji = -B(i, j) = 0$, quindi $ij = -ji$. Ponendo $k = ij$ otteniamo $k^2 = -1$, $ik + ki = 0 = kj + jk$. Perciò $k \in A'$ ed è ortogonale a i, j. Ne segue che $1, i, j, k$ sono linearmente indipendenti e $\mathbb{R} + \mathbb{R}i + \mathbb{R}j + \mathbb{R}k = \mathbb{H}$. Mostriamo che $A = \mathbb{H}$. In caso contrario esisterebbe $\ell \in A'$ tale che $Q(\ell) = 1$ e $\ell \perp i, j, k$. Allora $\ell i = -i\ell$, $\ell j = -j\ell$, $\ell k = -k\ell$, con $k = ij$. Tuttavia le prime due uguaglianze danno

$$\ell(ij) = (\ell i)j = -(i\ell)j = -i(\ell j) = i(j\ell) = (ij)\ell,$$

ossia $\ell k = k\ell$. Questa contraddizione mostra che $A = \mathbb{H}$, e abbiamo la terza possibilità. □

Passiamo al caso di un campo finito F. Se $|F| = q$ e V è uno spazio vettoriale di dimensione n su F, allora $|V| = q^n$. In particolare ogni algebra di divisione di dimensione finita su F è un anello di divisione finito. Viceversa, siano A un anello di divisione finito e F il suo centro. Allora F è un campo finito e A è un'algebra di dimensione finita su F. Nel 1905 J. H. M. Wedderburn scoprì il sorprendente fatto che ogni anello di divisione finito è commutativo. Il teorema di

[14] L. E. Dickson, *Linear Algebras*, pp. 10–12.

Wedderburn ha una notevole conseguenza per la geometria proiettiva. È noto infatti che le geometrie proiettive desarguesiane, cioè quelle in cui vale il teorema di Desargues, ammettono coordinate in un anello di divisione, e che tale anello è commutativo se e solo se vale il teorema di Pappo.[15] Dal teorema di Wedderburn segue dunque che il teorema di Pappo vale in ogni geometria proiettiva finita nella quale valga il teorema di Desargues. Dimostriamo ora il teorema.

Teorema di Wedderburn. *Ogni anello di divisione finito è commutativo.*

Dimostrazione. Sia F il centro dell'anello di divisione finito A, con $|F| = q$ e $[A : F] = n$. Allora $|A| = q^n$. Dobbiamo mostrare che $n = 1$. Sia A^* il gruppo moltiplicativo degli elementi non nulli di A. L'equazione delle classi dà

$$|A^*| = q^n - 1 = \sum_i [A^* : \operatorname{Stab} x_i], \tag{107}$$

dove gli x_i percorrono un sistema di rappresentanti delle classi di coniugio di A^*, un elemento per classe. Se $x_i \in F$, allora $\operatorname{Stab} x_i = A^*$ e il corrispondente termine della somma vale 1. Il numero di tali $x_i \in F \cap A^*$ è $q - 1$, sicché il loro contributo complessivo è $q - 1$. Sia ora $x_i \notin F$. Gli elementi di A che commutano con x_i formano un sottoanello di divisione F_i di A contenente F. Dunque $|F_i| = q^{d_i}$ con $d_i = [F_i : F]$, e $d_i < n$, poiché $x_i \notin F$ e quindi F_i è proprio in A. È chiaro che $\operatorname{Stab} x_i = F_i \cap A^*$, dunque $|\operatorname{Stab} x_i| = q^{d_i} - 1$. La (107) diviene

$$q^n - 1 = (q - 1) + \sum_i \frac{q^n - 1}{q^{d_i} - 1}, \tag{108}$$

dove ogni $d_i < n$. Inoltre $d_i \mid n$. Infatti A si può considerare in modo naturale come uno spazio vettoriale sinistro su F_i, e F_i a sua volta come spazio vettoriale su F. Vale la formula del prodotto delle dimensioni, come per i campi (Teorema 4.2). Quindi la dimensione n di A su F è divisibile per la dimensione d_i di F_i su F. Fino a questo punto la dimostrazione è quella di Wedderburn. Resta da provare che, se $d_i \mid n$ e $d_i < n$, la (108) è impossibile, tranne per $n = 1$. L'argomento che daremo è dovuto a E. Witt.

Consideriamo $\lambda^n - 1$ e $\lambda^{d_i} - 1$, con λ indeterminata. Ricordiamo che, se si definisce l'n-esimo polinomio ciclotomico $l_n(\lambda) = \prod(\lambda - z)$, dove z percorre le radici primitive n-esime dell'unità in $\mathbb{C}$, allora $\lambda^n - 1 = \prod_{d|n} l_d(\lambda)$ (§ 4.11). Abbiamo anche visto che i polinomi l_d sono monici a coefficienti interi. Ne segue che, per $d_i \mid n$ e $d_i < n$, il quoziente $(\lambda^n - 1)/(\lambda^{d_i} - 1)$ è un polinomio a coefficienti interi divisibile in $\mathbb{Z}[\lambda]$ per $l_n(\lambda)$. Quindi tanto $(q^n - 1)/(q^{d_i} - 1)$ quanto $q^n - 1$ sono divisibili per l'intero $l_n(q)$. Dalla (108) segue allora $l_n(q) \mid q - 1$. Supponiamo ora $n > 1$ e consideriamo la fattorizzazione $l_n(\lambda) = \prod(\lambda - z)$ sulle radici primitive n-esime dell'unità. Nessuna di queste è 1. La distanza del punto q sull'asse reale da ciascun z supera dunque la distanza di q dal punto 1. Perciò $|q - z| > q - 1$ e $|l_n(q)| = \prod |q - z| > q - 1$, contro la divisibilità $l_n(q) \mid q - 1$. Pertanto $n = 1$ e $A = F$ è commutativo. □

Esercizio.

1. Dimostrare la seguente estensione del teorema di Frobenius alle algebre alternative di divisione. Le uniche algebre alternative di divisione di dimensione finita su $\mathbb{R}$ sono (a) $\mathbb{R}$, (b) $\mathbb{C}$, (c) $\mathbb{H}$, (d) $\mathbb{O}$. (Applicare il teorema generalizzato di Hurwitz.)

[15] Si veda E. Artin, *Geometric Algebra*, New York, Wiley, 1957, p. 73.

Capitolo 8
Reticoli e algebre di Boole

A ogni insieme S è associato il suo insieme delle parti $\mathcal{P}(S)$, ossia l'insieme dei suoi sottoinsiemi, e l'algebra (in senso non tecnico) di $\mathcal{P}(S)$ fondata sull'intersezione $A \cap B$ e sull'unione $A \cup B$. Cercando di descrivere le proprietà fondamentali della struttura $(\mathcal{P}(S), \cap, \cup)$ si giunge al concetto astratto di algebra di Boole. George Boole riconobbe per primo che una tale algebra si poteva usare per analizzare il calcolo delle proposizioni della logica e che svolgeva un ruolo fondamentale nella teoria della probabilità.[1] Un concetto più generale è quello di reticolo, introdotto da Dedekind nello studio della divisibilità negli anelli commutativi e delle proprietà combinatorie degli ideali rispetto all'intersezione $A \cap B$ e alla somma $A + B$.[2]

In questo capitolo daremo un'introduzione ai reticoli e alle algebre di Boole. Vogliamo familiarizzare il lettore con i concetti e i risultati elementari di questa teoria che trovano applicazione in altre parti dell'algebra. Alcuni ci occorreranno nel secondo volume per lo studio dell'algebra universale. Oltre alle definizioni fondamentali tratteremo principalmente il teorema di Jordan–Hölder per i reticoli semimodulari, il «teorema fondamentale della geometria proiettiva», che determina gli isomorfismi fra i reticoli dei sottospazi di spazi vettoriali, l'equivalenza fra algebre di Boole e anelli booleani, e la funzione di Möbius di un insieme parzialmente ordinato.

8.1 Insiemi parzialmente ordinati e reticoli

Il concetto più generale che considereremo in questo capitolo è quello di insieme parzialmente ordinato. Ricordiamo che una relazione binaria su un insieme S è un sottoinsieme R del prodotto $S \times S$ (Introduzione, p. 10). Diciamo che a è in relazione R con b, e scriviamo aRb, se e solo se $(a, b) \in R$. Diamo ora la seguente definizione.

Definizione 8.1. *Un insieme parzialmente ordinato è un insieme S munito di una relazione binaria $a \geq b$ che soddisfa le condizioni seguenti.*

PO1 $a \geq a$ (riflessività).

PO2 Se $a \geq b$ e $b \geq a$, allora $a = b$ (antisimmetria).

PO3 Se $a \geq b$ e $b \geq c$, allora $a \geq c$ (transitività).

Se $a \geq b$ e $a \neq b$, scriviamo $a > b$. Scriviamo anche $a \leq b$ per $b \geq a$ e $a < b$ per $b > a$. In generale, dati $a, b \in S$, può non valere né $a \geq b$ né $b \geq a$. Se per ogni coppia di elementi si ha $a \geq b$ oppure $b \geq a$, diciamo che S è *totalmente ordinato* o che è una *catena*.

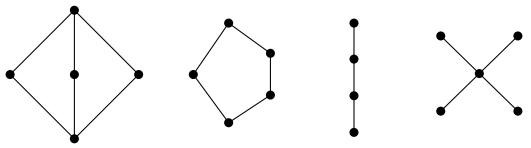
Abbiamo già incontrato parecchi esempi di insiemi parzialmente ordinati: l'insieme $\mathcal{P}(S)$ dei sottoinsiemi di S, in cui $A \geq B$ significa $A \supseteq B$; l'insieme dei sottoanelli di un anello, quello dei sottogruppi di un gruppo, quello degli ideali di un anello, e così via, tutti ordinati per inclusione. In generale, se $(S, \geq)$ è parzialmente ordinato, ogni suo sottoinsieme T è parzialmente ordinato dalla restrizione a T della relazione di S. Altri esempi interessanti provengono dalla divisibilità nei monoidi e negli anelli. Nel monoide moltiplicativo degli interi positivi poniamo $a \geq b$ quando $a \mid b$ (a è un divisore di b). Valgono PO1–PO3. Più generalmente, sia S un monoide commutativo

[1] George Boole, *The Mathematical Analysis of Logic*, 1847 (ristampa Barnes and Noble, 1965), e *An Investigation of the Laws of Thought*, 1854 (ristampa Dover, 1953).

[2] Richard Dedekind, «Über Zerlegungen von Zahlen durch ihre grössten gemeinsamen Teiler», in *Gesammelte mathematische Werke*, vol. 2, 1931, pp. 103–147, e «Über die von drei Moduln erzeugte Dualgruppe», ivi, pp. 236–272.

in cui vale la legge di cancellazione. Si dice *ridotto* se 1 è il solo elemento invertibile. Allora $a \mid b$ e $b \mid a$ implicano $a = b$, e la relazione $a \geq b$ definita da $a \mid b$ ordina parzialmente S. Se S non è ridotto, ponendo $a \sim b$ quando $a = bu$ con u invertibile otteniamo una congruenza non banale. Il monoide quoziente $\bar{S}$ rispetto a questa congruenza è ridotto ed è parzialmente ordinato dalla divisibilità.

In un insieme parzialmente ordinato finito si può esprimere $>$ per mezzo della relazione di copertura. Diciamo che a_1 *copre* a_2 se $a_1 > a_2$ e non esiste u tale che $a_1 > u > a_2$. In un insieme finito si ha $a > b$ se e solo se esiste una successione $a = a_1, a_2, \ldots, a_n = b$ in cui ogni a_i copre a_{i+1}. Questo suggerisce una rappresentazione mediante diagrammi. Rappresentiamo gli elementi di S con punti. Se a_1 copre a_2, collochiamo a_1 sopra a_2 e congiungiamo i due punti con un segmento. Allora $a > b$ se e solo se una spezzata discendente congiunge a a b. Se $a \neq b$ e nessuna spezzata monotona li congiunge, i due elementi sono inconfrontabili, cioè non vale né $a \geq b$ né $b \geq a$. Ecco alcuni esempi.



Il terzo rappresenta un insieme totalmente ordinato.

Un elemento u di un insieme parzialmente ordinato S è un *maggiorante* di un sottoinsieme A se $u \geq a$ per ogni $a \in A$. È un *estremo superiore*, o $\sup A$, se è un maggiorante e $u \leq v$ per ogni maggiorante v di A. Per PO2, se $\sup A$ esiste è unico. Si definiscono analogamente i *minoranti* e l'*estremo inferiore*, o $\inf A$, di A. Anche $\inf A$, quando esiste, è unico.

Definizione 8.2. *Un reticolo è un insieme parzialmente ordinato in cui ogni coppia di elementi ammette un estremo superiore e un estremo inferiore.*

Denotiamo con $a \vee b$ l'estremo superiore di a, b (la loro unione reticolare) e con $a \wedge b$ l'estremo inferiore (la loro intersezione reticolare). Per a, b, c in un reticolo L si ha $(a \vee b) \vee c \geq a, b, c$ e, se $v \geq a, b, c$, allora $v \geq a \vee b, c$, per cui $v \geq (a \vee b) \vee c$. Dunque $(a \vee b) \vee c$ è l'estremo superiore di a, b, c. Per induzione ogni insieme finito non vuoto di elementi ha estremo superiore, e analogamente estremo inferiore. Indichiamo questi due elementi per $a_1, \ldots, a_n$ rispettivamente con

$$a_1 \vee a_2 \vee \cdots \vee a_n, \qquad a_1 \wedge a_2 \wedge \cdots \wedge a_n.$$

Ogni insieme totalmente ordinato è un reticolo. Infatti, presi due elementi, si ha $a \geq b$ oppure $b \geq a$. Nel primo caso $a \vee b = a$ e $a \wedge b = b$, nel secondo $a \vee b = b$ e $a \wedge b = a$.

Un insieme parzialmente ordinato si chiama *reticolo completo* se ogni sottoinsieme $A = \{a_\alpha\}$ ammette estremo superiore ed estremo inferiore. Li denotiamo con $\bigvee_\alpha a_\alpha$ e $\bigwedge_\alpha a_\alpha$. Se $\{a_\alpha\}$ coincide con l'insieme sottostante a L, allora $0 = \bigwedge_\alpha a_\alpha$ è il minimo di L e $1 = \bigvee_\alpha a_\alpha$ il suo massimo: $0 \leq a$ e $1 \geq a$ per ogni $a \in L$. È utile il seguente criterio per riconoscere i reticoli completi.

Teorema 8.1. *Un insieme parzialmente ordinato dotato di massimo* 1*, in cui ogni sottoinsieme non vuoto* $\{a_\alpha\}$ *ha estremo inferiore, è un reticolo completo. Dualmente, un insieme parzialmente ordinato dotato di minimo* 0*, in cui ogni sottoinsieme non vuoto ha estremo superiore, è un reticolo completo.*

Dimostrazione. Supposte le prime ipotesi, dobbiamo mostrare che ogni $A = \{a_\alpha\}$ ha estremo superiore. Poiché $1 \geq a_\alpha$, l'insieme B dei maggioranti di A non è vuoto. Sia $b = \inf B$. Allora $b = \sup A$. La seconda affermazione segue per simmetria. □

Esempi

1. Per ogni insieme S, $\mathcal{P}(S)$ è un reticolo completo. Qui $1 = S$ e $0 = \varnothing$.

2. L'insieme dei sottogruppi di un gruppo G, ordinato per inclusione. Poiché G è un sottogruppo e l'intersezione di una famiglia qualsiasi di sottogruppi è un sottogruppo, questo insieme è un reticolo completo. Dalla dimostrazione del teorema 8.1 risulta che l'estremo superiore della famiglia $\{H_\alpha\}$ è l'intersezione di tutti i sottogruppi che contengono ciascun H_α, ossia il sottogruppo da essi generato.

   I quattro esempi seguenti sono analoghi al secondo: sono reticoli completi nei quali $\geq$ indica l'inclusione.

3. L'insieme dei sottogruppi normali di un gruppo. L'estremo superiore di una famiglia di sottogruppi normali è il sottogruppo che essi generano.

4. L'insieme dei sottospazi di uno spazio vettoriale, ordinato per inclusione. L'estremo inferiore è l'intersezione insiemistica e l'estremo superiore è il sottospazio generato dai sottospazi dati.

5. L'insieme degli ideali di un anello R. L'estremo inferiore è l'intersezione e quello superiore è l'ideale generato. Per due ideali I_1, I_2 quest'ultimo è $I_1 + I_2$, l'insieme delle somme $b_1 + b_2$ con $b_i \in I_i$.

6. L'insieme degli ideali sinistri (o destri) di un anello.

7. Gli interi positivi ordinati per divisibilità, $a \geq b \iff a \mid b$. Qui $a \vee b$ è il massimo comune divisore di a, b e $a \wedge b$ il loro minimo comune multiplo. Questo reticolo non è completo.

8. Tutti i diagrammi precedenti tranne l'ultimo rappresentano reticoli (necessariamente completi, essendo finiti).

9. L'insieme $\mathbb{Q}$ dei razionali con l'usuale significato di $a \geq b$. È totalmente ordinato ed è quindi un reticolo, ma non è completo.

10. Neppure il sottoinsieme dei razionali fra 0 e 1 è completo. L'intervallo reale $[0, 1]$, con l'ordine usuale, è invece un reticolo completo.

Esaminiamo le proprietà fondamentali delle operazioni binarie $a \wedge b$ e $a \vee b$. Otterremo una definizione alternativa di reticolo per mezzo di condizioni su due operazioni binarie. Per definizione $a \vee b$ e $a \wedge b$ sono simmetrici nei due argomenti, per cui si hanno le leggi commutative $a \vee b = b \vee a$ e $a \wedge b = b \wedge a$. Inoltre $(a \vee b) \vee c$ è l'estremo superiore di a, b, c. Essendo questo simmetrico nei tre elementi, $(a \vee b) \vee c = a \vee (b \vee c)$, e analogamente $(a \wedge b) \wedge c = a \wedge (b \wedge c)$. Ogni a è idempotente rispetto a entrambe le operazioni: $a \vee a = a$, $a \wedge a = a$. Se $a \geq b$, allora $a \vee b = a$ e $a \wedge b = b$. Pertanto, per a, b qualsiasi, $(a \vee b) \wedge a = a$ e $(a \wedge b) \vee a = a$.

Viceversa, sia L un insieme su cui siano definite due operazioni binarie $\vee$ e $\wedge$ soddisfacenti le condizioni appena incontrate:

L1	$a \vee b = b \vee a$,	$a \wedge b = b \wedge a$,
L2	$(a \vee b) \vee c = a \vee (b \vee c)$,	$(a \wedge b) \wedge c = a \wedge (b \wedge c)$,
L3	$a \vee a = a$,	$a \wedge a = a$,
L4	$(a \vee b) \wedge a = a$,	$(a \wedge b) \vee a = a$.

Mostreremo che, definendo opportunamente $\geq$, L è un reticolo in cui $a \vee b$ e $a \wedge b$ sono l'estremo superiore e quello inferiore di a, b.

Prima della dimostrazione osserviamo che le ipotesi sulle due operazioni sono perfettamente simmetriche. Ne segue il *principio di dualità*: se un enunciato S è deducibile dai nostri assiomi, lo è anche l'*enunciato duale* S' ottenuto scambiando ovunque $\vee$ con $\wedge$. Per $a, b \in L$, dagli assiomi L1–L4 segue inoltre che le condizioni $a \vee b = a$ e $a \wedge b = b$ sono equivalenti. Definiamo in L la relazione $a \geq b$ come $a \vee b = a$, ossia $a \wedge b = b$. Nel passaggio al duale, $a \geq b$ va quindi sostituita con $b \geq a$.

Verifichiamo PO1–PO3. Poiché $a \vee a = a$, si ha $a \geq a$, ossia PO1. Se $a \geq b$ e $b \geq a$, allora $a \vee b = a$ e $b \vee a = b$; per la commutatività ne segue $a = b$, ossia PO2. Se $a \geq b$ e $b \geq c$, allora $a \vee b = a$ e $b \vee c = b$. Quindi

$$a \vee c = (a \vee b) \vee c = a \vee (b \vee c) = a \vee b = a,$$

vale a dire $a \geq c$. È provata PO3.

Da $(a \vee b) \wedge a = a$ per L4 segue $a \vee b \geq a$. Analogamente $a \vee b \geq b$. Se $c \geq a$ e $c \geq b$, allora $a \vee c = c$ e $b \vee c = c$, per cui

$$(a \vee b) \vee c = a \vee (b \vee c) = a \vee c = c,$$

ossia $c \geq a \vee b$. Dunque $a \vee b$ è l'estremo superiore di a, b. Per dualità $a \wedge b$ è l'estremo inferiore. Ciò dimostra che un insieme con due operazioni binarie che verificano L1–L4 è un reticolo con gli estremi indicati.

Un sottoinsieme M di L è un *sottoreticolo* se è chiuso rispetto a $\vee$ e $\wedge$. È chiaro che un sottoreticolo è un reticolo per le operazioni indotte. D'altra parte un sottoinsieme di L può essere un reticolo per l'ordine indotto senza essere un sottoreticolo. Per esempio, il reticolo dei sottogruppi di G non è in generale un sottoreticolo di $\mathcal{P}(G)$, poiché $H_1 \cup H_2$ non è in generale un sottogruppo. Se $a \in L$ è fissato, gli elementi $x \geq a$ (oppure $x \leq a$) formano un sottoreticolo. Se $a \leq b$, gli elementi $x \in L$ con $a \leq x \leq b$ formano un sottoreticolo detto *intervallo* e denotato con $I[a, b]$.

La definizione mediante L1–L4 suggerisce di chiamare *omomorfismo* di un reticolo L in un reticolo L' un'applicazione $a \mapsto a'$ tale che $(a \vee b)' = a' \vee b'$ e $(a \wedge b)' = a' \wedge b'$. Da $a \geq b$ segue $a \vee b = a$, quindi $a' \vee b' = a'$ e $a' \geq b'$. Un'applicazione tra insiemi parzialmente ordinati con questa proprietà si dice *monotona* o *che conserva l'ordine*. Ogni omomorfismo di reticoli conserva quindi l'ordine. Il viceversa non vale in generale. Un omomorfismo biiettivo di reticoli si dice *isomorfismo*. Lo si può caratterizzare in termini di ordine come segue.

Teorema 8.2. *Un'applicazione biiettiva di un reticolo L su un reticolo L' è un isomorfismo di reticoli se e solo se sia essa sia la sua inversa conservano l'ordine.*

Dimostrazione. Se $a \mapsto a'$ è un isomorfismo di reticoli conserva l'ordine, come abbiamo visto. Anche l'inversa è un isomorfismo di L' su L e conserva l'ordine. Viceversa, supponiamo che $a \mapsto a'$ sia biiettiva e che l'applicazione e l'inversa conservino l'ordine. Allora $a \geq b$ in L se e solo se $a' \geq b'$ in L'. Sia $d = a \vee b$. Da $d \geq a, b$ segue $d' \geq a', b'$. Sia $e' \geq a', b'$ e sia e la sua controimmagine. Allora $e \geq a, b$, quindi $e \geq d$ ed $e' \geq d'$. Abbiamo provato $d' = a' \vee b'$. Analogamente si dimostra $(a \wedge b)' = a' \wedge b'$. □

Esercizi

1. Mostrare che il reticolo dei sottogruppi di un gruppo ciclico il cui ordine è una potenza di un primo è totalmente ordinato.

2. Che cosa si può dire del viceversa dell'esercizio 1?

3. Costruire i diagrammi dei seguenti insiemi parzialmente ordinati: (i) $\mathcal{P}(S)$ per $S = \{1, 2, 3\}$, (ii) il reticolo dei sottogruppi di un gruppo ciclico di ordine 6, (iii) il reticolo dei sottogruppi del gruppo simmetrico S_3.

4. Sia S l'insieme delle funzioni reali continue su $[0, 1]$. Poniamo $f \geq g$ se $f(x) \geq g(x)$ per ogni $x \in [0, 1]$. Mostrare che S è un reticolo rispetto a questa relazione. È completo?

5. Il *duale* di $(S, \geq)$ è $(S, \geq')$, dove $a \geq' b$ se e solo se $a \leq b$. Descrivere la relazione tra il diagramma del duale $S' = (S, \geq')$ di un insieme parzialmente ordinato finito e il diagramma di S.

6. Costruendo i diagrammi, determinare tutti i reticoli con al più cinque elementi. Quali sono *autoduali*, ossia isomorfi ai loro duali?

7. Siano L_1, L_2 insiemi parzialmente ordinati. Sul prodotto si definisca $(a_1, a_2) \geq (b_1, b_2)$ se e solo se $a_1 \geq b_1$ e $a_2 \geq b_2$. Mostrare che, se L_1, L_2 sono reticoli, anche $L_1 \times L_2$ è un reticolo.

8. Siano S un insieme ed $(L, \geq)$ un insieme parzialmente ordinato. Nell'insieme L^S delle applicazioni $s \mapsto f(s)$ di S in L si ponga $f \geq g$ se $f(s) \geq g(s)$ per ogni s. Mostrare che si ottiene un ordine parziale su L^S e che L^S è un reticolo se L lo è.

9. Dare un esempio di due reticoli L_1, L_2 per i quali esista un'applicazione biiettiva di L_1 su L_2 che conserva l'ordine e che non è un isomorfismo.

8.2 Distributività e modularità

Si può considerare una delle operazioni del reticolo come analoga all'addizione di un anello e l'altra come analoga alla moltiplicazione. Secondo la scelta otteniamo la legge distributiva

$$\text{D} \qquad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

e la sua duale

$$\text{D}' \qquad a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c).$$

È un po' sorprendente che, come mostreremo ora, le due condizioni siano equivalenti. Supposta D, si ha

$$\begin{aligned}(a \vee b) \wedge (a \vee c) &= ((a \vee b) \wedge a) \vee ((a \vee b) \wedge c) \\ &= a \vee ((a \vee b) \wedge c) \\ &= a \vee ((a \wedge c) \vee (b \wedge c)) \\ &= (a \vee (a \wedge c)) \vee (b \wedge c) \\ &= a \vee (b \wedge c),\end{aligned}$$

che è D′. Dualmente D′ implica D. Un reticolo in cui valgono queste leggi si dice *distributivo*. Un primo esempio importante, come abbiamo mostrato nell'Introduzione (p. 4), è il reticolo $\mathcal{P}(S)$ dei sottoinsiemi di S. Un secondo è dato dal lemma seguente.

Lemma. *Ogni insieme totalmente ordinato è un reticolo distributivo.*

Dimostrazione. Per verificare D per tre elementi qualsiasi a, b, c, distinguiamo due casi: (1) $a \geq b$ e $a \geq c$; (2) $a \leq b$ oppure $a \leq c$. Nel caso (1) entrambi i membri, $a \wedge (b \vee c)$ e $(a \wedge b) \vee (a \wedge c)$, sono uguali a $b \vee c$. Nel caso (2) entrambi sono uguali ad a. Dunque D vale sempre. □

Il lemma permette di mostrare che gli interi positivi ordinati per divisibilità formano un reticolo distributivo. Qui $a \vee b = (a, b)$ è il massimo comune divisore e $a \wedge b = [a, b]$ il minimo

comune multiplo. Se $a = p_1^{a_1} \cdots p_k^{a_k}$, $b = p_1^{b_1} \cdots p_k^{b_k}$, con p_i primi distinti e a_i, b_i interi non negativi, allora $(a,b) = \prod_i p_i^{\min(a_i,b_i)}$, $[a,b] = \prod_i p_i^{\max(a_i,b_i)}$. Se $c = p_1^{c_1} \cdots p_k^{c_k}$ con c_i interi non negativi, allora

$$[a,(b,c)] = \prod_i p_i^{\max(a_i,\min(b_i,c_i))}, \qquad ([a,b],[a,c]) = \prod_i p_i^{\min(\max(a_i,b_i),\max(a_i,c_i))}.$$

Gli interi non negativi, col loro ordine naturale, sono totalmente ordinati, e in questo reticolo $\max(a_i,b_i) = a_i \vee b_i$, $\min(a_i,b_i) = a_i \wedge b_i$. La legge distributiva D$'$ dà allora

$$\max(a_i, \min(b_i,c_i)) = \min(\max(a_i,b_i), \max(a_i,c_i)).$$

Quindi $[a,(b,c)] = ([a,b],[a,c])$, che è D per gli interi positivi ordinati per divisibilità. Lo stesso ragionamento vale per ogni monoide fattoriale ridotto (cfr. § 2.14, p. 140).

Osserviamo inoltre che in ogni reticolo $a \wedge (b \vee c) \geq a \wedge b$ e $a \wedge (b \vee c) \geq a \wedge c$, dunque $a \wedge (b \vee c) \geq (a \wedge b) \vee (a \wedge c)$. Per stabilire la distributività basta quindi verificare la disuguaglianza opposta

$$a \wedge (b \vee c) \leq (a \wedge b) \vee (a \wedge c). \tag{1}$$

I reticoli più importanti dell'algebra, tra cui quelli dei sottomoduli di un modulo e dei sottogruppi normali di un gruppo, non sono in generale distributivi. Indichiamo per esempio con $L(V)$ il reticolo dei sottospazi di uno spazio vettoriale V su un campo F. Sia $\dim V \geq 2$ e siano x, y linearmente indipendenti. Si ha $F(x+y) \cap (Fx + Fy) = F(x+y)$, ma $F(x+y) \cap Fx = 0$ e $F(x+y) \cap Fy = 0$. Pertanto

$$F(x+y) \cap (Fx + Fy) \neq (F(x+y) \cap Fx) + (F(x+y) \cap Fy).$$

Vedremo tra poco che $L(V)$ soddisfa una forma più debole della distributività, formulata per la prima volta da Dedekind:

$$\text{M} \qquad \text{se } a \geq b, \quad a \wedge (b \vee c) = b \vee (a \wedge c).$$

Poiché $b = a \wedge b$, il secondo membro si può sostituire con $(a \wedge b) \vee (a \wedge c)$. La condizione M equivale quindi a D nel caso particolare $a \geq b$ (o $a \geq c$). Si chiama *modularità* e un reticolo che la soddisfa si dice *modulare*. La condizione duale M$'$ dice che, se $a \leq b$, allora $a \vee (b \wedge c) = b \wedge (a \vee c)$. È la stessa condizione M. Dunque il principio di dualità vale per i reticoli modulari, come per quelli distributivi.

L'importanza dei reticoli modulari in algebra deriva dal seguente teorema.

Teorema 8.3. *Il reticolo dei sottogruppi normali di un gruppo è modulare. Il reticolo dei sottomoduli di un modulo è modulare.*

Dimostrazione. Il sottogruppo normale generato da due sottogruppi normali H_1, H_2 di G è $H_1H_2 = H_2H_1$. Occorre dunque provare che, per tre sottogruppi normali H_1, H_2, H_3 con $H_1 \supseteq H_2$,

$$H_1 \cap (H_2H_3) = H_2(H_1 \cap H_3).$$

Per l'osservazione sulla legge distributiva basta provare l'inclusione da sinistra a destra. Se $a \in H_1 \cap (H_2H_3)$, allora $a = h_1 = h_2h_3$ con $h_i \in H_i$. Poiché $H_1 \supseteq H_2$, abbiamo $h_3 = h_2^{-1}h_1 \in H_1$. Quindi $h_3 \in H_1 \cap H_3$ e $a = h_2h_3 \in H_2(H_1 \cap H_3)$, come volevamo. Per i moduli il ragionamento è analogo e più semplice, e lo omettiamo. □

Una definizione alternativa di modularità, a volte utile, si ricava dal seguente teorema.

Teorema 8.4. *Un reticolo L è modulare se e solo se da $a \geq b$, $a \wedge c = b \wedge c$ e $a \vee c = b \vee c$ per un $c \in L$ segue $a = b$.*

Dimostrazione. Sia L modulare e siano a, b, c come nell'enunciato. Allora

$$a = a \wedge (a \vee c) = a \wedge (b \vee c) = b \vee (a \wedge c) = b \vee (b \wedge c) = b.$$

Viceversa, sia L un reticolo con la proprietà enunciata. Per $a, b, c \in L$ con $a \geq b$ sappiamo che $a \wedge (b \vee c) \geq b \vee (a \wedge c)$. Inoltre

$$(a \wedge (b \vee c)) \wedge c = a \wedge ((b \vee c) \wedge c) = a \wedge c$$

e

$$a \wedge c = (a \wedge c) \wedge c \leq (b \vee (a \wedge c)) \wedge c \leq a \wedge c.$$

Pertanto $(b \vee (a \wedge c)) \wedge c = a \wedge c$. Siccome $b \leq a$, la duale della prima relazione è $(b \vee (a \wedge c)) \vee c = b \vee c$ e la duale della seconda è $(a \wedge (b \vee c)) \vee c = b \vee c$. Abbiamo dunque

$$(a \wedge (b \vee c)) \wedge c = (b \vee (a \wedge c)) \wedge c,$$
$$(a \wedge (b \vee c)) \vee c = (b \vee (a \wedge c)) \vee c.$$

La proprietà ipotizzata implica $a \wedge (b \vee c) = b \vee (a \wedge c)$, cioè l'assioma modulare. □

Dimostriamo ora per i reticoli modulari un analogo del secondo teorema di isomorfismo dei gruppi (teorema 1.9, p. 65).

Teorema 8.5. *Se a, b sono elementi di un reticolo modulare, l'applicazione $x \mapsto x \wedge b$ è un isomorfismo dell'intervallo $I[a, a \vee b]$ su $I[a \wedge b, b]$. L'isomorfismo inverso è $y \mapsto y \vee a$.*

Dimostrazione. In ogni reticolo le applicazioni $x \mapsto x \vee a$ e $x \mapsto x \wedge a$ conservano l'ordine. Infatti $x \geq y$ se e solo se $x \vee y = x$, e anche se e solo se $x \wedge y = y$. Da $x \vee y = x$ segue

$$(x \vee a) \vee (y \vee a) = (x \vee y) \vee (a \vee a) = (x \vee y) \vee a = x \vee a.$$

Dunque $x \geq y$ implica $x \vee a \geq y \vee a$, e analogamente $x \wedge a \geq y \wedge a$. Se $a \leq x \leq a \vee b$, allora $a \wedge b \leq x \wedge b \leq b = (a \vee b) \wedge b$. Se $a \wedge b \leq y \leq b$, allora $a = a \vee (a \wedge b) \leq y \vee a \leq a \vee b$. Pertanto le due applicazioni mandano rispettivamente $I[a, a \vee b]$ in $I[a \wedge b, b]$ e $I[a \wedge b, b]$ in $I[a, a \vee b]$. Poiché conservano l'ordine, per il teorema 8.2 basta verificare che sono inverse. Sia $x \in I[a, a \vee b]$. Siccome $x \geq a$, per modularità $(x \wedge b) \vee a = x \wedge (a \vee b)$, e siccome $x \leq a \vee b$ ciò dà $(x \wedge b) \vee a = x$. Dualmente, se $y \in I[a \wedge b, b]$, allora $(y \vee a) \wedge b = y$. Questo prova la tesi. □

Il teorema suggerisce di introdurre per gli intervalli una nozione di equivalenza che nei reticoli modulari è più forte dell'isomorfismo. Diciamo che $I[u, v]$ e $I[w, t]$ sono *trasposti* se esistono a, b nel reticolo tali che uno dei due sia $I[a, a \vee b]$ e l'altro $I[a \wedge b, b]$. Si dicono *proiettivi* se esiste una successione finita

$$I[u, v] = I[u_1, v_1], I[u_2, v_2], \ldots, I[u_n, v_n] = I[w, t]$$

in cui ogni coppia di intervalli consecutivi $I[u_k, v_k], I[u_{k+1}, v_{k+1}]$ è formata da intervalli trasposti. È immediato che si tratta di una relazione di equivalenza. Dal teorema 8.5 risulta pure che in un reticolo modulare intervalli proiettivi sono isomorfi.

Esercizi

1. Mostrare che il reticolo dei sottogruppi di A_4 non è modulare.

2. Sia G un gruppo generato da x, y, con $x^{p^m} = 1$, $y^{p^r} = 1$, $y^{-1}xy = x^m$, dove $m^{p^r} \equiv 1 \pmod{p^m}$ e p è primo. Mostrare che, se H_1, H_2 sono sottogruppi di G, allora $H_1H_2 = H_2H_1$. Dedurne che il reticolo dei sottogruppi di G è modulare.[3]

3. Mostrare che un reticolo non distributivo contiene un sottoreticolo di cinque elementi il cui diagramma è il primo o il secondo dei diagrammi di p. 457 dell'originale (§ 8.1). Mostrare che ogni reticolo non modulare contiene un sottoreticolo con il secondo di tali diagrammi.

8.3 Il teorema di Jordan–Hölder–Dedekind

Un insieme parzialmente ordinato S si dice *di lunghezza finita* se le lunghezze delle sue catene, ossia dei suoi sottoinsiemi totalmente ordinati, sono limitate; qui la lunghezza di una catena è il numero dei suoi termini distinti. Se $a > b$ in un insieme parzialmente ordinato di lunghezza finita, esiste una successione finita $a = a_1, a_2, \ldots, a_n = b$ in cui ciascun a_i copre a_{i+1}. Una successione siffatta si chiama *catena satura* da a a b. È desiderabile che due catene sature qualsiasi da a a b abbiano la stessa lunghezza. Mostreremo che questo accade in un reticolo L di lunghezza finita se L è *semimodulare*, nel senso seguente: se $a \vee b$ copre sia a sia b, allora tanto a quanto b coprono $a \wedge b$. Se L è modulare, $I[a \wedge b, a]$ e $I[b, a \vee b]$ sono isomorfi, come abbiamo visto. È quindi chiaro che la modularità implica la semimodularità. Il teorema seguente è l'analogo, per i reticoli, del teorema di Jordan–Hölder per i gruppi finiti (p. 249).

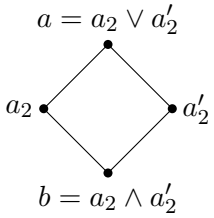
Teorema di Jordan–Hölder–Dedekind. *Sia L un reticolo semimodulare di lunghezza finita. Per $a > b$, due catene sature qualsiasi da a a b hanno la stessa lunghezza. Inoltre, se L è modulare e*

$$a = a_1 > a_2 > \cdots > a_{n+1} = b \tag{2}$$

$$a = a'_1 > a'_2 > \cdots > a'_{m+1} = b \tag{3}$$

sono due catene sature da a a b, gli intervalli $I[a_{i+1}, a_i]$ della prima e gli intervalli $I[a'_{j+1}, a'_j]$ della seconda possono essere abbinati in modo che gli intervalli di ogni coppia siano proiettivi.

Dimostrazione. Il ragionamento ricalca quello per i gruppi. Procediamo per induzione su n, dove $n+1$ è la lunghezza di una delle due catene sature. Se $n = 1$, a copre b e la tesi è evidente. Se $a_2 = a'_2$, abbiamo due catene sature da a_2 a b e la tesi segue per induzione. Supponiamo dunque $a_2 \neq a'_2$. Allora a_1 copre a_2 e a'_2, e quindi $a_2 \vee a'_2 = a_1$. Per semimodularità a_2 e a'_2 coprono $a''_3 = a_2 \wedge a'_2$. Inoltre $a''_3 \geq b$. Se $b = a''_3$, abbiamo il diagramma



[3] Come enunciata senza altre ipotesi, l'affermazione ammette controesempi per $p = 2$, fra cui il gruppo diedrale di ordine otto con l'azione di inversione.

In questo caso $m = n = 2$ e, se il reticolo è modulare, $I[a_2, a_1]$ e $I[b, a_2']$ sono trasposti, come pure $I[a_2', a_1]$ e $I[b, a_2]$. Se $a_3'' > b$, scegliamo una catena satura $a_3'', a_4'', \ldots, a_{q+1}'' = b$. Applichiamo l'induzione alle due catene $a_2, a_3, \ldots, a_{n+1} = b$ e $a_2, a_3'', \ldots, a_{q+1}'' = b$, e poi, usando $q = n$, alle catene $a_2', a_3'', \ldots, a_{q+1}'' = b$ e $a_2', a_3', \ldots, a_{m+1}' = b$. Nel caso modulare si usa inoltre il fatto che $I[a_2, a_1]$ e $I[a_3'', a_2']$ sono trasposti e che lo sono $I[a_2', a_1]$ e $I[a_3'', a_2]$, come nella dimostrazione per i gruppi. Lasciamo al lettore i dettagli rimanenti. □

Supponiamo ora L modulare, con minimo 0 e di lunghezza finita. Se $a_1 = a, a_2, \ldots, a_{n+1} = b$ è una catena satura da a a b, chiamiamo *lunghezza dell'intervallo* $I[b, a]$ il numero n, determinato univocamente da a e b. Denotiamo la lunghezza di $I[0, a]$ con $d(a)$ e la chiamiamo *rango* di a. Per $a \geq b$ si ha chiaramente

$$d(a) = d(b) + \text{lunghezza } I[b, a].$$

Per a, b qualsiasi di L risulta dunque

$$d(a \vee b) = d(a) + \text{lunghezza } I[a, a \vee b], \qquad d(b) = d(a \wedge b) + \text{lunghezza } I[a \wedge b, b].$$

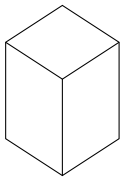
I due intervalli sono isomorfi e hanno quindi la stessa lunghezza. Ne segue $d(a \vee b) - d(a) = d(b) - d(a \wedge b)$, ovvero

$$d(a \vee b) = d(a) + d(b) - d(a \wedge b), \tag{4}$$

analoga alla formula delle dimensioni per i sottospazi di uno spazio vettoriale di dimensione finita.

Esercizi

1. Verificare che il reticolo con il diagramma seguente è semimodulare ma non modulare.



2. Osservare che, per definire il rango, basta che L abbia uno 0, sia di lunghezza finita e soddisfi la condizione di Jordan–Hölder–Dedekind, ossia la prima conclusione del teorema. Sia L un reticolo con 0 di lunghezza finita. Dimostrare che sono equivalenti: (i) L è modulare; (ii) L e il suo duale sono semimodulari; (iii) L soddisfa la condizione di Jordan–Hölder–Dedekind e la condizione (4) sul rango.

8.4 Il reticolo dei sottospazi di uno spazio vettoriale. Il teorema fondamentale della geometria proiettiva

Consideriamo anzitutto le proprietà fondamentali del reticolo $L(V)$ dei sottospazi di uno spazio vettoriale V. Supporremo V di dimensione finita. Poiché accresce considerevolmente la generalità senza aumentare la difficoltà, considereremo spazi vettoriali su anelli di divisione anziché, come abbiamo fatto abitualmente in questo libro, su campi. Sia dunque V uno spazio vettoriale (o modulo sinistro) su un anello di divisione Δ e sia $(e_1, \ldots, e_n)$ una sua base su Δ. Abbiamo già visto che $L(V)$ è modulare (teorema 8.3). Ogni sottospazio U ammette una base $(f_1, \ldots, f_m)$ con $m \leq n$, e $m = n$ se e solo se $U = V$. Ne segue che $L(V)$ è di lunghezza finita. Inoltre $L(V)$ ha massimo $1 = V$ e minimo $0 = 0$ (il sottospazio costituito dal solo vettore nullo). Un'altra

proprietà importante è l'esistenza di complementi: per ogni U esiste U' tale che $V = U + U'$ e $U \cap U' = 0$. Le due condizioni si abbreviano in $V = U \oplus U'$. Per dimostrare l'esistenza del complemento completiamo una base $(f_1, \dots, f_m)$ di U a una base $(f_1, \dots, f_m, f_{m+1}, \dots, f_n)$ di V. È immediato che per $U' = \sum_{j=m+1}^{n} \Delta f_j$ si ha $V = U \oplus U'$. Se $U \neq 0, V$, il complemento non è unico. Infatti $0 < m < n$ e anche $(f_1, \dots, f_m, f_1 + f_{m+1}, f_{m+2}, \dots, f_n)$ è una base, per cui $U'' = \Delta(f_1 + f_{m+1}) + \Delta f_{m+2} + \cdots + \Delta f_n$ è un complemento, chiaramente distinto da U'.

Un reticolo L con 0 e 1 si dice *complementato* se per ogni $a \in L$ esiste a' con $1 = a \vee a'$ e $a \wedge a' = 0$. L'elemento a' è un *complemento* di a. Abbiamo dunque mostrato che $L(V)$ è complementato.

Passiamo al problema dell'isomorfismo dei due reticoli $L(V_i)$, $i = 1, 2$, dove V_i è uno spazio vettoriale su un anello di divisione Δ_i. Vedremo, nell'ipotesi sulla dimensione precisata più avanti, che i reticoli sono isomorfi se e solo se lo sono i due anelli di divisione e gli spazi vettoriali hanno la stessa dimensione. Determineremo anche tutti gli isomorfismi di reticoli tra gli $L(V_i)$ quando ne esistono. Vari problemi, per esempio la determinazione degli automorfismi del gruppo delle applicazioni lineari biiettive di V su Δ, ossia del gruppo di matrici invertibili $\mathrm{GL}_n(\Delta)$, conducono a questi problemi sui reticoli.

Supponiamo dapprima $\Delta_1 \cong \Delta_2$ e $\dim V_1 = \dim V_2 = n$. Siano $(e_1, \dots, e_n)$ e $(f_1, \dots, f_n)$ basi rispettivamente di V_1 e V_2, e sia σ un isomorfismo di Δ_1 su Δ_2. Ogni $x \in V_1$ si scrive in un solo modo $x = \sum a_i e_i$, con $a_i \in \Delta_1$. Possiamo definire l'applicazione

$$\eta\colon \sum a_i e_i \longmapsto \sum \sigma(a_i) f_i.$$

Essa è additiva:

$$\eta(x + y) = \eta x + \eta y. \tag{5}$$

Per $a \in \Delta_1$ si ha $ax = \sum a a_i e_i$, dunque $\eta(ax) = \sum \sigma(a a_i) f_i = \sum \sigma(a)\sigma(a_i) f_i = \sigma(a)(\eta x)$. Pertanto

$$\eta(ax) = \sigma(a)\eta x, \qquad a \in \Delta_1, \quad x \in V_1. \tag{6}$$

Un'applicazione di V_1 su Δ_1 in V_2 su Δ_2 che soddisfa (5) e (6) si dice *semilineare con isomorfismo associato* σ, o *σ-semilineare*. Per l'applicazione η definita sopra l'inversa è $\sum b_i f_i \mapsto \sum \sigma^{-1}(b_i) e_i$ ed è σ^{-1}-semilineare.

Sia ora η una qualsiasi applicazione biiettiva σ-semilineare di V_1 su V_2. Affermiamo che induce un isomorfismo di $L(V_1)$ su $L(V_2)$. Se U_1 è un sottospazio di V_1, la sua immagine $\eta(U_1)$ è un sottospazio di V_2. Infatti due vettori qualunque di $\eta(U_1)$ sono della forma $\eta x, \eta y$ con $x, y \in U_1$ e $\eta x + \eta y = \eta(x + y) \in \eta(U_1)$ poiché $x + y \in U_1$. Se $b \in \Delta_2$, si ha $b = \sigma(a)$ per un $a \in \Delta_1$, per cui $b(\eta x) = \sigma(a)(\eta x) = \eta(ax) \in \eta(U_1)$ poiché $ax \in U_1$. Abbiamo dunque l'applicazione

$$[\eta]\colon U_1 \longmapsto \eta(U_1) \tag{7}$$

di $L(V_1)$ in $L(V_2)$. Evidentemente essa conserva l'ordine: se $U_1 \subseteq W_1$, allora $[\eta]U_1 \subseteq [\eta]W_1$. L'inversa η^{-1} è σ^{-1}-semilineare da V_2 su V_1, come si verifica, e induce $[\eta^{-1}]$ da $L(V_2)$ in $L(V_1)$. Per $U_i \in L(V_i)$ si ha $[\eta^{-1}][\eta]U_1 = U_1$ e $[\eta][\eta^{-1}]U_2 = U_2$. Dunque $[\eta]$ è biiettiva e, per il teorema 8.2, è un isomorfismo di reticoli. In conclusione, se Δ_1, Δ_2 sono isomorfi e V_1, V_2 hanno uguale dimensione, $L(V_1)$ e $L(V_2)$ sono isomorfi. Inoltre ogni applicazione semilineare biiettiva η di V_1 su V_2 dà mediante (7) un isomorfismo dei reticoli. Dimostreremo ora i viceversa, almeno quando le dimensioni sono maggiori o uguali a tre. È un risultato classico che apparve dapprima in geometria proiettiva in forma alquanto diversa.[4] In tale ambito prese il nome seguente.

[4] Cfr. E. Artin, *Geometric Algebra*, New York, Wiley, 1957, p. 88, oppure R. Baer, *Linear Algebra and Projective Geometry*, New York, Academic Press, 1952, p. 44.

Teorema fondamentale della geometria proiettiva. *Siano V_i, $i = 1, 2$, spazi vettoriali su anelli di divisione Δ_i. Se $L(V_1) \cong L(V_2)$ e $\dim V_1 \geq 3$, allora $\Delta_1 \cong \Delta_2$ e $\dim V_1 = \dim V_2$. Inoltre ogni isomorfismo di $L(V_1)$ su $L(V_2)$ è della forma $[\eta]$ di (7), dove η è un'applicazione semilineare biiettiva di V_1 su V_2.*

Dimostrazione (Artin). Prendiamo una base $(e_1, \dots, e_n)$ di V_1 e poniamo $V_{1i} = \sum_{j \geq i} \Delta_1 e_j$, $i = 1, \dots, n$. Allora

$$V_1 = V_{11} \supsetneq V_{12} \supsetneq \cdots \supsetneq V_{1n} \supsetneq 0 \tag{8}$$

è una catena satura in $L(V_1)$ da $1 = V_1$ a 0. Se ζ è un isomorfismo di $L(V_1)$ su $L(V_2)$, l'immagine di (8) è una catena satura

$$V_2 = V_{21} \supsetneq V_{22} \supsetneq \cdots \supsetneq V_{2n} \supsetneq 0.$$

I V_{2i} sono sottospazi e non ve ne sono di strettamente compresi tra V_{2i} e $V_{2,i+1}$. Dunque $\dim V_{2i} = \dim V_{2,i+1} + 1$ e $\dim V_2 = n$. Analogamente, se U è un sottospazio di dimensione m, $\dim \zeta(U) = m$. In particolare i sottospazi unidimensionali hanno immagine unidimensionale. Scriviamo $\zeta(\Delta_1 e_i) = \Delta_2 e_i' \neq 0$. Da $V_{1i} = \sum_{j \geq i} \Delta_1 e_j$ segue $V_{2i} = \sum_{j \geq i} \Delta_2 e_j'$ e $V_2 = \sum_{j=1}^n \Delta_2 e_j'$. Pertanto $(e_1', \dots, e_n')$ è una base di V_2. Sia $0 \neq a \in \Delta_1$. Allora $\Delta_1(e_1 + ae_2) \subseteq \Delta_1 e_1 + \Delta_1 e_2$ e questo sottospazio unidimensionale è distinto da entrambe le rette coordinate. Dunque $\zeta(\Delta_1(e_1 + ae_2)) = \Delta_2(e_1' + a'e_2')$, con $a' \neq 0$, e a' è unico poiché $\Delta_2(e_1' + a'e_2') \neq \Delta_2(e_1' + b'e_2')$ per $a' \neq b'$. Si ottiene un'applicazione $a \mapsto a'$ che estendiamo ponendo $0' = 0$. Sostituendo a e_2' il vettore $1'e_2'$ possiamo supporre $1' = 1$. Analogamente otteniamo un'applicazione $a \mapsto a''$ di Δ_1 in Δ_2 tale che $\Delta_1(e_1 + ae_3) \mapsto \Delta_2(e_1' + a''e_3')$, $0'' = 0$, $1'' = 1$. Affermiamo che $a' = a''$. Infatti per l'indipendenza lineare di e_1, e_2, e_3, se $a \neq 0$,

$$\Delta_1(e_2 - e_3) = \big(\Delta_1(e_1 + ae_2) + \Delta_1(e_1 + ae_3)\big) \cap (\Delta_1 e_2 + \Delta_1 e_3).$$

L'immagine di $\Delta_1(e_2 - e_3)$ è l'intersezione

$$\big(\Delta_2(e_1' + a'e_2') + \Delta_2(e_1' + a''e_3')\big) \cap (\Delta_2 e_2' + \Delta_2 e_3'),$$

che contiene $a'e_2' - a''e_3'$. Quindi $\Delta_1(e_2 - e_3) \mapsto \Delta_2(a'e_2' - a''e_3')$. Il primo membro è indipendente da a, e $1' = 1'' = 1$. Ne segue $a'' = a'$. Allo stesso modo si ottiene

$$\Delta_1(e_1 + ae_i) \longmapsto \Delta_2(e_1' + a'e_i'), \qquad 2 \leq i \leq n. \tag{9}$$

Proviamo quindi per induzione su $r = 2, \dots, n$ che

$$\Delta_1(e_1 + a_2e_2 + \cdots + a_re_r) \longmapsto \Delta_2(e_1' + a_2'e_2' + \cdots + a_r'e_r'). \tag{10}$$

Supponiamolo vero per r e consideriamo $\Delta_1(e_1 + a_2e_2 + \cdots + a_{r+1}e_{r+1})$. È l'intersezione di $\Delta_1(e_1 + a_2e_2 + \cdots + a_re_r) + \Delta_1 e_{r+1}$ e $\Delta_1(e_1 + a_{r+1}e_{r+1}) + \Delta_1 e_2 + \cdots + \Delta_1 e_r$. La sua immagine è dunque l'intersezione di $\Delta_2(e_1' + a_2'e_2' + \cdots + a_r'e_r') + \Delta_2 e_{r+1}'$ e $\Delta_2(e_1' + a_{r+1}'e_{r+1}') + \Delta_2 e_2' + \cdots + \Delta_2 e_r'$; questa è proprio $\Delta_2(e_1' + a_2'e_2' + \cdots + a_{r+1}'e_{r+1}')$. La (10) vale così per ogni r e in particolare per $r = n$. L'intersezione di $\Delta_1(e_1 + a_2e_2 + \cdots + a_ne_n) + \Delta_1 e_1$ e $\Delta_1 e_2 + \cdots + \Delta_1 e_n$ è $\Delta_1(a_2e_2 + \cdots + a_ne_n)$. Con un ragionamento dello stesso tipo si ottiene quindi

$$\Delta_1(a_2e_2 + \cdots + a_ne_n) \longmapsto \Delta_2(a_2'e_2' + \cdots + a_n'e_n').$$

Possiamo ora provare che $a \mapsto a'$ è un isomorfismo. Si osservi che

$$\Delta_1(e_1 + (a + b)e_2 + e_3) \subseteq \Delta_1(e_1 + ae_2) + \Delta_1(be_2 + e_3).$$

Pertanto $\Delta_2(e'_1 + (a+b)'e'_2 + e'_3) \subseteq \Delta_2(e'_1 + a'e'_2) + \Delta_2(b'e'_2 + e'_3)$. Il solo vettore di quest'ultimo spazio della forma $e'_1 + ce'_2 + e'_3$ è $e'_1 + (a' + b')e'_2 + e'_3$. Segue $(a+b)' = a' + b'$. Analogamente, usando $\Delta_1(e_1 + abe_2 + ae_3) \subseteq \Delta_1 e_1 + \Delta_1(be_2 + e_3)$, si ottiene $(ab)' = a'b'$. L'applicazione è dunque un omomorfismo e, poiché Δ_1 è un anello di divisione e $1' = 1$, è un monomorfismo. I sottospazi unidimensionali di V_1 hanno una delle due forme $\Delta_1(e_1 + a_2e_2 + \cdots + a_ne_n)$ oppure $\Delta_1(a_2e_2 + \cdots + a_ne_n)$, e le rispettive immagini sono $\Delta_2(e'_1 + a'_2e'_2 + \cdots + a'_ne'_n)$ e $\Delta_2(a'_2e'_2 + \cdots + a'_ne'_n)$. Poiché ζ è suriettiva, per ogni $b \in \Delta_2$ il sottospazio $\Delta_2(e'_1 + be'_2)$ è immagine di un sottospazio unidimensionale di V_1, necessariamente del primo tipo. Ne segue $\Delta_2(e'_1 + a'_2e'_2 + \cdots) = \Delta_2(e'_1 + be'_2)$, dunque $b = a'_2$ e $a \mapsto a'$ è un epimorfismo. Abbiamo così un isomorfismo σ di Δ_1 su Δ_2 e un'applicazione semilineare biiettiva $\eta\colon \sum a_ie_i \mapsto \sum a'_ie'_i$. Le applicazioni ζ e $[\eta]$ coincidono sui sottospazi unidimensionali; siccome ogni sottospazio è somma di sottospazi unidimensionali, $\zeta = [\eta]$. □

Osservazioni. L'ipotesi $\dim V_1 \geq 3$ è essenziale nella conclusione principale. Vediamo che cosa succede se la dimensione è al più due. Il reticolo $L(V)$ ha esattamente due elementi se e solo se $\dim V = 1$. Se $\dim V = 2$, due sottospazi distinti U_1, U_2, diversi da $0, V$, hanno intersezione nulla e somma V. Quindi, se $\dim V_1 = \dim V_2 = 2$, ogni biiezione fra gli insiemi dei sottospazi unidimensionali si estende, ponendo $V_1 \mapsto V_2$, $0 \mapsto 0$, a un isomorfismo di $L(V_1)$ su $L(V_2)$. Per spazi non nulli con $\dim V_1 \leq 2$ si ha dunque $L(V_1) \cong L(V_2)$ se e solo se o $\dim V_1 = \dim V_2 = 1$, oppure $\dim V_1 = \dim V_2 = 2$ e $|L(V_1)| = |L(V_2)|$. In dimensione due quest'ultima condizione equivale a $|\Delta_1| = |\Delta_2|$.

Consideriamo il caso $V = V_1 = V_2$ del teorema fondamentale. Gli automorfismi del reticolo $L(V)$ formano un gruppo di trasformazioni di $L(V)$. Abbiamo inoltre il gruppo $\mathrm{GS}(V)$ delle trasformazioni semilineari biiettive di V. Infatti, se $\eta_1\colon V_1 \to V_2$ è σ-semilineare e $\eta_2\colon V_2 \to V_3$ è τ-semilineare, la composta $\eta_2\eta_1$ è $\tau\sigma$-semilineare. Se η_1 è biiettiva, η_1^{-1} è σ^{-1}-semilineare. Ne segue che l'insieme $\mathrm{GS}(V)$ delle trasformazioni semilineari biiettive di uno spazio n-dimensionale su Δ è un gruppo di trasformazioni. Se $0 \neq a \in \Delta$, la moltiplicazione scalare $x \mapsto ax$ soddisfa $a(bx) = (aba^{-1})ax$ ed è perciò una trasformazione semilineare biiettiva associata all'automorfismo interno $b \mapsto aba^{-1}$ di Δ. Essa induce l'identità su $L(V)$. Abbiamo l'omomorfismo $\eta \mapsto [\eta]$ da $\mathrm{GS}(V)$ nel gruppo degli automorfismi del reticolo $L(V)$. Per il teorema fondamentale della geometria proiettiva, esso è suriettivo se $\dim V \geq 3$. Il nucleo contiene tutte le moltiplicazioni scalari non nulle, come appena visto; d'altra parte il ragionamento usato a p. 378 mostra che queste costituiscono l'intero nucleo. Indicando quest'ultimo con Δ_L^*, otteniamo il seguente risultato.

Corollario. *Se* $\dim V \geq 3$*, il gruppo degli automorfismi del reticolo dei sottospazi di* V *su un anello di divisione* Δ *è isomorfo a* $\mathrm{GS}(V)/\Delta_L^*$*, dove* $\mathrm{GS}(V)$ *è il gruppo delle trasformazioni semilineari biiettive di* V *e* Δ_L^* *è l'insieme delle moltiplicazioni scalari non nulle.*

Esercizi

1. Un *anti-isomorfismo* di un reticolo L su L' è una biiezione $a \mapsto a'$ tale che $(a \wedge b)' = a' \vee b'$ e $(a \vee b)' = a' \wedge b'$. Osservare che è lo stesso che un isomorfismo del duale di L su L'; per il teorema 8.2 un anti-isomorfismo è dunque una biiezione che inverte l'ordine e la cui inversa inverte anch'essa l'ordine ($a \leq b \iff a' \geq b'$). Sia V uno spazio vettoriale di dimensione finita su un anello di divisione Δ e sia V^* lo spazio vettoriale destro delle funzioni lineari su V. Per un sottospazio U poniamo $\operatorname{ann} U = \{f \in V^* \mid f(y) = 0 \text{ per ogni } y \in U\}$. Provare che $U \mapsto \operatorname{ann} U$ è un anti-isomorfismo di $L(V)$ su $L(V^*)$.

2. Mostrare che, se $\dim V_1 \geq 3$, i reticoli dei sottospazi di V_1, V_2 sono anti-isomorfi se e solo se sono anti-isomorfi gli anelli di divisione sottostanti e le dimensioni sono uguali.

3. Mostrare che, se $\dim V \geq 3$ e $L(V)$ ammette un antiautomorfismo ζ, allora Δ ammette un antiautomorfismo $a \mapsto \bar{a}$ ed esiste un'applicazione $g\colon V \times V \to \Delta$ con le seguenti proprietà:

$$\begin{array}{ll} \text{(i)} & g(x_1 + x_2, y) = g(x_1, y) + g(x_2, y), \\ \text{(ii)} & g(x, y_1 + y_2) = g(x, y_1) + g(x, y_2), \\ \text{(iii)} & g(ax, y) = ag(x, y), \\ \text{(iv)} & g(x, ay) = \overline{g(x, y)}\bar{a}, \\ \text{(v)} & g(y, x) = \overline{g(x, Qy)}. \end{array}$$

Qui Q è biiettiva e σ-semilineare, con σ l'inversa di $a \mapsto \bar{a}$, secondo la notazione dell'originale. (vi) g è non degenere nel senso che $g(z, x) = 0$ per ogni x se e solo se $z = 0$. (vii) Per ogni sottospazio U, $\zeta(U) = \{v \in V \mid g(v, u) = 0 \text{ per ogni } u \in U\}$.

4. Sia V di dimensione due su un campo F di q elementi. Contare i sottospazi unidimensionali di V e determinare l'ordine del gruppo $S_2(F)$ delle trasformazioni semilineari biiettive. Dedurne che esistono automorfismi di $L(V)$ che non provengono da trasformazioni semilineari biiettive di V.

5. Sia V uno spazio di dimensione tre su $\mathbb{Z}/(p)$, dove p è primo. Determinare il numero degli automorfismi del reticolo $L(V)$.

8.5 Algebre di Boole

Definizione 8.3. *Un'algebra di Boole*[5] *è un reticolo distributivo e complementato, dotato di massimo* 1 *e minimo* 0.

Gli esempi più importanti sono i reticoli dei sottoinsiemi di un insieme S. Più generalmente, ogni *campo di sottoinsiemi* di S è un'algebra di Boole, ossia ogni famiglia di sottoinsiemi chiusa rispetto a unione e intersezione, che contiene S, $\varnothing$ e il complementare di ciascuno dei suoi membri.

Il teorema seguente raccoglie le principali proprietà elementari dei complementi.

Teorema 8.6. *In un'algebra di Boole B ogni elemento a ha un unico complemento a'. L'applicazione $a \mapsto a'$ è un antiautomorfismo di periodo al più due. Si hanno*

$$(a \vee b)' = a' \wedge b', \qquad (a \wedge b)' = a' \vee b', \tag{11}$$

$$a'' = a. \tag{12}$$

Dimostrazione. Siano $a \in B$ e a', a_1 tali che $a \vee a' = 1$, $a \wedge a_1 = 0$. Allora

$$a_1 = a_1 \wedge 1 = a_1 \wedge (a \vee a') = (a_1 \wedge a) \vee (a_1 \wedge a') = a_1 \wedge a'.$$

Se inoltre $a \vee a_1 = 1$, $a \wedge a' = 0$, si ha anche $a' = a' \wedge a_1$ e quindi $a' = a_1$. Il complemento è perciò unico. È chiaro che a è il complemento di a'. Dunque $a'' = (a')' = a$, per cui l'applicazione ha periodo uno o due ed è biiettiva. Sia ora $a \leq b$. Allora $a \wedge b' \leq b \wedge b' = 0$, quindi

$$b' = b' \wedge 1 = b' \wedge (a \vee a') = (b' \wedge a) \vee (b' \wedge a') = b' \wedge a'.$$

Ne segue $b' \leq a'$. Poiché l'applicazione coincide con la propria inversa e inverte l'ordine, per il teorema 8.2 (si veda l'esercizio 1, p. 473) è un anti-isomorfismo del reticolo. □

[5] Per evitare il conflitto con la nozione tecnica di algebra sarebbe preferibile «reticolo booleano». Abbiamo adottato «algebra di Boole» perché è il termine più diffuso.

Storicamente, le algebre di Boole furono i primi reticoli a essere studiati. Boole le introdusse per formalizzare il calcolo delle proposizioni. Per molto tempo si pensò che l'algebra di questi sistemi fosse di natura diversa da quella dei sistemi numerici e delle loro generalizzazioni, ossia anelli e algebre nel senso tecnico. Solo piuttosto tardi M. H. Stone scoprì che non è così: ogni algebra di Boole, opportunamente considerata, diventa un anello di tipo particolare.

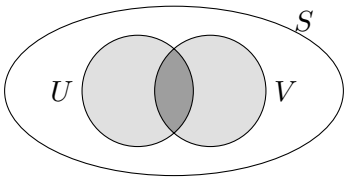
Per fare di B un anello introduciamo l'operazione

$$a + b = (a \wedge b') \vee (a' \wedge b),$$

detta *differenza simmetrica* di a, b. Si ha

$$\begin{aligned}(a \vee b) \wedge (a \wedge b)' &= (a \vee b) \wedge (a' \vee b') \\ &= ((a \vee b) \wedge a') \vee ((a \vee b) \wedge b') \\ &= ((a \wedge a') \vee (b \wedge a')) \vee ((a \wedge b') \vee (b \wedge b')) \\ &= (b \wedge a') \vee (a \wedge b') = a + b.\end{aligned} \tag{13}$$

La prima espressione mostra che nell'algebra dei sottoinsiemi $U + V$ è l'insieme degli elementi che appartengono a U oppure a V, ma non a entrambi.



Mostriamo che B è un anello per l'addizione appena definita, il prodotto $ab = a \wedge b$ e l'unità 1. L'addizione è evidentemente commutativa. Per l'associatività osserviamo anzitutto che, per (13),

$$(a + b)' = (a \vee b)' \vee (a \wedge b) = (a \wedge b) \vee (a' \wedge b').$$

Pertanto

$$\begin{aligned}(a + b) + c &= [((a \wedge b') \vee (a' \wedge b)) \wedge c'] \vee [((a \wedge b) \vee (a' \wedge b')) \wedge c] \\ &= [(a \wedge b' \wedge c') \vee (a' \wedge b \wedge c')] \vee [(a \wedge b \wedge c) \vee (a' \wedge b' \wedge c)] \\ &= (a \wedge b' \wedge c') \vee (a' \wedge b \wedge c') \vee (a \wedge b \wedge c) \vee (a' \wedge b' \wedge c).\end{aligned}$$

L'espressione è simmetrica in a, b, c. In particolare $(a + b) + c = (c + b) + a$, da cui per commutatività segue l'associatività di $+$. Evidentemente $a + 0 = (a \wedge 1) \vee (a' \wedge 0) = a$ e $a + a = (a \wedge a') \vee (a' \wedge a) = 0$. Dunque $(B, +, 0)$ è un gruppo commutativo.

Il prodotto $\cdot = \wedge$ è associativo e commutativo, e $a \cdot 1 = 1 \cdot a = a \wedge 1 = a$ per ogni $a \in B$. Resta da verificare una delle leggi distributive. Si ha

$$\begin{aligned}(a + b)c &= ((a \wedge b') \vee (a' \wedge b)) \wedge c = (a \wedge b' \wedge c) \vee (a' \wedge b \wedge c), \\ ac + bc &= ((a \wedge c) \wedge (b \wedge c)') \vee ((a \wedge c)' \wedge (b \wedge c)) \\ &= ((a \wedge c) \wedge (b' \vee c')) \vee ((a' \vee c') \wedge (b \wedge c)) \\ &= (a \wedge c \wedge b') \vee (a' \wedge b \wedge c).\end{aligned}$$

Il confronto dà $(a + b)c = ac + bc$. Ne segue che $(B, +, \cdot, 0, 1)$ è un anello.

Abbiamo visto che l'anello è commutativo e ogni elemento ha ordine additivo al più due. Inoltre ogni elemento è idempotente: $a^2 = a \wedge a = a$. Queste proprietà non sono indipendenti.

Infatti, se tutti gli elementi di un anello sono idempotenti, allora esso è commutativo e $2a = 0$ per ogni a. Per dimostrarlo osserviamo che

$$a + b + ab + ba = a^2 + b^2 + ab + ba = (a + b)^2 = a + b.$$

Dunque $ab + ba = 0$. Ne segue $2a = 2a^2 = aa + aa = 0$, ossia $a = -a$, e allora $ab = -ba = ba$.

Definizione 8.4. *Un anello si dice booleano se tutti i suoi elementi sono idempotenti.*

Abbiamo visto che un tale anello, se non è nullo, ha caratteristica due. Proveremo che ogni anello booleano definisce un'algebra di Boole e che i due concetti sono equivalenti. Dato un anello booleano $(B, +, \cdot, 0, 1)$, per invertire il procedimento precedente definiamo

$$a \vee b = a + b - ab = 1 - (1 - a)(1 - b).$$

La seconda espressione mostra che, introducendo $\sigma\colon x \mapsto 1 - x$, vale $a \vee b = \sigma^{-1}(\sigma(a)\sigma(b))$, poiché $\sigma^2 = 1$. Dall'associatività della moltiplicazione segue quindi quella di $\vee$, e naturalmente l'operazione è commutativa. Inoltre $a \vee a = 2a - a^2 = -a^2 = a$. Definiamo $a \wedge b = ab$. L'associatività e la commutatività sono evidenti, e $a \wedge a = a$ perché ogni elemento è idempotente. Infine $(a \vee b) \wedge a = (a + b - ab)a = a$ e $(a \wedge b) \vee a = ab + a - a^2 b = a$. Valgono dunque le condizioni L1–L4 che definiscono un reticolo. È immediato che 1 e 0 dell'anello sono il massimo e il minimo del reticolo e che $1 - a$ è un complemento di a, poiché $a \vee (1 - a) = 1$ e $a \wedge (1 - a) = 0$. Il reticolo è distributivo:

$$(a \vee b) \wedge c = (a + b - ab)c = ac + bc - abc = ac + bc - acbc = (a \wedge c) \vee (b \wedge c).$$

Quindi $(B, \vee, \wedge, 0, 1, ')$ è un'algebra di Boole.

Resta da provare che i due procedimenti sono inversi. Partiamo da un'algebra di Boole $(B, \vee, \wedge, 0, 1, ')$. Otteniamo l'anello in cui $a + b = (a \wedge b') \vee (a' \wedge b)$ e $ab = a \wedge b$. Applicando il secondo procedimento, 1 e 0 sono gli stessi, il complemento è $a' = 1 - a$ e le nuove operazioni, denotate con $\bar\vee$ e $\bar\wedge$, sono

$$a \mathbin{\bar\vee} b = a + b - ab = 1 - (1 - a)(1 - b) = (a' \wedge b')' = a \vee b, \qquad a \mathbin{\bar\wedge} b = ab = a \wedge b.$$

Riotteniamo quindi l'algebra originaria. Viceversa, partendo da un anello booleano otteniamo l'algebra in cui $a \vee b = a + b - ab$, $a \wedge b = ab$, $0 = 0$, $1 = 1$, $a' = 1 - a$. Applicando il primo procedimento, la nuova addizione $\oplus$ e il nuovo prodotto $\odot$ sono

$$\begin{aligned} a \oplus b &= (a \wedge (1 - b)) \vee ((1 - a) \wedge b) \\ &= a(1 - b) \vee (1 - a)b = (a - ab) \vee (b - ab) \\ &= a - ab + b - ab - (a - ab)(b - ab) \\ &= a - ab + b - ab - ab + ab + ab - ab = a + b, \\ a \odot b &= a \wedge b = ab. \end{aligned}$$

Anche 1 e 0 coincidono e ritroviamo l'anello iniziale. È così provato il seguente teorema di Stone.

Teorema 8.7. *Le strutture astratte di algebra di Boole e di anello booleano sono equivalenti.*

Un'ulteriore osservazione. Nel passare da un'algebra di Boole a un anello booleano potremmo scambiare nella costruzione $\vee$ con $\wedge$ e 1 con 0, per il principio di dualità, applicabile alle algebre di Boole. Il procedimento conduce allora a un anello B' con lo stesso insieme sottostante B,

addizione $a +' b = (a \vee b') \wedge (a' \vee b)$ e prodotto $a \cdot' b = a \vee b$. Il nuovo zero è $0' = 1$ e la nuova unità è $1' = 0$. In termini dell'anello B si ha

$$\begin{aligned} a +' b &= (a + (1 - b) - a + ab)(b + (1 - a) - b + ab) \\ &= (1 - b + ab)(1 - a + ab) = 1 - (a + b), \\ a \cdot' b &= a + b - ab. \end{aligned}$$

Un *ideale* di un'algebra di Boole B è un ideale dell'anello booleano associato $(B, +, \cdot, 0, 1)$. Perché un sottoinsieme non vuoto I sia un ideale occorre e basta che (1) $u, v \in I$ implichi $u + v \in I$ e (2) per ogni $u \in I$ e ogni $a \in B$ si abbia $ua \in I$. Poiché $ua = u \wedge a$ ed è uguale ad a se e solo se $a \leq u$, la seconda condizione equivale alla seguente: per ogni $u \in I$ appartiene a I ogni $b \leq u$. Poiché $u \vee v = u + v + uv$, abbiamo $u \vee v \in I$ per $u, v \in I$. Viceversa, sia I chiuso rispetto a $\vee$ e contenga ogni $b \leq u$ per ogni $u \in I$. Allora $u \wedge v'$ e $v \wedge u'$ appartengono a I, e così pure $u + v = (u \wedge v') \vee (v \wedge u')$. Dunque un sottoinsieme non vuoto di un'algebra di Boole è un ideale se e solo se è chiuso per $\vee$ e contiene con ogni elemento tutti quelli inferiori.

Un ideale I è *proprio* se $I \neq B$, ossia se e solo se $1 \notin I$. Per $u \in B$, $(u) = \{x \in B \mid x \leq u\}$ è l'*ideale principale* generato da u. Un ideale I è *massimale* se è proprio e nessun ideale proprio lo contiene strettamente. Osserviamo che I è massimale se e solo se è proprio e per ogni $a \in B$ appartiene a I almeno uno degli elementi a, a'. Supponiamo infatti I massimale e $a \notin I$. L'insieme $\bar{I}$ degli elementi $u + b$ con $u \in I$ e $b \leq a$ è un ideale che contiene strettamente I. Per massimalità coincide con B. Si ha quindi $1 = b + u$ con $b \leq a$ e $u \in I$, da cui $b' = 1 + b = u \in I$. Poiché $a' \leq b'$, anche $a' \in I$. Viceversa, sia I proprio e contenga per ogni a uno fra a e a'. Se un ideale $\bar{I}$ contiene strettamente I, prendiamo $a \in \bar{I}$, $a \notin I$. Allora $a' \in I \subseteq \bar{I}$, per cui $1 = a + a' \in \bar{I}$. Quindi $\bar{I} = B$ e I è massimale.

Possiamo passare al duale applicando le medesime considerazioni al secondo anello $B' = (B, +', \cdot', 0', 1')$ associato a B. Un *filtro* (o *ideale duale*) di B è per definizione un ideale di B'. Il criterio duale del precedente dice che un sottoinsieme non vuoto F è un filtro se e solo se è chiuso per $\wedge$ e contiene con ogni $u \in F$ ogni $b \geq u$. Dalle relazioni $(a \wedge b)' = a' \vee b'$ e $(a \vee b)' = a' \wedge b'$ si vede che F è un filtro se e solo se l'insieme F' dei complementi a', $a \in F$, è un ideale. La chiusura per $\wedge$ equivale alla chiusura rispetto alle intersezioni finite. Un filtro è *proprio*, ossia diverso da B, se e solo se $0 \notin F$. Un ideale massimale di B' si dice *ultrafiltro* di B. Un filtro F è un ultrafiltro se e solo se (1) $0 \notin F$ e (2) per ogni $a \in B$ si ha $a \in F$ oppure $a' \in F$. Gli elementi $x \geq a$ formano il *filtro principale* generato da a. Concludiamo questa breve introduzione con due esempi di filtri.

Esempi

1. Sia $\mathbb{R}$ la retta reale con la topologia usuale, e sia S la famiglia dei suoi aperti non vuoti. Il testo originale afferma che tale famiglia ha la proprietà dell'intersezione finita e che la famiglia $\bar{S}$ dei sottoinsiemi di $\mathbb{R}$ che contengono un aperto non vuoto è un filtro.[6]

2. Sia S un insieme qualsiasi, $B = \mathcal{P}(S)$, e sia I l'insieme dei sottoinsiemi finiti di S. È un ideale di B, quindi i complementi dei sottoinsiemi finiti formano un filtro.

[6] L'esempio, così stampato, non vale: due intervalli aperti disgiunti hanno intersezione vuota. Si ottiene un esempio corretto prendendo gli aperti che contengono un punto fissato p. I loro sovrainsiemi costituiscono il filtro degli intorni di p.

Esercizi

1. Se e, f sono due idempotenti di un anello e commutano, mostrare che ef ed $e \circ f = e+f-ef$ sono idempotenti. Provare che gli idempotenti contenuti nel centro formano un'algebra di Boole rispetto a $e \vee f = e + f - ef$, $e \wedge f = ef$, $e' = 1 - e$.

2. Provare che, se R è un anello in cui $pa = 0$ e $a^p = a$ per ogni $a \in R$, dove p è un primo, allora R è commutativo.

3. Mostrare che la cardinalità di un'algebra di Boole finita è una potenza di due.

4. (Seligman). Siano $e_1, \dots, e_n$ idempotenti a due a due commutanti di un anello R, e sia $s = \sum_{i=1}^n e_i$. Mostrare che $\prod_{j=0}^n (s - j\,1_R) = 0$.[7]

8.6 La funzione di Möbius di un insieme parzialmente ordinato

Daremo un'applicazione degli insiemi parzialmente ordinati a problemi di enumerazione.[8] Nei problemi considerati compare una somma su un insieme parzialmente ordinato. Invertendo questa somma si ottiene la formula di enumerazione cercata. Eccone un esempio.

Problema 1. Vogliamo contare gli *scompigli*, ossia le permutazioni di un insieme finito S prive di punti fissi. Per un sottoinsieme T di S, indichiamo con $f(T)$ il numero delle permutazioni di S che fissano tutti gli elementi di T e nessun elemento del suo complementare T'. Indichiamo con $g(T)$ il numero di quelle che fissano tutti gli elementi di T ed eventualmente altri elementi. Allora

$$g(T) = \sum_{U \supseteq T} f(U), \tag{14}$$

dove, naturalmente, $U \in \mathcal{P}(S)$. Vogliamo «invertire» (14), ossia esprimere $f(U)$ in funzione dei $g(T)$. Ne ricaveremo $f(\varnothing)$, il numero degli scompigli di S, poiché $g(T)$ è semplicemente il numero delle permutazioni di T', cioè $|T'|!$.

Più in generale, consideriamo un insieme parzialmente ordinato finito S e due funzioni f, g su S a valori in un gruppo commutativo A, tali che

$$g(y) = \sum_{\substack{x \in S \\ x \geq y}} f(x), \qquad y \in S. \tag{15}$$

Anche qui vogliamo esprimere f mediante g. Ci serve il lemma seguente.

Lemma (Szpilrajn–Marczewski). *Si possono ordinare totalmente gli elementi di S, diciamo $x_1, x_2, \dots, x_n$, in modo che $x_i < x_j$ nell'ordine parziale originario implichi $i < j$.*

Dimostrazione. Poiché S è finito contiene un elemento minimale x_1. Proseguiamo scegliendo induttivamente x_{i+1} minimale nel complementare di $\{x_1, \dots, x_i\}$. L'ordine $x_1, \dots, x_n$ così ottenuto ha la proprietà richiesta. Infatti, se $x_i < x_j$, nella costruzione x_j non può essere stato scelto prima di x_i. Dunque $i < j$. □

Per $x, y \in S$ definiamo ora la funzione di due variabili a valori interi

$$\zeta(x, y) = \begin{cases} 1 & \text{se } x \leq y, \\ 0 & \text{altrimenti.} \end{cases} \tag{16}$$

[7] Nella stampa compare $s - e_j$, con e_0 non definito; la formula usuale riguarda le differenze $s - j\,1_R$.

[8] Sono debitore a Neil White per il materiale di questo paragrafo.

Con l'ordine totale del lemma, la (15) si può scrivere come sistema di equazioni $g(x_i) = \sum_{j=1}^{n} \zeta(x_i, x_j) f(x_j)$ o in forma matriciale

$$\begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \\ g(x_n) \end{pmatrix} = \begin{pmatrix} \zeta(x_1,x_1) & \zeta(x_1,x_2) & \cdots & \zeta(x_1,x_n) \\ \zeta(x_2,x_1) & \zeta(x_2,x_2) & \cdots & \zeta(x_2,x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \zeta(x_n,x_1) & \zeta(x_n,x_2) & \cdots & \zeta(x_n,x_n) \end{pmatrix} \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}. \tag{17}$$

Ricordiamo che $f(x), g(x)$ appartengono al gruppo abeliano A, che si può riguardare in modo naturale come uno $\mathbb{Z}$-modulo (p. 166). Si ha $\zeta(x_i, x_i) = 1$. Se $i > j$ non può aversi $x_i \le x_j$, dunque $\zeta(x_i, x_j) = 0$. La matrice $Z = (\zeta_{ij}) = (\zeta(x_i, x_j))$ è triangolare superiore con tutti 1 sulla diagonale: è della forma $1 - N$, dove

$$N = \begin{pmatrix} 0 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix}. \tag{18}$$

Si ha $N = (\nu_{ij})$ con $\nu_{ij} = 0$ se $i \ge j$ e $\nu_{ij} = -\zeta_{ij}$ se $i < j$. Per induzione si vede subito che l'elemento (i, j) di N^k è zero per $i \ge j - k + 1$. Dunque $N^n = 0$ e $Z = 1 - N$ è invertibile con inversa

$$M = 1 + N + N^2 + \cdots + N^{n-1}. \tag{19}$$

In forma abbreviata (17) si scrive $G = ZF$, dove G, F sono i vettori colonna (matrici $n \times 1$) $(g(x_i)), (f(x_i))$. Invertendo otteniamo $F = MG$. Se $M = (\mu_{ij})$, abbiamo dunque

$$f(x_i) = \sum_j \mu_{ij} g(x_j). \tag{20}$$

La matrice M definisce la *funzione di Möbius* dell'insieme parzialmente ordinato S, a valori in $\mathbb{Z}$, mediante

$$\mu(x_i, x_j) = \mu_{ij} \qquad \text{per ogni } x_i, x_j. \tag{21}$$

La (20) si riscrive allora

$$f(y) = \sum_{x \in S} \mu(y, x) g(x). \tag{20$'$}$$

Teorema 8.8. *Per ogni insieme parzialmente ordinato finito S esiste un'unica funzione $\mu\colon S \times S \to \mathbb{Z}$ tale che, se A è un gruppo commutativo e $f, g\colon S \to A$ soddisfano la (15), ossia $g(y) = \sum_{x \ge y} f(x)$ per ogni $y \in S$, allora*

$$f(y) = \sum_{x \in S} \mu(y, x) g(x), \qquad y \in S. \tag{22}$$

Dimostrazione. Abbiamo provato l'esistenza. Per l'unicità possiamo supporre gli x_i ordinati come nel lemma di Szpilrajn–Marczewski. Prendiamo $A = (\mathbb{Z}, +)$ e sia $\delta_k\colon S \to \mathbb{Z}$ la funzione tale che $\delta_k(x_i) = \delta_{ik}$, uguale a 1 per $i = k$ e a 0 altrimenti. Sia ε_k la funzione corrispondente definita dalla (15), o equivalentemente dalla (17). Allora

$$\varepsilon_k(x_i) = \sum_j \zeta(x_i, x_j) \delta_k(x_j) = \zeta(x_i, x_k).$$

Per (20), $\delta_{ik} = \delta_k(x_i) = \sum_j \mu_{ij}\varepsilon_k(x_j) = \sum_j \mu_{ij}\zeta(x_j, x_k)$. Abbiamo quindi l'equazione matriciale $MZ = 1$, con $M = (\mu_{ij})$, $Z = (\zeta(x_i, x_j))$. Ne segue che $M = Z^{-1}$ e quindi μ è univocamente determinata. □

In maniera analoga si trattano sistemi della forma

$$g(y) = \sum_{\substack{x \in S \\ x \leq y}} f(x), \qquad y \in S. \tag{23}$$

Definendo Z come sopra, la (23) equivale all'equazione fra vettori riga

$$(g(x_1), g(x_2), \ldots, g(x_n)) = (f(x_1), f(x_2), \ldots, f(x_n))Z.$$

Pertanto $(f(x_1), \ldots, f(x_n)) = (g(x_1), \ldots, g(x_n))M$, dove $M = Z^{-1}$. Ne ricaviamo il seguente risultato.

Corollario 1. *Se $f, g\colon S \to A$, con A abeliano, soddisfano $g(y) = \sum_{x \leq y} f(x)$, allora*

$$f(y) = \sum_{x \in S} \mu(x, y)g(x). \tag{24}$$

Si può determinare la funzione di Möbius per ricorrenza.

Corollario 2. *La funzione di Möbius è l'unica funzione di $S \times S$ in $\mathbb{Z}$ che soddisfa $\mu(x, y) = 0$ se $x \not\leq y$, e la relazione di ricorrenza*

$$\sum_{x \leq y \leq z} \mu(x, y) = \delta(x, z), \tag{25}$$

dove $\delta(x, z)$ vale 1 se $x = z$ e 0 se $x \neq z$. In alternativa, alla (25) si può sostituire

$$\sum_{x \leq y \leq z} \mu(y, z) = \delta(x, z). \tag{26}$$

Dimostrazione. Le due condizioni corrispondono rispettivamente alle equazioni matriciali $MZ = 1$ e $ZM = 1$, dove 1 è la matrice identità $n \times n$. □

Corollario 3. *Se in S gli intervalli $I[x, z]$ e $I[w, t]$ sono isomorfi, allora $\mu(x, z) = \mu(w, t)$.*

Dimostrazione. Dalla (25) segue $\mu(x, z) = \delta(x, z) - \sum_{x \leq y < z} \mu(x, y)$. La tesi si ottiene per induzione sulla lunghezza dell'intervallo. □

Tra poco applicheremo questo metodo per risolvere il problema posto a p. 480. Formuliamo prima un secondo problema dello stesso tipo.

Problema 2. La colorazione delle carte. Una *carta* è un piano diviso in un numero finito di regioni connesse non sovrapposte, dette *paesi*, da un numero finito di archi che si incontrano soltanto agli estremi. Due paesi sono *adiacenti* se hanno in comune un arco di frontiera. Una *colorazione propria* assegna colori ai paesi in modo che paesi adiacenti abbiano colori diversi. Data una carta e dato k, ci si può domandare in quanti modi la si possa colorare propriamente usando k colori. Il celebre problema dei quattro colori, che il testo attribuisce a De Morgan nel 1850 e che fu risolto impiegando circa 1200 ore di calcolo, chiede se ogni carta possa essere colorata con quattro colori. In altri termini, è positivo il numero delle colorazioni proprie di ogni carta per $k = 4$? La risposta è sì. Mostreremo che, per ogni carta data, esiste un polinomio in k a coefficienti interi, detto *polinomio cromatico* della carta, che dà il numero delle sue colorazioni proprie con k colori.

Una *sottocarta* Δ di una carta Γ è una carta ottenuta cancellando alcuni confini. Ordiniamo parzialmente l'insieme S delle sottocarte di Γ ponendo $E \le \Delta$ se E è una sottocarta di Δ. Per $\Delta \in S$ indichiamo con $f(\Delta)$ il numero delle colorazioni proprie di Δ con k colori e con $g(\Delta)$ il numero di tutte le colorazioni con k colori. Se $c(\Delta)$ è il numero di paesi di Δ, allora $g(\Delta) = k^{c(\Delta)}$. Inoltre ogni colorazione di Δ corrisponde a una colorazione propria di una sottocarta E di Δ (si cancellano i confini fra paesi dello stesso colore). Quindi $g(\Delta) = \sum_{E \le \Delta} f(E)$ e per il corollario 1

$$f(\Delta) = \sum_{\substack{E \in S \\ E \le \Delta}} \mu(E, \Delta) k^{c(E)}, \tag{27}$$

dove μ è la funzione di Möbius di S. Questo è il polinomio cromatico di Δ.

Consideriamo ora il calcolo della funzione di Möbius per alcuni insiemi parzialmente ordinati.

Teorema 8.9. *Per la catena $C = \{0, 1, \dots, n\}$ di lunghezza n con l'ordine naturale, la funzione di Möbius è data da*

$$\mu(i, i) = 1, \qquad \mu(i-1, i) = -1, \qquad \mu(j, i) = 0 \quad \text{per } j \ne i, i-1. \tag{28}$$

Dimostrazione. Da $g(i) = \sum_{j \le i} f(j) = \sum_{j=0}^{i} f(j)$ segue $f(i) = g(i) - g(i-1)$ per $i \ge 1$, mentre $f(0) = g(0)$. Per il corollario 1 si ottiene (28). □

Vediamo come ridurre il calcolo della funzione di Möbius del prodotto di due insiemi parzialmente ordinati al calcolo delle funzioni dei due fattori. Ricordiamo (esercizio 7, p. 461) che $S_1 \times S_2$ è ordinato da $(x_1, x_2) \le (y_1, y_2)$ se e solo se $x_1 \le y_1$ e $x_2 \le y_2$.

Teorema 8.10. *Sia $S = S_1 \times S_2$ e siano μ, μ_1, μ_2 le funzioni di Möbius rispettivamente di S, S_1, S_2. Allora*

$$\mu((x_1, x_2), (y_1, y_2)) = \mu_1(x_1, y_1)\mu_2(x_2, y_2) \tag{29}$$

per ogni $x_1, y_1 \in S_1$, $x_2, y_2 \in S_2$.

Dimostrazione. Siano $\delta_1, \delta_2, \delta$ le funzioni delta di S_1, S_2, S. Allora $\delta((x_1, x_2), (y_1, y_2)) = \delta_1(x_1, y_1)\delta_2(x_2, y_2)$. Inoltre

$$\begin{aligned}
&\sum_{(x_1,x_2) \le (y_1,y_2) \le (z_1,z_2)} \mu_1(y_1, z_1)\mu_2(y_2, z_2) \\
&= \sum_{x_1 \le y_1 \le z_1} \sum_{x_2 \le y_2 \le z_2} \mu_1(y_1, z_1)\mu_2(y_2, z_2) \\
&= \left(\sum_{x_1 \le y_1 \le z_1} \mu_1(y_1, z_1) \right) \left(\sum_{x_2 \le y_2 \le z_2} \mu_2(y_2, z_2) \right) \\
&= \delta_1(x_1, z_1)\delta_2(x_2, z_2) = \delta((x_1, x_2), (z_1, z_2)).
\end{aligned}$$

Il prodotto e la funzione di Möbius di S soddisfano così la medesima ricorrenza del corollario 2, e per unicità coincidono. □

Questo risultato e il teorema 8.9 permettono di calcolare la funzione di Möbius dell'algebra di Boole $\mathcal{P}(S)$ dei sottoinsiemi di un insieme finito $S = \{1, 2, \dots, n\}$.

Corollario. *La funzione di Möbius di $\mathcal{P}(S)$ è*

$$\mu(U, V) = \begin{cases} (-1)^{|V-U|} & U \subseteq V, \\ 0 & U \not\subseteq V, \end{cases}$$

dove $V - U = V \cap U'$ è l'insieme degli elementi di V che non appartengono a U.

Dimostrazione. Il reticolo $\mathcal{P}(S)$ è isomorfo al prodotto di n copie della catena $C = \{0, 1\}$. Infatti a ogni sottoinsieme U associamo la funzione caratteristica $\chi_U \colon S \to \{0, 1\}$, uguale a 1 su U e a 0 fuori di U, rappresentata dal vettore $(\chi_U(1), \ldots, \chi_U(n))$. L'applicazione $U \mapsto (\chi_U(1), \ldots, \chi_U(n))$ è un isomorfismo di $\mathcal{P}(S)$ su $C_1 \times \cdots \times C_n$, con $C_i = \{0, 1\}$ nell'ordine naturale. Dal teorema 8.10 applicato più volte e dal teorema 8.9 risulta

$$\mu(U, V) = \prod_{i=1}^{n} \mu_i(\chi_U(i), \chi_V(i)) = \begin{cases} (-1)^{|V-U|} & U \subseteq V, \\ 0 & U \not\subseteq V. \end{cases} \qquad \square$$

L'impiego della funzione di Möbius di $\mathcal{P}(S)$ è spesso detto *metodo di inclusione ed esclusione*. Possiamo ora dare la soluzione del primo problema.

Soluzione del problema 1. Il numero degli scompigli di $S = \{1, \ldots, n\}$ è

$$\begin{aligned} f(\varnothing) &= \sum_{U \in \mathcal{P}(S)} \mu(\varnothing, U) g(U) = \sum_{U} (-1)^{|U|} |U'|! \\ &= \sum_{i=0}^{n} (-1)^i \binom{n}{i} (n-i)! = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!}. \end{aligned}$$

Questa quantità è asintoticamente uguale a $n!/e$. La probabilità che una permutazione scelta a caso sia uno scompiglio è dunque molto vicina a $1/e$, sostanzialmente indipendente da n.

Consideriamo infine l'esempio classico da cui ha avuto origine la teoria.

Problema 3. Sia n un intero positivo e sia D_n il reticolo dei suoi divisori positivi ordinati per divisibilità, con $a \leq b$ se $a \mid b$.[9] Se $n = p_1^{e_1} \cdots p_h^{e_h}$ con p_i primi distinti ed $e_i > 0$, l'applicazione

$$d = p_1^{d_1} \cdots p_h^{d_h} \longmapsto (d_1, \ldots, d_h)$$

è un isomorfismo di D_n su $C_1 \times \cdots \times C_h$, dove C_i è la catena $\{0, 1, \ldots, e_i\}$. Se $c = p_1^{c_1} \cdots p_h^{c_h} \mid d$, per cui $c_i \leq d_i$, allora

$$\mu(c, d) = \prod_{i=1}^{h} \mu(c_i, d_i) = \begin{cases} (-1)^l & \text{se } d/c \text{ è prodotto di } l \text{ primi distinti,} \\ 0 & \text{se } d/c \text{ ha un fattore quadrato maggiore di 1.} \end{cases}$$

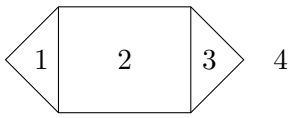
Si osservi che $\mu(c, d) = \mu(1, d/c)$, valore della classica funzione di Möbius della teoria dei numeri, denotata con $\mu(d/c)$. La formula di inversione corrispondente è quella scoperta da Möbius.

Esercizi

1. Sia $\varphi(n)$ la funzione di Eulero, numero degli interi positivi fino a n e primi con n. Usare l'inversione per dedurre
$$\varphi(n) = n \prod_{\substack{p \text{ primo} \\ p \mid n}} \left(1 - \frac{1}{p}\right).$$

2. Determinare l'insieme parzialmente ordinato delle sottocarte della carta Γ rappresentata qui e determinare il suo polinomio cromatico.

[9] Qui le formule dell'originale usano l'ordine usuale dei divisori, come indicano la parametrizzazione per esponenti crescenti e la formula di inversione.



3. Sia $L(V)$ il reticolo dei sottospazi di uno spazio V di dimensione n su un campo di q elementi. Il *coefficiente gaussiano* $\begin{bmatrix} n \\ k \end{bmatrix}_q$ denota il numero dei sottospazi di dimensione k. Ricavare le formule

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q^n-1)(q^n-q)\cdots(q^n-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}, \tag{30}$$

$$\begin{bmatrix} n \\ k \end{bmatrix}_q = q^k \begin{bmatrix} n-1 \\ k \end{bmatrix}_q + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q. \tag{31}$$

4. Per $X, Y \in L(V)$ come nell'esercizio 3 e $X \subseteq Y$, $\mu(X,Y)$ dipende soltanto da $m = \dim Y - \dim X$, per il corollario 3 al teorema 8.8. Scriviamo perciò $\mu(X,Y) = \mu(m)$. Provare che

$$\mu(m) = (-1)^m q^{m(m-1)/2}. \tag{32}$$

5. Sia W un sottospazio di V di dimensione l nelle ipotesi dell'esercizio 3. Mostrare che il numero dei sottospazi U di dimensione k tali che $U \cap W = 0$ è

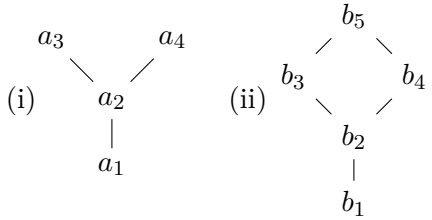
$$\sum_{i=0}^{l} (-1)^i q^{i(i-1)/2} \begin{bmatrix} n-i \\ k-i \end{bmatrix}_q \begin{bmatrix} l \\ i \end{bmatrix}_q. \tag{33}$$

6. Trovare il numero totale degli insiemi di vettori che generano V, nelle ipotesi dell'esercizio 3.

7. Sia $G_{n,p}$ il prodotto diretto di n gruppi ciclici di ordine primo p. Sia H un suo sottogruppo isomorfo a $G_{k,p}$. Trovare il numero degli omomorfismi iniettivi $\eta\colon G_{l,p} \to G_{n,p}$ tali che $\eta(G_{l,p}) \cap H = 0$.

8. Per due partizioni π, ρ di S, si dice che π è un *raffinamento* di ρ se ogni classe di π (si veda p. 11) è contenuta in una classe di ρ. Mostrare che l'insieme P delle partizioni di S, ordinato per raffinamento, è un reticolo e determinarne la funzione μ.

9. Determinare le funzioni di Möbius dei seguenti insiemi parzialmente ordinati.



Appendice

Alcuni argomenti per lo studio individuale

1. Domini euclidei

Riferimenti. (1) G. H. Hardy ed E. M. Wright, *An Introduction to the Theory of Numbers*, 5ª ed., Oxford University Press, New York, 1975, pp. 212–217. (2) H. Chatland, «On the Euclidean algorithm», *Bulletin of the American Mathematical Society*, 55 (1949), pp. 948–953. (3) T. Motzkin, «The Euclidean algorithm», ivi, 55 (1949), pp. 1142–1156. (4) P. Samuel, «About Euclidean rings», *Journal of Algebra*, 19 (1975), pp. 282–301.

I primi tre lavori discutono il problema di determinare quali anelli I degli elementi interi dei campi quadratici $\mathbb{Q}(\sqrt{m})$ siano euclidei (si vedano le pp. 147–151 e gli esercizi 3–5 a p. 287 dell'originale). L'articolo di Samuel sviluppa una teoria generale dei domini euclidei.

2. Domini non commutativi a ideali principali

Riferimenti. (1) N. Jacobson, *Theory of Rings*, American Mathematical Society Surveys, n. 2, Providence, Rhode Island, 1943, cap. 3. (2) P. M. Cohn, *Free Rings and their Relations*, Academic Press, Londra e New York, 1971, cap. 8.

Si tratta di estendere ai domini non commutativi la teoria presentata alle pp. 147–151 dell'originale. Esempi sono l'anello dei polinomi in un'indeterminata a coefficienti in un anello con divisione e l'anello dei polinomi differenziali formali in un'indeterminata.

3. Il teorema dei quattro quadrati e i quaternioni interi

Riferimenti. (1) G. H. Hardy ed E. M. Wright, *An Introduction to the Theory of Numbers*, 5ª ed., Oxford University Press, New York, 1975, pp. 300–310. (2) I. N. Herstein, *Topics in Algebra*, 2ª ed., John Wiley and Sons, New York, 1975, pp. 371–377.

Il teorema, dovuto a Lagrange, afferma che ogni intero positivo è somma di quattro quadrati. L'anello dei quaternioni interi è l'anello I definito nell'esercizio 5 a p. 100 dell'originale. Esso è un dominio euclideo non commutativo, quindi a ideali principali, la cui aritmetica si può usare per dimostrare il teorema dei quattro quadrati. Un passo importante della dimostrazione consiste nel verificare che, se $N(x)$ è la forma norma di un'algebra di quaternioni su un campo di caratteristica $p \neq 2$, nel caso finito in esame esiste $x \neq 0$ con $N(x) = 0$.[10] Herstein lo dimostra invocando il teorema di Wedderburn sulla commutatività degli anelli finiti con divisione. Una dimostrazione più diretta segue dall'esercizio 6 a p. 361 dell'originale.

4. Gruppi cristallografici del piano

Riferimenti. (1) H. Weyl, *Symmetry*, Princeton University Press, Princeton, New Jersey, 1952, pp. 83–115. (2) H. S. M. Coxeter, *Introduction to Geometry*, John Wiley and Sons, New York, Londra e Sydney, 1961, cap. 4. (3) R. Schwarzenberger, *N-dimensional Crystallography*, Pitman Publishing Program, San Francisco, Londra e Melbourne, 1980, pp. 1–10.

Sono i gruppi discontinui (discreti) di movimenti euclidei del piano. Analiticamente un tale movimento è un'applicazione $(x, y) \mapsto (x', y')$ con $x' = ax + by + h$, $y' = cx + dy + k$, dove

[10] *Nota alla traduzione.* L'enunciato stampato omette l'ipotesi di finitezza del campo, necessaria in questo passaggio: la norma di un'algebra di quaternioni con divisione è anisotropa.

$a, b, c, d, h, k \in \mathbb{R}$ e la matrice $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ è ortogonale. Discontinuità significa che esiste un intorno dell'applicazione identica che non contiene alcun elemento del gruppo diverso dall'identità. Con una naturale nozione di equivalenza (si veda il libro di Weyl) il problema si riconduce a classificare i gruppi in cui i coefficienti $a, b, \ldots$ sono interi e il sottogruppo con $h = k = 0$ è finito. Il caso interessante è quello in cui h e k assumono tutti i valori interi. A meno di equivalenza unimodulare vi sono 17 gruppi di questo tipo. Sono i possibili gruppi di simmetria degli ornamenti piani, come i motivi delle carte da parati. Ne illumina il significato storico questo passo di Weyl (p. 103):

> Esempi di tutti e diciassette i gruppi di simmetria si trovano nei motivi decorativi dell'antichità, in particolare negli ornamenti egizi. È difficile sopravvalutare la profondità dell'immaginazione geometrica e dell'inventiva che vi si riflette. La loro costruzione è tutt'altro che banale dal punto di vista matematico. L'arte dell'ornamento contiene in forma implicita il più antico esempio di matematica superiore a noi noto.

5. Gruppi finiti di riflessioni

Riferimenti. (1) R. Steinberg, *Lectures on Chevalley Groups*, Yale University Lecture Notes, Department of Mathematics, Yale University, New Haven, Connecticut, 1967, appendice. (2) C. T. Benson e L. C. Grove, *Finite Reflection Groups*, Bogden and Quigley, Tarrytown-on-Hudson, New York, e Belmont, California, 1971. (3) R. Carter, *Simple Groups of Lie Type*, Wiley-Interscience, New York, 1972, cap. 2.

Sono i gruppi finiti generati da riflessioni (ossia da simmetrie nel senso definito a p. 363 dell'originale) in uno spazio euclideo. Sotto la veste di «gruppi di Weyl» hanno un ruolo fondamentale nella teoria delle algebre di Lie semplici e dei gruppi di Lie semplici.

6. Gruppi di Mathieu

Riferimenti. (1) E. Witt, «Die 5-fach transitiven Gruppen von Mathieu» e «Über Steinersche Systeme», *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, 12 (1938), pp. 256–264 e 265–275. (2) J. A. Todd, «On representation of the Mathieu group M_{24} as a collineation group», *Annali di Matematica Pura ed Applicata* (IV), 71 (1966), pp. 199–238. (3) N. L. Biggs e A. T. White, *Permutation Groups and Combinatorial Structures*, Cambridge University Press, Londra, 1979, p. 57 e pp. 70–74.

I cinque gruppi $M_{11}, M_{12}, M_{22}, M_{23}, M_{24}$ sono gruppi semplici molteplicemente transitivi scoperti da E. L. Mathieu nel 1861 e nel 1873. Il pedice n indica il grado del gruppo di permutazioni (M_n è un sottogruppo di S_n). Burnside li chiamò gruppi semplici «sporadici», perché non appartengono ad alcuna delle classi infinite di gruppi semplici (come i gruppi alterni o i gruppi semplici definiti a partire da $GL_n(F)$ per F finito). Fra il 1966 e il 1981 sono stati trovati altri 21 gruppi semplici sporadici, e grazie al lavoro di molti matematici si è giunti alla classificazione completa dei gruppi semplici finiti. L'elenco comprende (1) i gruppi ciclici di ordine primo, (2) i gruppi alterni A_n per $n \geq 5$, (3) i gruppi di tipo Lie definiti nel terzo riferimento del punto 5 e (4) i 26 gruppi semplici sporadici. Un particolare reticolo, il «reticolo di Leech», ha un ruolo importante nella definizione di gran parte dei nuovi gruppi sporadici. È collegato ai sistemi di Steiner che intervengono nella definizione dei gruppi di Mathieu. Per la definizione del reticolo di Leech, i suoi rapporti con i sistemi di Steiner e la definizione del gruppo sporadico di Conway si può consultare un articolo di J. Conway nel *Bulletin of the London Mathematical Society*, 1 (1969).

7. Campi finiti

Riferimenti. (1) L. E. Dickson, *Linear Groups with an Exposition of Galois Field Theory*, 1900; ristampa Dover Publications, 1958, pp. 1–54. (2) A. A. Albert, *Fundamental Concepts of Higher Algebra*, University of Chicago Press, Chicago, 1956, cap. 5.

Questi libri espongono molte proprietà particolari dei campi finiti che non figurano nei trattati generali di algebra. I campi finiti hanno importanti applicazioni, per esempio all'informatica e alla crittografia.

8. Il teorema di irriducibilità di Hilbert

Riferimenti. (1) D. Hilbert, «Über die Irreduzibilität ganzer rationaler Funktionen mit ganzzahligen Koeffizienten», *Journal für die reine und angewandte Mathematik*, 110 (1892), pp. 104–129; oppure *Gesammelte Abhandlungen*, vol. 2, Springer-Verlag, Berlino, 1933. (2) S. Lang, *Diophantine Geometry*, Wiley-Interscience, New York, 1962, cap. 8. (3) C. R. Hadlock, *Field Theory and its Classical Problems*, Carus Mathematical Monographs, Mathematical Association of America, 1978, cap. 4.

Il teorema di Hilbert afferma che, se $f(t_1, \ldots, t_n, x) \in D = \mathbb{Q}[t_1, \ldots, t_n, x]$ è irriducibile in D, esistono infinite scelte dei $t_i = a_i \in \mathbb{Q}$ per le quali $f(a_1, \ldots, a_n, x)$ è irriducibile in $\mathbb{Q}[x]$. Il terzo riferimento dà una dimostrazione relativamente semplice. Nel secondo il risultato è dimostrato per polinomi a coefficienti in un campo di numeri algebrici su $\mathbb{Q}$. Hilbert si servì del teorema per provare l'esistenza di infiniti polinomi a coefficienti razionali con gruppo di Galois S_n o A_n.

9. Gruppi di Galois di alcuni polinomi classici

Riferimenti. (1) I. Schur, «Gleichungen ohne Affekt», *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse*, 1930, pp. 443–449; oppure *Gesammelte Abhandlungen*, vol. 3, pp. 191–197. (2) I. Schur, «Affektlose Gleichungen in der Theorie der Laguerreschen und Hermiteschen Polynome», *Journal für die reine und angewandte Mathematik*, 165 (1931), pp. 52–58; oppure *Gesammelte Abhandlungen*, vol. 3, pp. 227–233.

Questi lavori determinano i gruppi di Galois su $\mathbb{Q}$ dei polinomi di Laguerre e di Hermite, dei polinomi

$$E_n(x) = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}$$

e di polinomi affini. In tutti i casi il gruppo di Galois è S_n oppure A_n.

10. Equazioni di Plücker

Riferimenti. (1) W. V. D. Hodge e D. Pedoe, *Methods of Algebraic Geometry*, vol. 1, Cambridge University Press, Cambridge, 1947, pp. 286–315. (2) N. Jacobson e D. Saltman, *Finite Dimensional Division Algebras*, volume annunciato, cap. 3.

Le equazioni di Plücker sono equazioni algebriche nelle coordinate di Plücker di un elemento ω della componente omogenea V^r dell'algebra esterna $E(V)$; esprimono condizioni necessarie e sufficienti affinché ω sia decomponibile. Esse conferiscono all'insieme dei vettori decomponibili una struttura algebrico-geometrica: definiscono una varietà grassmanniana, corrispondente all'insieme dei sottospazi di dimensione r dello spazio vettoriale V.

Indice analitico

I numeri si riferiscono alle pagine della seconda edizione inglese. Nel PDF, ogni numero apre il paragrafo italiano corrispondente all'inizio della pagina o dell'intervallo citato.

A

Algebra di gruppo, 127, 408.

Algebre, alternative, 443; associative, 407; di Boole, 474; di Cayley–Graves, 449; di composizione, 440; delle derivazioni, 435; esterne, 411–422; libere, 127; di gruppo, 127, 408; omomorfismi, 409; ideali, 409; di Jordan, 435; di Lie, 434; di trasformazioni lineari, 422; rappresentazioni matriciali, 424; non associative, 409; quoziente, 424; rappresentazione, 403.

Algoritmo della divisione, 23, 128–129, 317.

Amitsur–Levitzki, teorema di, 422.

Anello, 86; gruppo additivo, 87; antiomomorfismo, 114; anti-isomorfismo, 112; booleano, 476; caratteristica, 109; commutativo, 90; somma diretta, 110; con divisione, 91; degli endomorfismi di un gruppo abeliano, 158–162; degli endomorfismi di un modulo, 169, 204–208; delle serie formali di potenze, 127; degli interi di Gauss, 87; gruppo delle unità, 91; omomorfismo, 106; automorfismo interno, 112; teoremi di isomorfismo, 108; omomorfismo di Jordan, 114; di matrici, 92–96; moltiplicazioni, 162; monoide moltiplicativo, 87; opposto, 113; di polinomi, 119–126; quoziente, 101–102; delle classi di resti modulo un intero, 103–105; semplice, 110; senza unità, 155–156.

Anello booleano, 476.

Anello con divisione, 91.

Anello quoziente, 101.

Anello semplice, 110.

Anisotropo, 359.

Annullatore, 168.

Anti-isomorfismo, 109.

Antiomomorfismo, 114.

Applicazione, 5; antilineare, 402; biiettiva, 7; codominio, 5; composizione, 7; dominio, 5; grafico, 5; identica, 8; immagine, 6; iniettiva, 7; immagine inversa, 12; inversa, 9; naturale, 12; che conserva l'ordine, 460; semilineare, 470; suriettiva, 7.

Artin–Chevalley, teorema di, 137.

Associati, 140.

Associatività, 8, 17; generalizzata, 39–40; delle potenze, 433.

Associatore, 435.

Automorfismo, 60; di un'estensione di campi, 234; gruppo degli automorfismi, 60; gruppo degli automorfismi esterni, 63; interno, 63, 112.

Azione di un gruppo su un insieme, 71–79; fedele, 72; azioni equivalenti, 73; nucleo, 72; primitiva, 77; transitiva, 75.

B

Base, 170; cartesiana, 361; duale (o complementare), 344; normale, 291; ortogonale, 356; simplettica, 391.

Base normale, teorema della, 294.

Base simplettica, 391.

Binomio, teorema del, 89.

Blocco, 11.

Boole, algebra di, 474.

Budan, teorema di, 316.

Buon ordinamento dei numeri naturali, 18.

C

Campo, 91; algebricamente chiuso, 224; ciclotomico, 252, 271–276; finito, 277–278, 287–290, 491; delle frazioni, 115–118; ordinabile, 307; ordinato, 307; perfetto, 233; primo, 213; reale chiuso, 308; di spezzamento, 224.

Campo ordinato, 307.

Campo reale chiuso, 308.

Cancellazione, legge di, 17, 36; ristretta, 90–91.

Caratteri, 291.

Caratterizzazione di S_5, 82–83.

Cardano, formule di, 266.

Cartan–Brauer–Hua, teorema di, 100.

Cartan–Dieudonné, teorema di, 372.

Cartesio, regola dei segni di, 316.

Casus irreducibilis, 267.

Catena ascendente di ideali, 102, 147.

Cayley, teorema di, 38; analogo per le algebre, 423; analogo per gli anelli, 161.

Centralizzante, 41.

Centro, 41, 97.

Cicli, 48; decomposizione in cicli, 48; disgiunti, 48.

Classi di coniugio, 74; nel gruppo simmetrico, 74.

Classi laterali, 52; spazio delle classi laterali, 73.

Cogredienti (congruenti), 349.

Colorazione delle carte geografiche, 484.

Commutatività, 17, 40–41; dei diagrammi, 8.

Commutatore, 245; sottogruppo dei commutatori, 238–245.

Complemento, 10, 469.

Composizione r-aria, 10.

Congruenze, 54–57; intersezione, 58.

Coniugati, 255, 291.

Contenuto di un polinomio, 151.

Copertura, 318, 457.

Corrispondenza, 15; biunivoca, 7.

Costruibilità euclidea, 216–224, 271–274; poligoni regolari costruibili, 272–273.

Costruzioni euclidee, 216–224, 271–274.

D

Decomposizione in fratti semplici, 150.

Dedekind, teorema di indipendenza di, 291.

Derivazione, 434; ordinaria, 230.

Determinanti, 95, 396–400.

Dickson–Dieudonné, teorema di, 389.

Differenza simmetrica, 475.

Discriminante, 258, 345, 349.

Distributività, leggi di, 4.

Divisore, 22; massimo comune divisore, 23, 144; dello zero, 90.

Divisore dello zero, 90.

Divisore elementare, 193.

Dominio, 90; euclideo, 148, 489; fattoriale, 141; a ideali principali, 130, 147–149.

Dualità, principio di, 459.

E

Eisenstein, criterio di, 154.

Elementi algebricamente indipendenti, 126.

Elemento, 3; algebrico, 124; elementi algebricamente indipendenti, 126; primo, 22, 143; quasi-invertibile, 156; trascendente, 124.

Elemento algebrico, 124.

Elemento antisimmetrico, 113.

Elemento idempotente, 91.

Elemento irriducibile, 141.

Elemento nilpotente, 91.

Elemento primitivo di un'estensione, 214, 290.

Elemento quasi-invertibile, 156.

Elemento simmetrico, 113.

Elemento trascendente, 124.

Endomorfismi, 60; anello degli endomorfismi, 158–162, 169, 204–208.

Engel, proprietà di, 251.

Epimorfismo, 59.

Equazione delle classi, 76.

Equazione generale, 262–266.

Equivalenza, classe, 11; di matrici, 181; relazione, 11.

Esponente di un gruppo finito, 46.

Espressione razionale simmetrica, 242.

Estensioni di campi, abeliane, 253; algebriche, 216; cicliche, 253; di Galois, 253; normali, 238; separabili, 238; semplici, 214.

Euclide, algoritmo di, 150, 316–320.

Eulero, funzione φ di, 47, 105, 110.

Eulero, teorema di, 102–105.

F

Fattore, 22, 140.

Fattore invariante, 193.

Fermat, teorema di, 105.

Fibra, 12.

Filtro, 479.

Forma, 354; bilineare, 344; hermitiana, 401–403; quadratica, 354–361. Vedi anche Forma bilineare.

Forma (o matrice) canonica razionale, 198–200.

Forma bilineare, 344; alternante, 348, 349–353; anisotropa, 359; discriminante, 345, 349–350; equivalente, 348; isotropa, 359; matrice, 345; non degenere, 346; nulla, 359; radicali, 346; simmetrica, 347–348, 354–361; universale, 359.

Forme quadratiche, 355; radicale bilineare, 355; che ammettono composizione, 440; definite positive, 361; radicale, 356.

Frazioni, 116.

Frobenius, p-automorfismo di, 304.

Frobenius, teorema di, sulle matrici che commutano, 207; sulle algebre reali con divisione, 452.

Funzione, lineare, 343; polinomiale, 134–137, 354; trilineare, 435.

G

Galois, gruppo di, 234–242, 252, 256–260; dei campi ciclotomici, 276.

Gauss, lemma di, 152.

Geometria, ortogonale, 361–366; simplettica,

391–396; unitaria, 401–403.

Geometria simplettica, 391–396.

Geometria unitaria, 401–403.

Grado, di un polinomio, 128; di un'espressione razionale, 243; totale, 137.

Gruppi cristallografici, 490.

Gruppi di riflessioni, 491.

Gruppo, 31; abeliano, 41, 46–47, 194; operante su un insieme, 71–78; alterno, 51, 71, 247; di automorfismi, 60; dei commutatori (derivato), 245; ciclico, 42–47; diedrale, 34; prodotto diretto, 35; quoziente, 56; finitamente generato, 69; finitamente presentato, 70; libero, 69; abeliano libero, 67; lineare generale, 95, 375–382; generatori, 42; omomorfismo, 58; isomorfismo, 36–37; nilpotente, 261; ortogonale, 362; p-gruppo, 245; periodico, 47; di permutazioni, 32; proiettivo unimodulare, 377; prodotto semidiretto, 79; semplice, 247; risolubile, 244–245; simmetrico, 31, 70; simplettico, 391; di trasformazioni, 32; delle unità, 31, 91; prodotto intrecciato, 79.

Gruppo di trasformazioni, 32; relazione di equivalenza associata, 51; transitivo, 51–52.

Gruppo lineare, 95.

Gruppo quoziente, 56.

Gruppo semplice, 57.

H

Hamilton–Cayley–Frobenius, teorema di, 201.

Hilbert, teorema di irriducibilità di, 492.

Hua, identità di, 92.

Hua, teorema di, 114.

Hurwitz, problema di, 438–449; teorema generalizzato, 447.

I

Ideale, 101; sinistro e destro, 103; massimale, 111; dell'ordine (annullatore), 169; primo, 111; principale, 102, 478.

Inclusione-esclusione, metodo di, 486.

Indeterminate, 122, 125.

Induzione matematica, 16, 18.

Insieme parzialmente ordinato, 456; funzione di Möbius, 480–487.

Insieme totalmente ordinato (catena), 456.

Insiemi, 3; prodotto cartesiano, 4; funzioni caratteristiche, 5, 386; intersezione, 3; insieme delle parti, 3; quoziente, 12; totalmente ordinati, 456; unione, 4.

Interi algebrici, 279.

Intervalli in un reticolo, 460.

Inverso, 9, 31; elemento invertibile (o unità), 31.

Involuzione, 82, 108, 112.

Isometria, 362.

Isomorfismo, indipendenza algebrica di isomorfismi, 294; primo e secondo teorema di isomorfismo per i gruppi, 65; per gli anelli, 108; di gruppi, 36–37; di reticoli, 460; di monoidi, 37; di anelli, 94–95.

Isomorfismo d'ordine, 308.

Isotropo, 359.

J

Jacobi, identità di, 432.

Jacobson–Rickart, teorema di, 114.

Jordan, forma (o matrice) canonica di, 200.

Jordan, identità di, 433.

Jordan, omomorfismo di, 114.

Jordan, prodotto di, 432.

Jordan–Hölder, teorema di, 249.

Jordan–Hölder–Dedekind, teorema di, 466.

L

Lagrange, risolvente di, 253.

Lagrange, teorema di, 52.

Laplace, sviluppo di un determinante secondo, 417.

Legendre, polinomio di, 316.

Legendre, simbolo di, 133–134.

Lie, prodotto di (commutatore additivo), 431.

Lindemann–Weierstrass, teorema di, 277, 278.

Lunghezza, di un intervallo in un reticolo modulare, 467; in un dominio a ideali principali, 183–184.

M

Malcev, esempi di, 119.

Massimo comune divisore, 23, 144.

Mathieu, gruppi di, 491.

Matrici, 90, 173; somma, 90; aggiunta classica, 96; alternanti, 350; polinomio caratteristico, 196; cogredienti, 349; compagna, 197; composte, 417; diagonali, 94; elementari, 181–182; equivalenti, 181; hermitiane, 401; inversa, 96; forma canonica di Jordan, 200; anello di matrici, 92–96; unità matriciali, 94; moltiplicazione, 93; forma normale, 184; forma canonica razionale, 198, 200; simili, 195; traccia, 196, 423; trasposta, 111.

Matrici canoniche, di Jordan, 200; razionali, 198–200.

Minimo comune multiplo, 23, 144.

Möbius, formula di inversione di, 151, 487.

Möbius, funzione di, di un insieme parzialmente ordinato, 480–487; degli interi positivi, 151.

Modulo, 163, 164; ciclico, 168; definito mediante una trasformazione lineare, 165; somma diretta, 175–178; libero, 170; omomorfismo, 168; indecomponibile, 193; irriducibile, 169; primario, 191; quoziente, 167; un anello considerato come modulo, 165; di torsione, 189, 191.

Modulo quoziente, 167.

Moltiplicazione sinistra, 162.

Monade, 28.

Monoide, 28; prodotto diretto, 35; fattoriale, 140–146; libero, 68; generatori, 42; omomorfismo, 58–59; isomorfismo, 37; tavola di moltiplicazione, 30; ordine, 29; quoziente, 55; di trasformazioni, 29.

Monomio, 125.

Monomorfismo, 59.

Moufang, identità di, 443, 450.

Multiplo, 42; minimo comune multiplo, 23, 144.

N

Newton, identità di, 140.

Norme, 296, 424; transitività, 426.

Nucleo anisotropo, 370.

Nucleo di un omomorfismo, 61, 72, 106, 168.

Numeri algebrici, 279.

Numeri cardinali, 24–25.

Numeri interi (il sistema $\mathbb{Z}$), 19.

Numeri naturali, 15.

O

Oggetti liberi, 67–70, 127, 171.

Olomorfo, 63.

Omomorfismo, teorema fondamentale per i moduli, 168; per monoidi e gruppi, 61; per gli anelli, 107; di gruppi, 58; indotto, 61, 106; nucleo, 61, 106, 168; di moduli, 168; di monoidi, 58–59; di restrizione, 59; di anelli, 106.

Orbite, 51, 74.

Ordine, 18; di un elemento di un gruppo, 44; di un monoide, 30; dei gruppi ortogonali e simplettici, 399, 400; di $\mathrm{PSL}_n(F)$ e di $\mathrm{SL}_n(F)$, 381.

Ordine lessicografico, 138, 282.

Ore, teorema di immersione di, 119.

Ortogonale, 347; base, 356; complemento, 349; geometria, 361–366; gruppo, 362; trasformazione, 362.

Ottonioni, 446.

P

Partizione, 11.

Peano, assiomi di, 16.

Permutazione, 31; pari e dispari, 51.

Permutazioni senza punti fissi, 480.

Pfaffiano, 352.

Piano iperbolico, 365.

Plücker, coordinate di, 420.

Plücker, equazioni di, 492.

Polinomio, 119, 122, 125; cromatico, 484; contenuto, 152; ciclotomico, 252, 271–276; grado, 128; funzioni polinomiali, 134–137; gruppo di Galois, 251; omogeneo, 138; di Legendre, 316; p-polinomio, 234; primitivo, 152; radice, 131; separabile, 233; simmetrico, 138–139; grado totale, 137.

Polinomio caratteristico, 196.

Polinomio minimo, 131.

Polinomio monico, 131.

Polinomio simmetrico, 138–139.

Potenza, 40; associatività delle potenze, 433.

Primi fra loro, 145.

Primitività di un'azione di gruppo, 77.

Primo, 22, 142; campo primo di un campo, 213; anello primo di un anello, 108.

Principio dei cassetti, 25.

Prodotto diretto, di gruppi, 35; di monoidi, 35.

Prodotto intrecciato, 79.

Q

Quaternioni, 98–100, 446, 450.

Quattro colori, problema dei, 484.

Quattro quadrati, teorema dei, 490.

R

Radici multiple, 229.

Rappresentazioni, 424; matriciali, 424; regolare, 424.

Relazione binaria, 10; di equivalenza, 11.

Relazione fra le dimensioni, 346, 359, 365; per le estensioni di campi, 214.

Reticolo, 457; anti-isomorfismo, 473; complementato, 469; completo, 458; distributivo, 462; isomorfismo, 460; modulare,

463; semimodulare, 466; dei sottospazi di uno spazio vettoriale, 468–473.

Ricorsione, teorema di, 16.

Riduzione modulo p, 301.

Risolubilità per radicali di un'equazione, 261.

Risultante, 326.

Rotazione, 363.

Ruffini–Abel, teorema di, 264.

S

Schröder–Bernstein, teorema di, 25.

Schur, lemma di, 170.

Segnatura, 359, 403.

Seidenberg, metodo di decisione di, 327–385.

Semigruppo, 28.

Semplicità, dei gruppi alterni, 247; criterio, 379; dei gruppi ortogonali proiettivi, 389; dei gruppi simplettici proiettivi, 397; di $\mathrm{PSL}_n(F)$, 380.

Serie, centrale, 251; di composizione, 248; derivata, 245; normale, 244.

Serie di composizione, 248; fattori di composizione, 248.

Serie formali di potenze, 127.

Simmetria, 363.

Somma diretta, di moduli, 175–178; di anelli, 110.

Sottoanello, 87; generato da un sottoinsieme, 85, 110; primo, 108.

Sottocampo generato da un sottoinsieme, 110.

Sottogruppo, 31; generato da un sottoinsieme, 43; indice, 53; normale, 55; di Sylow, 80–81.

Sottoinsieme, 3.

Sottomodulo, 166.

Sottomonoide, 29; generato da un sottoinsieme, 43.

Sottospazio, isotropo, 365; non degenere, 349; totalmente isotropo, 365.

Spazio proiettivo, 377.

Spazio vettoriale, 165; coniugato, 344.

Stabilizzatore, 78.

Sturm, teorema di, 311; versione parametrica, 320.

Successione di Sturm, 312; ordinaria, 313.

Successione euclidea, 319.

Sylow, teoremi di, 79–82; generalizzazione di Frobenius, 83–84.

Sylvester, teorema di, 359.

Szpilrajn–Marczewski, lemma di, 480.

T

Tarski, teorema di, 335.

Tavola di moltiplicazione, 30.

Teorema 90 di Hilbert, 297–298.

Teorema cinese del resto, 110.

Teorema del fattore, 130.

Teorema del resto, 130.

Teorema fondamentale, dell'aritmetica, 22; della teoria di Galois, 239; degli omomorfismi, 61, 107, 168; della geometria proiettiva, 468.

Teorema fondamentale dell'algebra, 224, 309.

Torre di radicali, 251.

Torre di radici quadrate, 219.

Traccia, funzione traccia, 396, 442; di una matrice, 196, 423; transitività, 426.

Transitività, dei determinanti, 427; delle norme, 426; delle tracce, 426.

Transitività multipla (di grado k**)**, 79.

Trasformazione, 7; gruppo, 32; lineare, 165; monoide, 28; semilineare, 472. Vedi anche Trasformazioni lineari.

Trasformazioni lineari, 165, 194–201; aggiunte, 349, 354; ortogonali, 362, 363; autoaggiunte, 366; antiaggiunte, 371; unipotenti, 366.

Traslazione sinistra, 38.

Trasposizione, 49.

Trasvezione, 381; simplettica, 392.

Trasvezione simplettica, 392.

U

Unità, 28.

W

Wedderburn, teorema di (sugli anelli finiti con divisione), 453.

Wilson, teorema di, 133.

Witt, indice di, 369.

Witt, teorema di cancellazione di, 367.

Witt, teorema di estensione di, 369.