ALGEBRA LINEARE

ISTRUZIONI PER L'USO.

(III) vers.3

Placido Longo

Indice

Capitolo 1

Un problema insignificante

Il problema da cui prendere le mosse non è di quelli che sembrano poter cambiare la vita all'ingegnere, e neppure al matematico. Vedremo che non è esattamente così.

Consideriamo la funzione su $\mathbb{R}^n$ definita ponendo

$$H(x) \;=\; \sum_{i=1}^{n} \alpha_{ij} x_i x_j$$

ove $x_1, x_2, ..., x_n$ sono le componenti di x rispetto alla base canonica $e_1, e_2, ..., e_n$, e $(\alpha_{ij}) \in \mathbb{R}^{n\times n}$. Semplici esempi potrebbero essere $x^2 + xy$ su $\mathbb{R}^2$ oppure $x_7^2 - 4x_3x_5 + 2x_1x_4$ su $\mathbb{R}^{12}$. Il problema da cui partiremo sarà quello dello studio del segno di tali espressioni, che ricorrono spesso e che hanno meritato un nome *ad hoc*, le *forme quadratiche*, anche se sono dei semplici polinomi omogenei di secondo grado. È incredibile il numero di questioni fondamentali, teoriche e pratiche, nelle quali sono coinvolte direttamente o indirettamente, così come verranno qui trattate o in modo generalizzato. Un piccolo saggio di tali applicazioni verrà fornito in coda a questo scritto.
Ancora più miserevole è la scintilla che accende il fuoco: lo studio del segno al quale fingeremo di essere interessatissimi è banalmente semplice se la forma H ha una particolare struttura "diagonale". Se $\alpha_{ij} = 0 \quad$ per $i \neq j$, infatti, risulta

$$H(x) \;=\; \sum_{i=1}^{n} \alpha_{ii} x_i^2$$

Poiché $H(e_i) = \alpha_{ii}$, ne segue subito che se esistono coefficienti α_{ii} di segno discorde, allora H cambia segno, mentre se quelli non nulli sono tutti concordi, essendo i quadrati tutti non negativi, anche H sarà nulla o con essi concorde. Osserviamo di passaggio che, se tutti i coefficienti sono non nulli e concordi, anche H è ovunque non nulla, con l'unica eccezione dell'origine.
Vale la pena di riassumere queste semplici osservazioni in un teorema, perché rappresentano, in certo senso che preciseremo, il caso generale.

Teorema 1 *Sia H una forma quadratica diagonale su $\mathbb{R}^n$.*

- *Se due coefficienti sono discordi (non nulli), allora H cambia segno.*
- *Se i coefficienti sono tutti non nulli e fra loro concordi, allora H è non nulla fuori dall'origine e concorde ai coefficienti.*
- *Se qualcuno dei coefficienti è nullo e gli altri sono tutti concordi, H o è nulla o è concorde ad essi.*

Osserviamo che, salvo che nel secondo caso, esistono punti di $\mathbb{R}^n$ non nulli sui quali H si annulla: se un coefficiente è nullo basta scegliere il vettore della base canonica corrispondente; se no, supposto $\alpha_{ii} > 0$ e $\alpha_{jj} < 0$, basta scegliere $\overline{x} = \sqrt{|\alpha_{jj}|}e_i + \sqrt{\alpha_{ii}}e_j$ per ottenere subito $H(\overline{x}) = 0$.

Purtroppo tanta semplicità si perde del tutto nel caso generale: già negli esempi presentati più su non si capisce gran che su come poter fare, visto che il segno dei prodotti $x_i x_j$ dipende dal segno delle componenti coinvolte, e in ogni caso il segno della somma non è immediatamente ricavabile da quello degli addendi (ad esempio, $x^2 - 2xy + y^2 \geq 0 \quad \forall (x, y) \in \mathbb{R}^2$, indipendentemente da segni negativi di coefficienti e componenti). A dire il vero, si sa perfettamente cosa fare nel primo esempio: raccogliere la x! Il metodo non funziona, però, in tutti i casi nei quali non esistano fattori comuni. Se la forma quadratica dipende solo da due variabili x e y, si può pensare di dividere, con tutte le precauzioni del caso, per x^2 e/o per y^2 e ridurre lo studio a quello del segno di un polinomio di secondo grado nella (sola) variabile x/y (e/o y/x), ma anche questa tecnica non funziona, in generale, se le variabili sono più di due.
L'idea centrale che seguiremo sarà quella di cambiare base: il fatto di essere o meno "diagonale", infatti, non è una proprietà intrinseca della forma quadratica in istudio, ma dipende dalla base prescelta (in questo caso la base canonica). Con una scelta opportuna della base si può, per una vasta classe di forme quadratiche, trasformarne l'espressione in una di tipo "diagonale" e calcolarne i coefficienti o quanto meno, nei casi più sfortunati, stabilire se i loro segni siano concordi o no, e se qualcuno di essi sia nullo: tutto il necessario per completare lo studio!
Il cammino sarà piuttosto lungo e occorrerà, come Newton, "appoggiarsi sulle spalle dei giganti" per ottenere dei resultati utili. In compenso si potranno studiare i massimi e minimi per le funzioni di più variabili, si potranno classificare le quadriche, che generalizzano allo spazio le coniche di Archimede e Apollonio, si capirà meglio la struttura degli sforzi all'interno di un materiale elastico o dell'inerzia di un corpo che ruota, e molte, molte altre cose! Nessuna di esse verrà trattata in dettaglio qui, ma è bene sapere che proprio tali questioni hanno contribuito a motivare le riflessioni che seguono.

Capitolo 2

La diagonalizzazione e lo spettro

2.1 Forme bilineari e forme quadratiche

Per introdurre in modo conveniente il problema iniziale, e visto che abbiamo in mente di cambiare base, diamo una definizione intrinseca di forma quadratica, indipendente dalla base scelta.

Definizione 1 *Dato uno spazio vettoriale reale X, chiameremo* forma bilineare *su X una qualunque funzione $\alpha : X \times X \to \mathbb{R}$, tale che la funzione $u \to \alpha(u, v)$ risulti lineare per ogni fissato $v \in X$, e l'altra $v \to \alpha(u, v)$ risulti lineare per ogni fissato $u \in X$.*
Verrà detta simmetrica *se e solo se $\alpha(u, v) = \alpha(v, u) \quad \forall u, v \in X$.*

Definizione 2 *Dato uno spazio vettoriale reale X, una funzione $H : X \to \mathbb{R}$ verrà detta forma quadratica su X se esiste una forma bilineare α su X tale che $H(u) = \alpha(u, u) \quad \forall u \in X$*

L'esempio più importante di forme bilineari (simmetriche) è costituito dai prodotti scalari, e quello di forme quadratiche dai quadrati delle norme, negli spazi euclidei reali.
Non è immediato riconoscere che le forme quadratiche così introdotte abbiano qualche cosa a che vedere con quelle viste nella sezione precedente. Ciò è conseguenza del seguente

Teorema 2 *Sia X uno spazio reale di dimensione n, e sia α una forma bilineare su X. Sia $e_1, e_2, ..., e_n$ una base e denotiamo con u_i, $i = 1..n$, le componenti di $u \in X$ rispetto ad essa. Posto allora*

$$\alpha_{ij} \; = \; \alpha(e_i, e_j) \qquad i, j = 1..n$$

si ha

$$\alpha(u, v) \; = \; \sum_{i,j=1}^{n} \alpha_{ij} u_i v_j \qquad \forall u, v \in X$$

La matrice $A = (\alpha_{ij})$ così definita è detta *matrice associata ad* (o anche *matrice che rappresenta*) α rispetto alla base $e_1, e_2, ..., e_n$.

Prima di dimostrare quanto asserito, osserviamo subito che, essendo ogni forma quadratica definita a partire da una bilineare prendendone gli argomenti uguali, ne segue che ogni forma quadratica verifica, per un'opportuna matrice (α_{ij})

$$H(u) \; = \; \alpha(u, u) \; = \; \sum_{i,j=1}^{n} \alpha_{ij} u_i u_j$$

ed eccoci di nuovo al problema iniziale!
Dimostrazione.

Si ha $u = \sum_{i=1}^n u_i e_i$, $v = \sum_{j=1}^n v_j e_j$, e dunque

$$\alpha(u,v) \; = \; \alpha(\sum_{i=1}^n u_i e_i \; , \; \sum_{j=1}^n v_j e_j) \; =$$

(utilizzando la linearità rispetto al secondo argomento, fissato il primo)

$$= \; \sum_{j=1}^n v_j \alpha(\sum_{i=1}^n u_i e_i \; , \; e_j) \; =$$

(dalla linearità rispetto al primo argomento, fissati i secondi)

$$= \sum_{i,j=1}^n u_i v_j \alpha(e_i, e_j)$$

Dunque le forme bilineari sono completamente individuate dai loro valori sulle coppie di vettori della base. Inoltre

Teorema 3 *La matrice associata ad una forma bilineare è simmetrica se e solo se la forma è simmetrica.*

Dimostrazione.
Se la forma è simmmetrica, allora $\alpha_{ij} \equiv \alpha(e_i, e_j) = \alpha(e_j, e_i) = \alpha_{ji}$, e dunque la matrice associata è simmetrica. Analogamente, se la matrice è simmetrica, per il teorema precedente $\alpha(u,v) = \sum_{i,j=1}^n u_i v_j \alpha_{ij} = \sum_{i,j=1}^n v_j u_i \alpha_{ji} = \alpha(v,u)$

Per quanto segue è però opportuno spingere oltre questa linea di ragionamento.
Osserviamo infatti che di forme bilineari associate ad una fissata forma quadratica ce ne possono essere diverse: tornando all'esempio iniziale, basta scrivere $x^2 - xy$ sotto la forma $x^2 - \frac{1}{2}xy - \frac{1}{2}yx$ per dare luogo a due diverse forme bilineari $u_1v_1 - u_1v_2$ e $u_1v_1 - \frac{1}{2}u_1v_2 - \frac{1}{2}u_2v_1$, che però coincidono per $u = v = (x, y)$. In realtà, ogni forma quadratica che abbia termini "non diagonali", come xy si può ottenere da infinite forme bilineari distinte, ripartendo il coefficiente in proporzioni diverse fra il termine xy e quello yx. Se però si decide di ripartirlo in parti uguali, la forma è unica. Questo ragionamento può essere ripetuto anche in astratto:

Teorema 4 *Per ogni forma quadratica H su X esiste un'unica forma bilineare* simmetrica *β su X tale che*

$$H(u) \; = \; \beta(u,u) \qquad \forall u \in X$$

Dimostrazione. Dato che H è una forma quadratica, esiste una forma bilineare $\alpha(u,v)$ su X tale che $H(u) = \alpha(u,u)$ per ogni u. Si definisce allora

$$\beta(u,v) \; = \; \frac{1}{2}[\alpha(u,v) \; + \; \alpha(v,u)]$$

e si verifica immediatamente che β è una forma bilineare simmetrica coincidente con α per $u = v$. Resta da provare che è unica. Supponiamo che β e β' siano due forme bilineari simmetriche coincidenti con H per $u = v$. Allora

$$\beta(e_i + e_j, e_i + e_j) \; = \; H(e_i + e_j) \; = \; \beta'(e_i + e_j, e_i + e_j)$$

Poiché, per la bilinearità e la simmetria,

$$\beta(e_i, e_i) \; + \; \beta(e_j, e_j) \; + \; 2\beta(e_i, e_j) \; = \; \beta(e_i + e_j, e_i + e_j) \; =$$

$$= \beta'(e_i + e_j, e_i + e_j) \;=\; \beta'(e_i, e_i) \;+\; \beta'(e_j, e_j) \;+\; 2\beta'(e_i, e_j)$$

e inoltre

$$\beta(e_i, e_i) \;=\; \beta'(e_i, e_i) \qquad \beta(e_j, e_j) \;=\; \beta'(e_j, e_j)$$

ne segue infine

$$\beta(e_i, e_j) \;=\; \beta'(e_i, e_j) \qquad \forall i, j = 1..n$$

Dal teorema precedente, ne segue immediatamente che β e β' assumono valori identici su ogni coppia di vettori della base e, di conseguenza, su ogni coppia di vettori.

La simmetria della forma non è un simpatico accessorio: è un'ipotesi essenziale che verrà messa a profitto più avanti. Il teorema precedente esprime in astratto quanto risulta immediatamente nello studio del segno delle forme quadratiche su $\mathbb{R}^n$— il nostro problema iniziale — : ci si può sempre ricondurre ad una forma definita da una matrice simmetrica.

2.2 Diagonalizabilità

Quanto visto più su suggerisce la definizione seguente:

Definizione 3 *Una forma bilineare $\alpha(u, v)$ (o quadratica $\alpha(u, u)$) su uno spazio reale X di dimensione finita n verrà detta* diagonale *rispetto ad una base $e_1, e_2, ..., e_n$ se la matrice $\alpha_{ij} \;=\; \alpha(e_i, e_j)$ che la rappresenta è diagonale.*
Verrà inoltre detta diagonalizzabile *se esiste una base rispetto alla quale essa è diagonale.*

Per individuare condizioni convenienti di diagonalizzabilità, supponiamo che lo spazio X, oltre ad essere reale di dimensione finita, sia anche euclideo, ossia dotato di un prodotto scalare. In tal caso si può dare alle forme bilineari una diversa espressione facendo uso proprio del prodotto scalare.

Fissiamo allora una base ortonormale $e_1, e_2, ..., e_n$ e sia (α_{ij}) la matrice associata alla forma bilineare rispetto a tale base, definita più su. Consideriamo anche l'operatore lineare su X rappresentato dalla matrice (α_{ij}), definito da

$$\mathcal{A}v \;=\; \sum_{i,j=1}^{n} \alpha_{ij} v_j e_i$$

e osserviamo che, dalla bilinearità del prodotto scalare,

$$u(\mathcal{A}v) = (\sum_{h=1}^{n} u_h e_h)(\sum_{i,j=1}^{n} \alpha_{ij} v_j e_i) \;=\; \sum_{i,j,h=1}^{n} \alpha_{ij} u_h v_j e_h e_i$$

Poiché la base è ortonormale, $e_h e_i = 0$, salvo che per $h = i$, caso in cui vale 1. Dunque

$$u(\mathcal{A}v) \;=\; \sum_{i,j}^{n} \alpha_{ij} u_i v_j \;=\; \alpha(u, v)$$

Questa particolare rappresentazione delle forme bilineari come prodotti scalari nasconde un (piccolo) problema: la definizione di $\mathcal{A}$ da noi adottata dipende, almeno in apparenza, dalla base ortonormale scelta. Cosa accade se si cambia base? Nulla! Supponiamo che due operatori $\mathcal{A}$ e $\mathcal{A}$' siano il risultato di due diverse scelte di basi ortonormali. Si avrà comunque $\alpha(u, v) = u\mathcal{A}v =$ $= u\mathcal{A}'v \quad \forall u, v \in X$, da cui $u(\mathcal{A}v - \mathcal{A}'v) = 0 \quad \forall u, v \in X$. Basta scegliere nella relazione precedente $u = \mathcal{A}v - \mathcal{A}'v$ per ottenere $|\mathcal{A}v - \mathcal{A}'v|^2 = 0$ da cui infine $\mathcal{A}v = \mathcal{A}'v \quad \forall v \in X$. Dunque gli operatori $\mathcal{A}$ e $\mathcal{A}$' potranno pure essere definiti da formule diverse, ma assumono valori coincidenti per ogni vettore $v \in X$, e dunque coincidono. Riassumiamo quanto detto in un teorema.

Teorema 5 *Per ogni forma bilineare α su uno spazio euclideo reale X di dimensione n esiste un unico operatore lineare $\mathcal{A} : X \to X$ tale che*

$$\alpha(u, v) \ = \ u(\mathcal{A}v) \quad \forall u, v \in X$$

espresso, per ogni fissata base ortonormale $e_1, e_2, ..., e_n$, da

$$\mathcal{A}v \ = \ \sum_{i,j=1}^{n} \alpha_{ij} v_j e_i \quad \forall v \in X$$

ove v_j sono le componenti di v rispetto alla base, e $\alpha_{ij} \ = \ \alpha(e_i, e_j)$.

Si può ora provare il resultato principale di questa sezione.

Teorema 6 *Condizione necessaria e sufficiente perché la matrice associata alla forma bilineare α e alla base ortonormale $e_1, e_2, ..., e_n$ sia diagonale è che esistano $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{R}$ tali che*

$$\mathcal{A}e_i \ = \ \lambda_i e_i \quad \forall i = 1..n$$

Dimostrazione.
Poiché $e_1, e_2, ..., e_n$ formano una base ortonormale, decomponendo $\mathcal{A}e_j$ si avrà

$$\mathcal{A}e_j = \sum_{h=1}^{n} [(\mathcal{A}e_j)e_h] \, e_h$$

Poiché, inoltre, $\alpha_{ij} = \alpha(e_i, e_j) = e_i(\mathcal{A}e_j)$, ne segue che $\alpha_{ij} = 0$ per $i \neq j$ se e solo nella decomposizione ortogonale precedente $(\mathcal{A}e_j)e_h = 0 \quad \forall h \neq j$. La tesi segue immediatamente ponendo $\lambda_j = (\mathcal{A}e_j)e_j \quad j = 1..n$.

Un'osservazione fondamentale per il seguito è che, una volta trovata la base che diagonalizza la forma si ha, oltre che $\alpha_{ij} = 0$ se $i \neq j$, anche $\alpha_{ii} = e_i(\mathcal{A}e_i) = \lambda_i$, e dunque i termini diagonali si debbono ricercare fra i valori λ per i quali l'equazione vettoriale $\mathcal{A}u = \lambda u$ ha soluzioni non nulle.

La diagonalizzabilità, ossia l'esistenza di una base rispetto alla quale la matrice associata alla forma bilineare è diagonale, è dunque equivalente al fatto che esistano basi ortonormali i vettori delle quali sono mutati da $\mathcal{A}$ in loro multipli. Non sembra un gran guadagno, ma è questo il problema che verrà effettivamente affrontato e in gran parte risolto nelle pagine che seguono. Verrà risolto del tutto per le forme simmetriche. Visto che l'apparato teorico esposto finora non è del tutto rassicurante, è opportuno riprendere gli esempi iniziali alla luce delle considerazioni sin qui svolte.

Data una forma quadratica "classica" su $\mathbb{R}^n$, ossia un polinomio omogeneo di grado 2 in n variabili,

$$\alpha(x, x) \ = \ \sum_{i,j=1}^{n} \alpha_{ij} x_i x_j$$

ove $A \equiv (\alpha_{ij})$ è una matrice $n \times n$ data, si osserva subito che la matrice associata ad essa rispetto alla base canonica è proprio (α_{ij}), e dunque l'operatore corrispondente è $\mathcal{A}(x) = \sum_{i,j=1}^{n} \alpha_{ij} x_j e_i$, che altro non è che l'operatore definito dal prodotto della matrice data A per il vettore $(x_1, x_2, ..., x_n)$ delle componenti. La matrice A rappresenta inoltre l'operatore $\mathcal{A}$ rispetto alla base canonica. Nessuna sorpresa, dunque, e nessun calcolo da fare: la matrice che definisce la forma quadratica è la stessa che rappresenta, rispetto alla base canonica, tanto la forma quadratica "astratta" associata alla forma "concreta" da cui siamo partiti, quanto l'operatore $\mathcal{A}$ ad essa associato.

2.3 Lo spettro di un operatore lineare

Definizione 4 *Sia X uno spazio vettoriale reale o complesso e sia $\mathcal{A} : X \to X$, lineare. Ogni scalare λ per il quale esistono soluzioni non nulle $u \in X$ dell'equazione $\mathcal{A}u = \lambda u$ verrà chiamato* autovalore *di $\mathcal{A}$, ed ognuna di tali soluzioni non nulle verrà chiamato* autovettore *relativo a λ; l'insieme costituito dallo zero e da tutti gli autovettori relativi ad un autovalore λ verrà chiamato* autospazio *di λ.*
L'insieme di tutti gli autovalori di $\mathcal{A}$ verrà chiamato spettro *di $\mathcal{A}$.*

Gli autovalori sono talvolta chiamati anche *valori propri* (eigenvalues, in inglese).
Osserviamo che l'autospazio di un fissato autovalore λ è per definizione il nucleo di $\mathcal{A} - \lambda I$ (ove I è l'operatore identico da X in sè) ed è dunque effettivamente un sottospazio vettoriale di X. In particolare, multipli di autovettori sono ancora autovettori.
Inoltre notiamo che 0 è autovalore di $\mathcal{A}$ se e solo se $Ker\mathcal{A}$ contiene vettori non nulli, che costituiscono gli autovettori relativi a 0. Osserviamo anche il legame con il problema della sezione precedente: essendo non nulli, gli autovettori sono i candidati ideali a formare le basi necessarie per diagonalizzare le forme. Dal teorema finale della sezione precedente scaturisce la

Definizione 5 *Dato uno spazio vettoriale (reale o complesso) X e un operatore $\mathcal{A} : X \to X$, diremo che tale operatore è* diagonalizzabile *se esistono basi di X formate da autovettori di $\mathcal{A}$.*

Una base di X formata da autovettori di $\mathcal{A}$ viene anche detta *base spettrale* di $\mathcal{A}$ e i teoremi di esistenza di tali basi vengono anche detti *teoremi spettrali.* Utilizzando questa terminologia, $\mathcal{A}$ è diagonalizzabile se e solo se X ammette basi spetttrali di $\mathcal{A}$.

L'esistenza degli autovalori è questione delicata, e verrà affrontata in breve. Per ora osserviamo che, data una matrice quadrata reale A, si possono definire *due* operatori ponendo $\mathcal{A}(u) = Au$, uno su $\mathbb{R}^n$ e uno su $\mathbb{C}^n$, che danno luogo a *due* ben diversi problemi agli autovalori, reali e complessi. Consideriamo l'esempio:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Posto $\mathcal{A}(u) = Au$ e $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, l'equazione vettoriale $\mathcal{A}(u) = \lambda u$ diventa

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \lambda u_1 \\ \lambda u_2 \end{pmatrix}$$

Eseguendo il prodotto e raccogliendo a primo membro si ottiene il sistema lineare

$$\begin{cases} -\lambda u_1 - u_2 = 0 \\ u_1 - \lambda u_2 = 0 \end{cases}$$

che ha soluzioni non nulle se e solo se è nullo il determinante dei coefficienti e cioè

$$\begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = 0$$

ovvero

$$\lambda^2 + 1 = 0$$

Tale equazione presenta un comportamento molto diverso a seconda che si cerchino soluzioni reali o ci si accontenti di quelle complesse: nel primo caso non ci sono autovalori, nel secondo ce ne sono due distinti i e $-i$. La straordinaria potenza del teorema di Gauss, non a caso chiamato anche teorema *fondamentale* dell'Algebra, garantisce che ogni polinomio di grado non nullo ha soluzioni in $\mathbb{C}$, mentre non può esistere un simile teorema nel campo reale, come mostra l'esempio precedente. Vedremo presto che il problema dell'esistenza degli autovettori nel campo reale comporta di conseguenza maggiori complicazioni, ineludibili, rispetto a quello negli spazi complessi.

Per esercizio, calcoliamo l'autospazio di $\lambda = i$ nell'esempio precedente.
Sostituendo nel sistema degli autovettori $\lambda = i$, si ottiene

$$\begin{cases} -iu_1 - u_2 = 0 \\ u_1 - iu_2 = 0 \end{cases}$$

Poiché il determinante è nullo si può eliminare una delle equazioni, ad esempio la prima, dipendente dall'altra, e il sistema si riduce all'equazione $u_1 = iu_2$. Tale equazione è risolta da tutti i vettori $\begin{pmatrix} it \\ t \end{pmatrix}$, $\forall t \in \mathbb{C}$. L'autospazio relativo ad i, dunque, è generato da $\begin{pmatrix} i \\ 1 \end{pmatrix}$, e gli autovettori sono tutti multipli non nulli di tale vettore.

2.4 Indipendenza di autovettori

Prima di affrontare la questione dell'esistenza degli autovettori, proviamo un importante teorema che consente di ottenere immediatamente un primo resultato di diagonalizzazione.

Teorema 7 *Sia $\mathcal{A}$ un operatore su X, spazio vettoriale reale o complesso. Siano $u_1, u_2, ..., u_k$ autovettori relativi ad autovalori a due a due distinti.*
Allora $u_1, u_2, ..., u_k$ sono indipendenti.

Dimostrazione.
Proveremo la tesi per induzione su k.
Siano u e v autovettori di $\mathcal{A}$ relativi rispettivamente a λ e μ, con $\mu \neq \lambda$. Se u e v fossero dipendenti, esisterebbe uno scalare α tale che $u = \alpha v$, da cui

$$\mathcal{A}(u) \ = \ \lambda u \ = \ \alpha\lambda v$$

ma anche

$$\mathcal{A}(u) \ = \ \mathcal{A}(\alpha v) \ = \ \alpha\mathcal{A}(v) \ = \ \alpha\mu v$$

e sottraendo membro a membro

$$\alpha(\lambda - \mu)v \ = \ 0$$

Essendo $v \neq 0$ in quanto autovettore, e $\lambda \neq \mu$, ne segue $\alpha = 0$ e dunque $u = 0v = 0$, il che contraddice l'ipotesi che u sia autovettore. La tesi è dunque verificata per $k = 2$.

Supponiamo ora la tesi verificata per ogni sistema di $k - 1$ autovettori appartenenti ad autospazi distinti e proviamolo per il nostro.
Supponiamo per assurdo che $u_1, u_2, ..., u_k$ siano dipendenti, e siano $\lambda_1, \lambda_2, ..., \lambda_k$ gli autovalori a due a due distinti ad essi relativi. Si avrebbe allora

$$\sum_{i=1}^{k} \beta_i u_i \ = \ 0$$

per opportuni $\beta_1, \beta_2, ..., \beta_k$ non tutti nulli. Scelto uno degli indici non nulli, e sia per semplicità β_1, risolvendo la precedente equazione rispetto a u_1 si ottiene

$$u_1 \ = \ \sum_{i=2}^{k} \beta'_i u_i$$

ove $\beta'_i \ = \ -\beta_i/\beta_1 \quad i = 2..k$, e non possono annullarsi tutti in quanto u_1, essendo un autovettore, è non nullo. Si ha dunque, da un canto

$$\mathcal{A}(u_1) \ = \ \lambda_1 u_1 \ = \ \lambda_1 \sum_{i=2}^{k} \beta'_i u_i \ = \ \sum_{i=2}^{k} \beta'_i \lambda_1 u_i$$

e dall'altro

$$\mathcal{A}(u_1) \; = \; \mathcal{A}(\sum_{i=2}^{k} \beta_i' u_i) \; = \; \sum_{i=2}^{k} \beta_i' \mathcal{A}(u_i) \; = \; \sum_{i=2}^{k} \beta_i' \lambda_i u_i \; = \; \sum_{i=2}^{k} \beta_i' \lambda_i u_i$$

da cui, sottraendo membro a membro segue infine

$$\sum_{i=2}^{k} \beta_i' (\lambda_i - \lambda_1) u_i \; = \; 0$$

Poiché è stato osservato che i β_i non sono tutti nulli, e le differenze in parentesi sono tutte diverse da zero per ipotesi, ne segue che i coefficienti della combinazione lineare che appare nell'ultima equazione sono non tutti nulli e quindi i $k-1$ autovettori $u_2, .., u_k$ sono dipendenti, contro l'ipotesi induttiva.

Si osservi che non è mai stato necessario assumere che X sia di dimensione finita. In qualunque spazio, autovettori di autospazi distinti (in numero finito) sono indipendenti. Vedremo più avanti un esempio molto interessante di insospettabile applicazione di tale resultato.
Prima di fare ciò però, passiamo a riscuotere.

Teorema 8 *Sia $\mathcal{A} : X \to X$ uno operatore lineare su uno spazio X di dimensione n, e supponiamo che $\mathcal{A}$ possieda n autovalori a due a due distinti.*
Allora X possiede una base formata da autovettori di $\mathcal{A}$, che è dunque diagonalizzabile.

Dimostrazione.
Infatti, essendo gli autovalori distinti a due a due, si può applicare il teorema precedente, e basta scegliere in ogni autospazio un qualsiasi autovettore per ottenere un sistema di n vettori indipendenti. Essendo X di dimensione n, essi costituiscono una base di X, e la tesi segue dalla condizione caratteristica di diagonalizzabilità.

Nel caso l'equazione caratteristica, risolta dagli autovalori (vedi più avanti), abbia radici multiple, non sarà purtroppo possibile ottenere risultati così semplici ed eleganti, e bisognerà introdurre nuove ipotesi. Infatti:

Teorema 9 *Esistono operatori non diagonalizzabili.*

Dimostrazione.
Sia $\mathcal{A}(u) = Au$ ove $A \in \mathbb{R}^{2\times 2}$ è definita da

$$A \; = \; \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Proviamo che $\mathcal{A}$ non è diagonalizzabile. Posto $u = \begin{pmatrix} x \\ y \end{pmatrix}$ l'equazione $\mathcal{A}(u) = \lambda u$ diventa

$$\begin{cases} -\lambda x + y = 0 \\ -\lambda y = 0 \end{cases}$$

Il sistema precedente ha soluzioni non nulle se e solo se si annulla il determinante dei coefficienti e cioè se e solo se $-\lambda^2 = 0$ e dunque se e solo se $\lambda = 0$. Sostituendo nel sistema precedente si ottiene

$$y = 0$$

e dunque le sue soluzioni non nulle sono tutte e sole quelle del tipo $\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ con $\alpha \neq 0$. Evidentemente, ogni vettore di $X = \mathbb{R}^2$ con la seconda componente non nulla non può essere generato da nessuna combinazione di autovettori aventi tutti la seconda componente nulla. Dunque, non esistono basi formate solo da autovettori di $\mathcal{A}$.

Può sembrare strano che tutti gli esempi "concreti" visti sinora riguardino matrici non simmetriche che, tutto sommato, ci interessano assai poco, visto che le forme quadratiche su $\mathbb{R}^n$ possono sempre essere trasformate in forme simmetriche. In realtà, la simmetria della matrice produce una serie di conseguenza benefiche; per esse, gli autovalori sono tutti reali e, anche quando non sono tutti distinti, consentono comunque di determinare basi spettrali, e dunque entrambi i contresempi prima presentati non potevano che riguardare matrici non simmetriche. La via è comunque ancora lunga ed è tempo di affrontare il problema dell'esistenza degli autovalori.

2.5 Il teorema di esistenza degli autovettori

Una conseguenza diretta del teorema fondamentale dell'Algebra di Gauss è il seguente resultato, chiave di volta della teoria della diagonalizzazione nel caso complesso.

Teorema 10 *Sia X uno spazio complesso di dimensione finita, non nulla, e sia $\mathcal{A}$ un operatore lineare da X in sè. Allora X contiene autovettori di $\mathcal{A}$, e cioè esistono $\lambda \in \mathbb{C}$, e $u \in X$ non nullo, tali che*

$$\mathcal{A}u \; = \; \lambda u$$

Dimostrazione.
Sia $x_1, x_2, ..., x_n$ una base di X, e sia α_{ij} la matrice che rappresenta $\mathcal{A}$ rispetto alla base $x_1, x_2, ..., x_n$, verificante dunque $\mathcal{A}(u) \; = \; \sum_{i,j=1}^{n} \alpha_{ij} u_j x_i$. Dato $u \in X$ e posto $u = \sum_{i=1}^{n} u_i x_i$, dall'indipendenza dei vettori della base segue che u è nullo se e solo se il vettore delle sue componenti (u_i) è il vettore nullo in $\mathbb{C}^n$. Dunque, $\mathcal{A}$ ha autovettori in X se e solo se esistono uno scalare $\lambda \in \mathbb{C}$ e un vettore $(u_i) \in \mathbb{C}^n$ non nullo, tali che

$$\sum_{i,j=1}^{n} \alpha_{ij} u_j x_i \; = \; \lambda \sum_{i=1}^{n} u_i x_i$$

e dunque

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{n} \alpha_{ij} u_j - \lambda u_i \right] x_i \; = \; 0$$

Essendo i vettori x_i indipendenti, ciò accade se e solo se i coefficienti degli x_i sono tutti nulli, e cioè se e solo se

$$\sum_{j=1}^{n} \alpha_{ij} u_j \; - \; \lambda u_i \; = \; 0 \qquad i = 1..n$$

Il sistema lineare omogeneo precedente ha soluzioni u_j non nulle se e solo se il determinante della matrice dei suoi coefficienti è nullo, ed essendo tale determinante un polinomio in λ di grado maggiore o uguale uno, per il teorema di Gauss si annulla certamente per qualche $\lambda \in \mathbb{C}$. Dunque tali λ saranno gli autovalori e, per ognuno di essi e per ogni soluzione non nulla $(u_1, u_2, ..., u_n) \in \mathbb{C}^n$ del precedente sistema si otterrà l'autovettore $u = \sum_{i=1}^{n} u_i x_i$ in X.

Il teorema precedente verrà nel seguito chiamato *teorema dei sottospazi invarianti*, in quanto asserisce l'esistenza di autovettori in ogni sottospazio del dominio dell'operatore che sia invariante, cioè mutato in sè, da $\mathcal{A}$. In tale veste verrà impiegato nella prova del teorema spettrale.

La condizione precedente di annullamento del determinante del sistema degli autovettori è cruciale in quanto individua gli autovalori di $\mathcal{A}$. Tale condizione sembra però dipendere dalla base scelta per rappresentare $\mathcal{A}$, il che non è. Prima di affrontare la questione, introduciamo la seguente

Definizione 6 *Data una matrice* $A \in \mathbb{R}^{n\times n}$*, si definisce* equazione caratteristica di A *l'equazione*

$$det(A - \lambda I) \;=\; 0$$

Il polinomio $p(\lambda) \;=\; det(A - \lambda I)$ *verrà detto* polinomio caratteristico *di* A*.*

Le soluzioni complesse dell'equazione caratteristica della matrice A associata ad un operatore $\mathcal{A}$ forniscono esattamente i valori di λ per i quali il determinante del sistema lineare risolto dalle componenti rispetto alla base degli autovettori $Au = \lambda u$ si annulla, e per i quali, di conseguenza, tale sistema possiede soluzioni non nulle.
Si è fatto cenno alla sgradevole possibilità che le soluzioni dell'equazione caratteristica dipendano dalla base scelta, al variare della quale varia la matrice $A = (\alpha_{ij})$ che appare nel sistema lineare. Non è così:

Teorema 11 *Il polinomio caratteristico della matrice di rappresentazione di un operatore lineare da uno spazio di dimensione finita in sè non dipende dalla base scelta per rappresentarlo.*

Dimostrazione.
Ricordando come cambia la matrice associata ad un operatore lineare al passaggio da una base ad un altra mediante una matrice M, segue subito dal teorema di Binet sul determinante di matrici prodotto e dell'inversa che

$$det\ (M^{-1}AM - \lambda I) \;=\; det\ [M^{-1}AM - \lambda M^{-1}IM] \;=\; det\ [M^{-1}(A - \lambda I)M] \;=$$
$$=\; det\ M^{-1}\,\det\,(A - \lambda I)\, det\ M \;=\; det\ (A - \lambda I)$$

e dunque i due polinomi caratteristici delle due matrici sono uguali e ogni soluzione dell'equazione caratteristica rispetto alla nuova base lo è anche rispetto alla vecchia e viceversa: in parole più semplici, gli autovalori di un operatore non dipendono dalla base scelta per rappresentarlo e, cosa utilissima per il seguito, possono essere calcolati su una qualunque delle matrici che lo rappresentano. Per questa ragione si può porre la

Definizione 7 *Dato un operatore* $\mathcal{A}$ *si definisce polinomio caratteristico di* $\mathcal{A}$ *il polinomio caratteristico della matrice che lo rappresenta rispetto a una qualunque base.*

Queste ultime considerazioni contribuiscono a rasserenare ulteriormente il quadro della situazione. Data una forma quadratica "classica", infatti, è stato già osservato che la matrice che la definisce rappresenta anche l'operatore $\mathcal{A}$ rispetto alla base canonica. Poiché gli autovalori della matrice di rappresentazione sono invarianti rispetto alla base, possono essere determinati risolvendo l'equazione caratteristica della matrice iniziale $\det(A - \lambda I) = 0$, senza ulteriori calcoli. In particolare, il loro segno (concorde o meno con gli altri) può essere studiato addirittura senza che sia necessario neppure calcolarli, come si vedrà più avanti.

2.6 Un'applicazione insolita

Prima di passare al teorema spettrale, presentiamo un'applicazione notevole del teorema dell'indipendenza degli autovettori in autospazi distinti.
Studiando le equazioni differenziali lineari a coefficienti costanti si incontra il problema di stabilire l'indipendenza delle funzioni

$$e^{\alpha_1 t}, e^{\alpha_2 t}, \cdots, e^{\alpha_n t}$$

quando le costanti α_i siano a due a due distinte. Se esse sono reali, il problema è ragionevolmente semplice, ma alcune equazioni di importanza fondamentale (l'oscillatore armonico, ad esempio) necessitano, o quanto meno beneficiano fortemente, dell'introduzione e della manipolazione di esponenziali complessi (vedi anche: R.Feynman "Lectures on Physics" vol I). L'indipendenza di funzioni esponenziali complesse di variabile reale a coefficienti distinti è assai più delicata. La soluzione classica (il determinante di Vandermonde) è considerevolmente macchinosa. La via che seguiremo qui è di provare che le funzioni esponenziali viste più su sono autovettori di un opportuno operatore su un opportuno spazio.

2.6.1 L'esponenziale complesso

Per prima cosa definiamo la funzione esponenziale complessa di variabile reale t.

Definizione 8 *Per ogni $t \in \mathbb{R}$ ed ogni $\lambda \in \mathbb{C}$, con $\lambda = a + ib$ si pone:*

$$e^{\lambda t} \equiv e^{at}[\cos(bt) + i\sin(bt)]$$

Poiché $|e^{\lambda t}| = |e^{at}||\cos(bt) + i\sin(bt)|$ e $|\cos(bt) + i\sin(bt)| = 1 \quad \forall t \in \mathbb{R}$, ne segue subito che $e^{\lambda t} \neq 0 \quad \forall t \in \mathbb{R}\ \forall \lambda \in \mathbb{C}$.

Più in generale, ad ogni funzione complessa f di variabile reale t si possono associare due funzioni reali g e h, la *parte reale* e la *parte immaginaria*, definite da

$$f(t) = g(t) + ih(t)$$

Poiché ad ogni numero complesso $f(t)$ sono associate biunivocamente la sua componente reale e quella immaginaria, ne segue che ogni funzione complessa è individuata dalle sue componenti al variare di t.

2.6.2 La derivata e lo spazio $C^\infty(\mathbb{R})$

Alle funzioni complesse di variabile reale, così come alle curve in $\mathbb{R}^n$, si possono estendere i concetti del calcolo differenziale.

Definizione 9 *Data una funzione $f(t) = g(t) + ih(t)$, si definisce la sua* derivata *ponendo*

$$f'(t) \equiv g'(t) + ih'(t)$$

in ogni punto nel quale entrambe le componenti reali g e h siano derivabili.
Si definisce lo spazio $C^\infty(\mathbb{R})$ come lo spazio vettoriale complesso di tutte le funzioni complesse di variabile reale derivabili quante volte si vuole su tutto $\mathbb{R}$. Somma e prodotto per uno scalare sono definiti punto per punto. Lo zero dello spazio è la funzione che sommata a qualunque altra la lascia inalterata, e quindi è la funzione identicamente nulla.

Denoteremo per brevità con X lo spazio vettoriale complesso or ora introdotto.

Teorema 12 *La derivata è un operatore lineare su X.*

Dimostrazione.
Siano $\phi = f + ig$ e $\psi = h + ik$ due funzioni di X, e sia anche $\lambda = a + ib$ uno scalare complesso. Si ha allora:

$$[\phi+\psi]' = [f+h+i(g+k)]' \equiv (f+h)'+i(g+k)' = f'+h'+i(g'+k') = f'+ig'+h'+ik' \equiv \phi'+\psi'$$

e anche, ricordando che a e b sono costanti,

$$(\lambda\phi)' = [(a+ib)(f+ig)]' = [(af-bg)+i(ag+bf)]' \equiv (af-bg)'+i(ag+bf)' =$$

$$= af'-bg'+i(ag'+bf') = (a+ib)(f'+ig') = \lambda\phi'$$

Ci attende ora un interessante esercizio: calcolare la derivata della funzione esponenziale complessa. Si può immaginare il resultato, ed infatti si ha:

$$(e^{\lambda t})' \equiv [e^{at}\cos(bt)]' + i[e^{at}\sin(bt)]' =$$

$$= ae^{at}\cos(bt) - be^{at}\sin(bt) + i[ae^{at}\sin(bt) + be^{at}\cos(bt)] = (a+ib)e^{at}[\cos(bt) + i\sin(bt)] = \lambda e^{\lambda t}$$

Non ci sono sorprese: l'esponenziale a valori complessi si deriva esattamente come quello reale. Il fatto che può però sfuggire ad una prima osservazione è che il "vettore non nullo" $e^{\lambda t}$ in X è un autovettore, relativo all'autovalore λ, dell'operatore definito su X dalla derivata : è proprio ciò che abbiamo appena calcolato! Ne segue dal teorema generale sull'indipendenza di autovettori, senza ulteriori calcoli, che funzioni esponenziali complesse, con coefficienti nell'esponente a due a due distinti e in numero finito arbitrario, sono linearmente indipendenti.
Non è facile apprezzare la potenza di questo resultato sino a che non si prova a dimostrarlo direttamente, per esempio per esponenziali con coefficienti complessi distinti di modulo unitario.

Capitolo 3

Il teorema spettrale per operatori autoaggiunti

L'esistenza di una base spettrale è abbastanza agevole da provare per una classe di operatori molto importanti in pratica: gli operatori autoaggiunti.

Inizieremo con l'introdurre la proprietà fondamentale che assumeremo come definizione; poi dedurremo da essa quanto è necessario per provare il teorema spettrale; infine, scopriremo che provare che un operatore è autoaggiunto richiede solo una facile verifica sulla matrice associata all'operatore rispetto ad una qualunque base e stabiliremo i legami fra operatori autoaggiunti e matrici simmetriche.

3.1 Gli operatori autoaggiunti

Sia X uno spazio euclideo complesso. Iniziamo con la definizione:

Definizione 10 *Sia $\mathcal{A}$ un operatore lineare da X in X. Allora $\mathcal{A}$ verrà detto* autoaggiunto *se e solo se verifica*

$$(\mathcal{A}u)v \;=\; u(\mathcal{A}v) \qquad \forall u, v \in X$$

I prodotti scalari vanno intesi in X.
La questione se tali operatori esistano o meno verrà affrontata e risolta alla fine del capitolo. Per ora limitiamoci a dedurre alcune interessanti conseguenza di tale proprietà.

3.1.1 Ogni autovalore è reale

Teorema 13 *Sia $\mathcal{A} : X \to X$ autoaggiunto. Allora ogni suo autovalore è reale.*

Dimostrazione. Siano $\lambda \in \mathbb{C}$ e $u \in X$ rispettivamente un autovalore e un corrispondente autovettore. Allora si ha

$$\lambda|u|^2 \;=\; (\lambda u)u \;=\; (\mathcal{A}u)u \;=\; u(\mathcal{A}u) \;=\; u(\lambda u) \;=\; \overline{\lambda}|u|^2$$

Essendo $u \neq 0$ per definizione di autovettore, dividendo per $|u|^2$ segue $\lambda = \overline{\lambda}$, e dunque λ è reale.

3.1.2 Il complemento ortogonale di un autovettore è invariante

Il prossimo risultato, sebbene elementare, è cruciale nella prova del teorema spettrale.

Teorema 14 *Sia $\mathcal{A} : X \to X$ autoaggiunto, e siano $\lambda \in \mathbb{C}$ e $u \in X$ rispettivamente un autovalore e un corrispondente autovettore. Allora, se w è ortogonale ad u, anche $\mathcal{A}w$ lo è.*

Dimostrazione.
Si ha:

$$u(\mathcal{A}w) \;=\; (\mathcal{A}u)w \;=\; (\lambda u)w \;=\; \lambda uw$$

da cui segue immediatamente la tesi.

Una conseguenza immediata è che, detto $W = \{w \in X : wu = 0\}$, da $w \in W$ segue subito $\mathcal{A}w \in W$, e dunque il complemento ortogonale di ogni autovettore è invariante, cioè mutato in sè, per $\mathcal{A}$.

3.1.3 Autovettori in autospazi distinti sono ortogonali

Il lemma seguente non è necessario alla dimostrazione del teorema spettrale, ma è interessante di per sè.

Teorema 15 *Sia $\mathcal{A} : X \to X$ autoaggiunto. Allora autovettori in autospazi distinti sono ortogonali.*

Dimostrazione.
Siano λ e μ due autovalori distinti e u e v due autovettori ad essi corrispondenti. Si ha:

$$\lambda uv \;=\; (\mathcal{A}u)v \;=\; u(\mathcal{A}v) \;=\; u(\mu v) \;=\; \overline{\mu}uv$$

Essendo μ reale per il resultato precedente, ne segue $\lambda uv = \mu uv$, e dunque $(\lambda - \mu)uv = 0$. Poiché $\lambda - \mu \neq 0$, ne segue la tesi.

Questo resultato rafforza quello precedente che asseriva l'*indipendenza* degli autovettori di autospazi distinti. Nel caso degli operatori autoaggiunti si ha in più l'ortogonalità.

3.2 Il teorema spettrale complesso

Si può ora enunciare e provare il teorema spettrale, per lo meno nel caso complesso che, ad onta del nome, risulta notevolmente più semplice da trattare di quello reale.

Teorema 16 *Sia $\mathcal{A} : X \to X$ un operatore lineare autoaggiunto su X, spazio vettoriale complesso di dimensione finita. Allora esiste una base ortonormale di X costituita da autovettori di $\mathcal{A}$.*

Dimostrazione.
Si procede per induzione sulla dimensione di X.
Poiché X è uno spazio invariante per $\mathcal{A}$, per il teorema di esistenza degli autovettori esso contiene un autovettore $u \neq 0$ di $\mathcal{A}$. Poiché i multipli non nulli di un autovettore sono a loro volta autovettori, si può sostituire ad u il suo versore, che indicheremo ancora con u.

Se $dim(X) = 1$, la base ortonormale richiesta è costituita da u stesso.

Supponiamo ora la tesi vera per gli spazi di dimensione strettamente minore di n, e proviamola per X, supposto di dimensione n.
Si consideri il complemento ortogonale di u:

$$W \;=\; \{w \in X : wu = 0\}$$

Dal lemma precedente, segue che $\mathcal{A}$ trasforma W in sè.

Poiché $u \in X$ ma $u \notin W$, ne segue che $dim\ W \;<\; dim\ X$. Per l'ipotesi induttiva, esisterà una base di W formata da autovettori di $\mathcal{A}$, di norma unitaria e a due a due ortogonali, e siano $w_1, \ldots, w_k$.

Proviamo che $u, w_1, \dots, w_k$, tutti autovettori, formano anche una base ortonormale per X.

Poiché $w_1, w_2, ..., w_k$ appartengono a W, sono tutti ortogonali ad u, ed essendo anche ortogonali fra loro e di norma unitaria per l'ipotesi induttiva, ne segue che $u, w_1, w_2, ..., w_k$ sono versori a due a due ortogonali, e quindi indipendenti.
Resta dunque da verificare solo che essi generano X. Basta osservare che, per ogni $x \in X$, si ha $x = P_u x + (x - P_u x)$, e dunque x è somma di un multiplo di u — la sua proiezione $P_u x$ su u — e di $w = x - P_u x$, ortogonale ad u per il teorema della proiezione, e di conseguenza appartenente a W, e combinazione di $w_1, \dots, w_k$.

È utile osservare che la costruzione stessa della base implica che ad ogni passo induttivo si aggiunga alla base precedente un vettore (u) ortogonale a tutto lo spazio da essa generato. Un lemma precedente garantiva sì l'ortogonalità, ma solo per gli autovettori relativi ad autovalori distinti. La dimostrazione presentata, invece, garantisce comunque l'ortogonalità degli autovettori della base spettrale, indipendentemente dal fatto che siano o meno relativi ad autovalori coincidenti, ossia a radici multiple dell'equazione caratteristica.

3.3 Esistono operatori autoaggiunti?

La semplicità e l'eleganza del teorema spettrale nel caso autoaggiunto non deve far dimenticare la questione di fondo se esistano o meno operatori autoaggiunti. Naturalmente, accanto a tale questione di principio ve n'è un'altra di pari importanza: come verificare in pratica se un operatore sia o no autoaggiunto.
La sezione seguente risponde ad entrambe le questioni, individuando delle condizioni di semplice verifica.

3.3.1 Una condizione caratteristica

In cerca di condizioni necessarie sulla matrice di rappresentazione, supponiamo che $\mathcal{A}$ sia un operatore autoaggiunto, che $x_1, x_2, ..., x_n$ sia una base ortonormale di X, e che (α_{ij}) sia la matrice che rappresenta $\mathcal{A}$ rispetto a tale base. Allora, dalla definizione di operatore autoaggiunto, segue subito che

$$(\mathcal{A}x_i)x_j = x_i(\mathcal{A}x_j)$$

Dalla definizione di matrice di rappresentazione si ha anche

$$\mathcal{A}u \; = \; \sum_{h,k=1}^{n} \alpha_{hk} u_k x_h$$

Sostituendo il generico vettore della base x_i al posto di u nella precedente espressione, e osservando che le componenti di x_i rispetto alla base, da sostituire al posto di u_k, valgono tutte 0 salvo u_i che vale 1, ne segue che nella somma precedente tutti i termini con $k \neq i$ sono nulli, e dunque

$$\mathcal{A}x_i \; = \; \sum_h \alpha_{hi} x_h$$

da cui, essendo $x_1, x_2, ..., x_n$ ortonormale, segue infine che

$$\alpha_{ji} \; = \; (\mathcal{A}x_i)x_j = x_i(\mathcal{A}x_j) = \overline{\alpha_{ij}}$$

Dunque, le matrici di rappresentazione degli operatori autoaggiunti rispetto alle basi ortonormali devono verificare la condizione $\alpha_{ji} \; = \; \overline{\alpha_{ij}}$. Osserviamo subito che la condizione è di verifica immediata per tutti gli operatori definiti da matrici, per i quali la matrice di rappresentazione rispetto alla base canonica è nota *a priori*, in quanto è la matrice stessa. Osserviamo anche che la condizione necessaria appena trovata è certamente verificata dalle matrici simmetriche reali.
Le cose, in realtà, vanno ancora meglio: la condizione, oltre che necessaria, è anche sufficiente.

Teorema 17 *Condizione necessaria e sufficiente perché un operatore $\mathcal{A} : X \to X$ sia autoaggiunto su uno spazio euclideo complesso di dimensione finita X è che la sua matrice di rappresentazione α_{ij} rispetto ad una qualunque base ortonormale, verifichi:*

$$\alpha_{ij} \;=\; \overline{\alpha_{ji}}$$

Dimostrazione.
Resta da provare solo la sufficienza della condizione. Sia dunque $\mathcal{A}$ un operatore e $x_1, x_2, ..., x_n$ una base rispetto alla quale esso è rappresentato dalla matrice (α_{ij}), verificante $\alpha_{ij} \;=\; \overline{\alpha_{ji}}$. Ne segue, ragionando come nella prova della necessità della condizione, che

$$(\mathcal{A}x_i)x_j \;=\; \alpha_{ji} \;=\; \overline{\alpha_{ij}} \;=\; x_i(\mathcal{A}x_j) \qquad \forall i, j = 1..n$$

Siano ora $u, v \in X$. Si ha $u = \sum_1^n u_i x_i$ e $v = \sum_1^n v_i x_i$ e dunque, ricordando che la base è ortonormale, e dunque $x_i x_h = 0$ se $i \neq h$ mentre $x_i x_h = 1$ se $i = h$, e che

$$\mathcal{A}u \;=\; \sum_{i,j=1}^{n} \alpha_{ij} u_j x_i$$

risulta

$$(\mathcal{A}u)v \;=\; (\sum_{i,j=1}^{n} \alpha_{ij} u_j x_i)(\sum_{h=1}^{n} v_h x_h) \;=\; \sum_{i,j,h=1}^{n} \alpha_{ij} u_j \overline{v_h} x_i x_h \;=\; \sum_{i,j=1}^{n} \alpha_{ij} u_j \overline{v_i}$$

mentre

$$u(\mathcal{A}v) \;=\; (\sum_{1}^{n} u_h x_h)(\sum_{i,j=1}^{n} \alpha_{ij} v_j x_i) \;=\; \sum_{i,j=1}^{n} \overline{\alpha_{ij}} u_i \overline{v_j} \;=\; \sum_{i,j=1}^{n} \alpha_{ji} u_i \overline{v_j}$$

e le due espressioni risultano eguali (cambia solo il nome degli indici) per ogni coppia di vettori u e v in X. Infine, essendo l'operatore autoaggiunto, verificherà la condizione necessaria sulla matrice di rappresentazione rispetto ad ogni base ortonormale, come è stato provato prima del teorema.

3.4 Operatori autoaggiunti reali

Ancora due parole sugli operatori autoaggiunti reali, cioè definiti da una matrice reale, strettamente correlati al problema iniziale. Dal teorema precedente segue immediatamente il

Teorema 18 *Sia $A \in \mathbb{R}^{n\times n}$ e si consideri l'operatore $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^n$ definito da $\mathcal{A}(u) = Au$. Allora*

- *$\mathcal{A}$ è rappresentato da A, rispetto alla base canonica (ortonormale in $\mathbb{C}^n$)*
- *$\mathcal{A}$ trasforma $\mathbb{R}^n$ in $\mathbb{R}^n$*
- *$\mathcal{A}$ è autoaggiunto se e solo se A è simmetrica*
- *Poichè i prodotti scalari su $\mathbb{C}^n$ ed $\mathbb{R}^n$ coincidono su $\mathbb{R}^n$, l'operatore $\mathcal{A}$ verifica*

$$(\mathcal{A}u)v = u(\mathcal{A}v) \qquad \forall u, v \in \mathbb{R}^n$$

Dunque, per le matrici simmetriche reali valgono tutti i resultati stabiliti per gli operatori autoaggiunti, almeno se si considerano gli operatori da esse definiti su $\mathbb{C}^n$. In realtà, gli autovalori sono già garantiti reali dal lemma precedente, ma non è altrettanto chiaro che lo siano anche gli autovettori. A tale proposito, dimostriamo il seguente lemma.

Teorema 19 *Nelle stesse ipotesi del resultato precedente, la parte reale e immaginaria di un autovettore verificano l'equazione degli autovettori relativi allo stesso autovalore e dunque, ogni autospazio di $\mathcal{A}$ contiene autovettori reali.*

Dimostrazione.
Siano $\lambda \in \mathbb{C}$ e $u \in \mathbb{C}^n$ verificanti $\mathcal{A}(u) = \lambda u$. Poiché $\mathcal{A}$ è autoaggiunto, λ è reale. Siano v e w i vettori in $\mathbb{R}^n$ (e in $\mathbb{C}^n$) le componenti dei quali sono rispettivamente le parti reali e immaginarie di quelle di u. Ne segue allora $u = v + iw$, da cui $\mathcal{A}(u) = \mathcal{A}(v) + i\mathcal{A}(w)$. L'equazione degli autovettori diventa

$$Av + iAw \;=\; \mathcal{A}(v) + i\mathcal{A}(w) \;=\; \mathcal{A}(u) \;=\; \lambda u \;=\; \lambda v + i\lambda w$$

Poiché λ, v, w, Av e Aw sono reali, l'equazione precedente è equivalente a

$$Av \;=\; \lambda v \qquad Aw \;=\; \lambda w$$

Dunque, v e w sono vettori reali verificanti l'equazione degli autovettori relativi a λ. Poiché u, essendo autovettore, non è nullo, almeno uno fra v e w deve essere non nullo, ed è quindi un autovettore reale di $\mathcal{A}$ relativo a λ.

3.5 Il teorema spettrale reale

In questa sezione raccogliamo il frutto di tutte le fatiche precedenti presentando il teorema di diagonalizzazione delle matrici simmetriche reali, che ci permetterà di risolvere completamente il problema dello studio del segno delle forme quadratiche, presentato all'inizio.
Nella dimostrazione, si costruisce una base ortogonale di autovettori reali ripercorrendo, in modo di necessità alquanto tortuoso, la via seguita per il caso complesso: si utilizza il teorema di esistenza degli autovalori, che però è valido solo negli spazi complessi; si utilizza la simmetria per assicurare che ci siano autovalori e autovettori reali; ad ogni passo, si separa la parte reale dall'immaginaria dell'autovettore trovato, *a priori* complesso, per ricavarne uno reale e ortogonale ai precedenti.

Per tutto il resto della sezione, sia $A \in \mathbb{R}^{n\times n}$ verificante $a_{ij} = a_{ji}$, e sia $\mathcal{A}(u) \equiv Au \quad \forall u \in \mathbb{C}^n$.

Iniziamo con un lemma elementare, ma importante.

Teorema 20 *Sia x un vettore di $\mathbb{C}^n$ a componenti reali (cioè un vettore in $\mathbb{R}^n$, pensato immerso in $\mathbb{C}^n$!), e sia $u \in \mathbb{C}^n$. Allora, posto $u = v + iw$, $ux = 0$ in $\mathbb{C}^n$ se e solo se $vx = 0$ e $wx = 0$ in $\mathbb{R}^n$.*

Dimostrazione.
Basta osservare che $0 = ux = (v + iw)x = vx + iwx$. Poiché i prodotti scalari in $\mathbb{C}^n$ fra i vettori reali v, w e x coincidono con quelli in $\mathbb{R}^n$ e sono reali, il numero complesso all'ultimo membro è nullo se e solo se sono nulle la sua parte reale vx e quella immaginaria wx.

In sostanza, un vettore complesso è ortogonale ad uno reale se e solo se lo sono la sua parte reale e immaginaria.

Possiamo ora enunciare e provare il teorema spettrale per le matrici simmetriche.

Teorema 21 *Sia $A \in \mathbb{R}^{n\times n}$ verificante $a_{ij} = a_{ji}$.*
Allora esiste una base ortonormale di $\mathbb{R}^n$ costituita da autovettori di $\mathcal{A}(u) = Au$.

Dimostrazione.
Ambienteremo il problema in $\mathbb{C}^n$ per poter beneficiare del teorema di esistenza degli autovalori, e costruiremo la base spettrale per induzione (finita).
Per costruire il primo elemento poniamo, come abbiamo fatto più su, $\mathcal{A}(u) = Au \quad \forall u \in \mathbb{C}^n$.
Poiché $\mathcal{A} : \mathbb{C}^n \to \mathbb{C}^n$, per il teorema dei sottospazi invarianti, possiede un autovalore λ ed un autovettore $u \in \mathbb{C}^n$, ad esso relativo.
Poiché A è simmetrica, per i teoremi della sezione precedente, $\mathcal{A}$ è autoaggiunto, e quindi λ è

reale, e l'autospazio relativo a λ contiene autovettori reali. Sceltone uno ad arbitrio, e diviso per la sua norma (coincidente in $\mathbb{R}^n$ e $\mathbb{C}^n$), si ottiene un versore reale u_1, autovettore di $\mathcal{A}$ relativo a λ, che costituirà il primo elemento della base spettrale.
Supponiamo ora, per induzione, di aver già individuato $k < n$ versori reali a due a due ortogonali, autovettori di $\mathcal{A}$, e costruiamo il $(k+1)$-esimo con le stesse proprietà come segue.
Sia W il complemento ortogonale in $\mathbb{C}^n$ di $\langle u_1, u_2, ..., u_k \rangle$, ossia poniamo

$$W \ = \ \{w \in \mathbb{C}^n \, : wu_i = 0 \quad \forall i = 1..k\}$$

Fino a che $k < n$, $W \neq \{0\}$ perché, decomponendo il generico vettore $x \in \mathbb{C}^n$ rispetto a $u_1, u_2, ..., u_k$ si ottiene $x = \sum_{h=1}^{k} (xu_h)u_h + \tilde{x}$ ove, per il teorema della proiezione ortogonale, $\tilde{x}$ è ortogonale a tutti gli $u_1, u_2, ..., u_k$ e dunque appartenente a W. Se, per assurdo, questo si riducesse al solo 0 vorrebbe dire che tutto $\mathbb{C}^n$ sarebbe generato da $u_1, u_2, ..., u_k$ con $k < n$, contro il fatto che è di dimensione n.
Il sottospazio W di $\mathbb{C}^n$ è di dimensione non nulla e invariante per $\mathcal{A}$, perché $w \in W$ equivale a $wu_i = 0, \ i = 1..k$. Essendo $u_1, u_2, ..., u_k$ autovettori, per i teoremi sugli operatori autoaggiunti, da $wu_i = 0, \ i = 1..k$ segue $(Aw)u_i = 0, \ i = 1..k$, e dunque $Aw \in W$.
Per il teorema sui sottospazi (complessi) invarianti per operatori autoaggiunti, W contiene un autovettore (in generale, complesso) v, relativo ad un autovalore, distinto o eventualmente coincidente con qualcuno degli autovalori precedenti, ma in ogni caso reale, perché l'operatore è autoaggiunto. Poiché i versori $u_1, u_2, ..., u_k$ sono tutti reali, e v è ortogonale a tutti loro in quanto appartiene a W, per il lemma precedente ne segue che le sue parti reale e immaginaria sono entrambe ortogonali a tutti i versori della base già costruiti e, per l'ultimo resultato della sezione precedente, almeno una fra esse è un autovettore reale di $\mathcal{A}$, che denotiamo con $\tilde{v}$. Per completare il passo induttivo ed ottenere la tesi basta porre $u_{k+1} = \frac{1}{|\tilde{v}|}\tilde{v}$, che risulta un versore reale, ortogonale a $u_1, u_2, ..., u_k$, e autovettore di $\mathcal{A}$.
La costruzione precedente può essere iterata sino a che $k < n$, e dunque si arresta solo quando $k = n$; gli n autovettori unitari $u_1, u_2, ..., u_n$, essendo mutuamente ortogonali, sono indipendenti, e dunque formano la base richiesta di $\mathbb{R}^n$.

Capitolo 4

La diagonalizzazione in pratica

4.1 Il problema insignificante ... risolto!

Si sarà notato come il problema "insignificante" dell'introduzione abbia richiesto una fatica davvero notevole, incluso l'impiego ineludibile di numeri complessi e relativo teorema di Gauss, per la sua risoluzione. Esaminiamo la questione in pratica.

Qual'è il segno di $x^2 + xy$? È chiaro che la risposta può essere ottenuta raccogliendo la x anche non sapendo nulla di diagonalizzazione, ma non è certo così per

$$H(x, y, z, u, v) = x^2 - xy - xz - xv + y^2 - v^2 - uz$$

Il primo passo è quello di trasformare, se occorre, la forma quadratica in una simmetrica, avvalendosi della proprietà commutativa del prodotto. Se si pone $(x, y) = (u_1, u_2)$ e si pensa alla forma quadratica $x^2 + xy$ come quella generata dalla forma bilineare $\alpha(u, v) = u_1v_1 + u_1v_2$ ponendovi $u = v$, la matrice (non simmetrica) associata $\alpha_{ij} = \alpha(e_i, e_j)$ sarà

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

ma nulla vieta di scrivere il prodotto "non diagonale" xy nella forma $\frac{1}{2}xy + \frac{1}{2}yx$ e in tal caso la forma quadratica sarà identica, ma la forma bilineare diventa $u_1v_1 + \frac{1}{2}u_1v_2 + \frac{1}{2}u_2v_1$ e la matrice di rappresentazione diventa simmetrica. Infatti essa vale

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$$

Usando l'accortezza di sommare i coefficienti dei termini "non diagonali" in u_iu_j e $u_ju_i, i \neq j$, e di distribuire la somma in "parti uguali" fra i due termini in u_iu_j e u_ju_i, si ottiene comunque una matrice simmetrica, il che è una notizia davvero ottima!!! In sostanza, basta osservare che $\alpha_{ij}u_iu_j + \alpha_{ji}u_iu_j = \frac{1}{2}[\alpha_{ij} + \alpha_{ji}]u_iu_j + \frac{1}{2}[\alpha_{ij} + \alpha_{ji}]u_ju_i$, e riconoscere che ora i coefficienti dei termini "simmetrici" di indici (i, j) e (j, i) sono uguali. La teoria sviluppata sinora ci assicura *a priori* che, rispetto ad un'opportuna base di autovettori che certamente esiste, la forma quadratica assumerà la forma diagonale con gli autovalori tutti reali della matrice originale sulla diagonale. Non occorre fare altro che scrivere l'equazione caratteristica della matrice e, nei casi fortunati, determinarne gli autovalori ed esaminarne il segno. Anche nei casi sfortunati nei quali non fosse possibile (o agevole) calcolare "a mano" gli autovalori risolvendo l'equazione caratteristica, è comunque facile stabilire se essi siano concordi o no, il che permette comunque di completare lo studio del segno, come vedremo nella prossima sezione.

4.2 Il segno delle forme quadratiche

Lo studio dei massimi e dei minimi per le funzioni di più variabili, nel solco dei risultati sul segno della derivata seconda già noti per le funzioni di una sola variabile, conduce a studiare il segno di una forma quadratica, detta forma hessiana, i coefficienti della quale sono abbastanza facili da determinare, perché richiedono solo il calcolo delle derivate seconde. Un importante teorema di Schwarz assicura miracolosamente che la forma è simmetrica, almeno quando le derivate seconde siano continue, ipotesi ragionevole e di facile verifica.
La formula di Taylor, infatti, consente di provare che il segno della differenza $f(x)-f(x_0)$ coincide con quello della forma hessiana, almeno quando esso è definito (cioè se essa non è nulla) e x è abbastanza vicino a x_0. Il problema è dunque identico a quello iniziale: "Che segno ha la forma hessiana"? Sappiamo già che occorre stabilire se gli autovalori della matrice delle derivate seconde nel punto siano o no concordi.
Il calcolo esplicito delle radici del polinomio caratteristico di tale matrice è facile per i polinomi di secondo grado, che si incontrano nello studio delle forme quadratiche in due variabili indipendenti, discretamente macchinoso per i polinomi di terzo grado (formula di Cardano), che si incontrano in presenza di tre variabili indipendenti, intollerabilmente macchinoso per i polinomi di quarto grado ("formula" di Ludovico Ferrari), presenti in quattro variabili indipendenti, e impossibile in generale dal quinto grado in su (teorema di Galois), quando si trattano cinque o più variabili indipendenti.
La buona notizia fra tante sciagure è che il calcolo esplicito delle radici non è necessario per stabilirne il segno, soprattutto se non interessa il segno di ogni radice, ma qualcosa di più grossolano: il segno della forma. Il teorema sui punti stazionari, infatti, suona approssimativamente così:

- Se gli autovalori dell'hessiana sono tutti strettamente positivi, il punto è di minimo.
- Se sono tutti strettamente negativi, il punto è di massimo.
- Se ci sono due autovalori non nulli e discordi, il punto è di sella.

che corrisponde a quanto già visto: la matrice è diagonalizzabile sui reali in quanto simmetrica; se gli autovalori sono tutti non nulli e concordi allora la forma è sempre non nulla fuori dall'origine e concorde ad essi (i primi due casi), mentre se esistono due autovalori di segno opposto allora la forma cambia segno: nella direzione di uno degli autovettori si sale perché il segno di $f(x)-f(x_0)$ è positivo e nell'altra si scende. Basta immaginare una sella, o un passo in montagna.
Notiamo che i tre casi precedenti non esauriscono tutti i casi possibili: resta fuori il caso delle forme l'equazione caratteristica delle quali ha la radice nulla, mentre tutte le altre sono concordi. In questo caso non si può adoperare il terzo criterio, perché non ci sono autovalori discordi, e neanche i primi due, perché gli autovalori non sono tutti non nulli. Discuteremo più avanti tale caso con degli esempi.
A dire il vero, i resultati della prossima sezione permetteranno comunque di completare lo studio del segno delle forme quadratiche, e quindi anche dell'hessiana, ma non di determinare il carattere dei punti critici con hessiana dotata di autovalori nulli e con autovalori non nulli concordi, caso corrispondente esattamente a quello infausto nel quale $f''(x_0)=0$ per le funzioni di una variabile: in entrambi i casi, il segno della differenza $f(x)-f(x_0)$ non è determinato solo da quello della f'' o della forma hessiana, coinvolge i termini di ordine superiore, e richiede indagini *ad hoc* che esulano dall'ambito di queste note. Ad ogni modo, i criteri seguenti permettono di identificare tali situazioni particolarmente "ostiche".

4.2.1 Il resultato fondamentale

Il resultato alla base di ciò che segue è:

Teorema 22 *Sia p un polinomio reale di grado n, avente n radici reali. Allora:*

- *Condizione necessaria e sufficiente perché le sue radici siano tutte strettamente negative è che i suoi coefficienti siano tutti non nulli e concordi.*

- *Condizione necessaria e sufficiente perché le sue radici siano tutte strettamente positive è che i suoi coefficienti siano tutti non nulli e a segni alterni.*

Questo resultato generalizza al grado n la nota regola dei segni di Cartesio per le equazioni di secondo grado.

Osserviamo subito che sappiamo già che tutte le radici del polinomio sono reali, ed è inoltre immediato riconoscere se il polinomio abbia o no la radice nulla: basta guardare il termine noto. Si può dunque decidere subito, con una semplice disamina dei coefficienti, se il punto sia di massimo (coefficienti tutti diversi da zero e concordi), di minimo (coefficienti tutti diversi da zero e a segni alterni), di sella (tutti gli altri casi purché con termine noto diverso da zero) oppure se si è in presenza del caso critico, nel quale nulla si può dire *a priori* (termine noto uguale a zero). Anche in tal caso, come si vedrà presto, i resultati precedenti possono ancora venire reimpiegati per decidere se le radici non nulle siano concordi (ed è il vero caso critico), oppure no, nel qual caso il punto sarà ancora di sella (degenere). Di tutto ciò si dirà meglio dopo la dimostrazione.

Dimostrazione.
Si inizierà col provare la condizione sufficiente del primo resultato: "Se i coefficienti sono non nulli e concordi, le radici sono strettamente negative".
Infatti, si consideri un polinomio

$$p(x) = \sum_{k=0}^{n} a_k x^k$$

a coefficienti strettamente positivi $a_k > 0 \quad \forall k = 0..n$.
Se $\overline{x} \geq 0$ ne segue subito $a_k \overline{x}^k \geq 0$ da cui $p(\overline{x}) \geq a_0 > 0$. Dunque nessun valore maggiore o uguale a zero può essere una radice. Se poi i coefficienti fossero tutti negativi, basta osservare che il polinomio $-p$ ha tutti i coefficienti positivi e le stesse radici di p.

Verrà provata ora la condizione necessaria: "Se le n radici sono strettamente negative, i coefficienti sono non nulli e concordi".
Si consideri il polinomio $\tilde{p}(x) = \frac{1}{a_n} p(x)$, che ha le stesse radici di p. Dal teorema di Ruffini, per induzione, si può provare che, se $x_1, x_2, ..., x_n$ sono le radici di p, (e di $\tilde{p}$), allora risulta

$$\tilde{p}(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$$

Eseguiamo ora il prodotto, sommando tutti i possibili prodotti ottenuti scegliendo in ogni fattore uno degli addendi, e ordiniamoli per potenze decrescenti di x. C'è un solo termine di grado n, ottenuto scegliendo in ogni fattore l'addendo x, e vale x^n. Ci sono n termini di grado $n-1$, ottenuti scegliendo in un solo fattore una delle costanti $-x_i$, $i = 1..n$, e prendendo la x in tutti gli altri; raccogliendo x^{n-1} si ottiene $x^{n-1} \sum_{k=1}^{n} (-x_k)$. Analogamente, per ottenere un termine in x^{n-k} occorre scegliere in k fattori, di indici $1 \leq i_1 < i_2 < \ldots i_k \leq n$, i termini noti $(-x_{i_1})$ $\ldots (-x_{i_k})$, e negli altri l'indeterminata x. Raccogliendo come sopra e ripetendo per tutti i gradi da 0 a n, si ottiene

$$\tilde{p}(x) = x^n \ + \ x^{n-1} \sum_{k=1}^{n} (-x_k) \ + \ \cdots$$

$$\cdots \ + \ x^{n-k} \sum_{1 \leq i_1 < i_2 < \ldots < i_k \leq n} [(-x_{i_1}) \cdots (-x_{i_k})] \quad + \ \cdots\cdots \ + [(-x_1) \cdots (-x_n)]$$

Dal principio di identità dei polinomi, che asserisce che due polinomi sono identici se e solo se hanno gli stessi coefficienti, i coefficienti delle potenze così calcolati sono proprio quelli di $\tilde{p}$ e, se le radici sono strettamente negative, essi sono tutti strettamente positivi. Poiché $p = a_n\tilde{p}$ ne segue subito che i coefficienti di p hanno tutti il segno di a_n e sono dunque concordi.

Il grosso della fatica è compiuto! Per provare l'altro resultato sulle radici positive, basta osservare che se $p(x)$ ha radici tutte strettamente positive allora $q(x) \equiv p(-x)$ ha radici tutte strettamente negative. Ne segue dal resultato precedente che $q(x) = \sum_{k=0}^{n} b_k x^k$ ha i coefficienti non nulli e concordi. Si ha allora

$$p(x) = q(-x) = \sum_{k=0}^{n} b_k(-1)^k x^k$$

e, per il già ricordato principio di identità dei polinomi, i coefficienti di p saranno $(-1)^k b_k$. Essendo i b_k non nulli e concordi, essi saranno allora non nulli e a segni alterni.

4.2.2 Il caso della radice nulla

In questa sezione verranno stabilite le condizioni sui coefficienti che assicurino che tutte le radici non nulle siano concordi. È stato già notato che è proprio il caso più difficile nello studio dei punti estremi. I resultati già dimostrati permettono di riconoscere facilmente tali casi.
Sia $p(x)$ tale che, accanto alla radice 0, abbia le altre radici tutte concordi. In tal caso, detto k il grado minimo dei termini non nulli di p e raccolto x^k si ottiene $p(x) = x^k\tilde{p}(x)$, ove $\tilde{p}$ non si annulla più in zero. Le radici di p sono quelle di $\tilde{p}$, con in più lo zero, e dunque $\tilde{p}$ ha tutte le radici non nulle e concordi. Per i teoremi precedenti i suoi coefficienti sono non nulli: concordi, se le radici sono negative, a segni alterni se positive. Per concludere, basta allora osservare che i coefficienti di $\tilde{p}$ sono quelli di p, anche se spostati di k posti, e dunque:

Teorema 23 *Nelle stesse ipotesi iniziali:*

- *Condizione necessaria e sufficiente affinché un polinomio avente* 0 *come radice di molteplicità k abbia tutte le altre radici negative è che i suoi coefficienti di grado maggiore o eguale a k siano non nulli e concordi.*

- *Condizione necessaria e sufficiente affinché un polinomio avente* 0 *come radice di molteplicità k abbia tutte le altre radici positive è che i suoi coefficienti di grado maggiore o eguale a k siano non nulli e a segni alterni.*

In tutti gli altri casi — termini mancanti o termini di segno diverso da quanto precisato nei teoremi precedenti — visto che le radici sono tutte reali per ipotesi, vuol dire che ci saranno radici discordi oltre a quella nulla, e ciò consente comunque di concludere che la forma cambia segno e, parlando di hessiane e punti critici, di identificare come punto di sella (degenere) il punto stazionario in esame.
In definitiva:

Teorema 24 *Data una forma quadratica simmetrica $\alpha(x) = \sum_{i,j=1}^{n} \alpha_{ij}x_ix_j$ e detto $p(\lambda)$ il polinomio caratteristico della matrice $A = (\alpha_{ij})$, allora*

- $$\alpha(x) \;<\; 0 \quad \forall x \neq 0$$

  se e solo se i coefficienti di p sono tutti non nulli e concordi.

  La forma viene in tal caso detta definita negativa*.*

- $$\alpha(x) \;>\; 0 \quad \forall x \neq 0$$
  se e solo se i coefficienti di p sono tutti non nulli e a segni alterni.

  La forma viene in tal caso detta definita positiva.

- $$\alpha(x) \;\leq\; 0 \quad \forall x \in \mathbb{R}^n$$
  se e solo se esiste $k \geq 0$ tale che i coefficienti di p di grado maggiore o uguale a k sono tutti non nulli e concordi, mentre quelli di grado minore sono nulli.

  Se $k > 0$, la forma viene detta semidefinita negativa*, e assume valore nullo anche in punti diversi dallo zero.*

- $$\alpha(x) \;\geq\; 0 \quad \forall x \in \mathbb{R}^n$$
  se e solo se esiste $k \geq 0$ tale che i coefficienti di p di grado maggiore o uguale a k sono tutti non nulli e a segni alterni, mentre quelli di grado minore sono nulli.

  Se $k > 0$, la forma viene detta semidefinita positiva*, e assume valore nullo anche fuori dallo zero.*

- *In ogni altro caso, esistono punti di $\mathbb{R}^n$ sui quali $\alpha(x)$ assume valori non nulli e discordi.*

  La forma viene in tal caso detta indefinita*, ed esistono vettori non nulli sui quali la forma si annulla (vedi il teorema iniziale di queste note).*

Esistono criteri più raffinati per studiare il segno degli autovalori: il teorema di Sylvester, ad esempio, consente abbastanza agevolmente di contare il numero di autovalori positivi e negativi, pur senza determinarli esplicitamente.
I criteri qui presentati, di certo più grossolani, consentono comunque di ricavare con molto maggiore facilità tutte le informazioni necessarie allo studio del segno delle forme e dei punti stazionari: il caso "critico" non è tale per difetto dei criteri esposti, ma per la natura delle cose. Infatti, $x^2 + y^4$ e $x^2 + y^3$ hanno entrambe hessiana identica ($\alpha(x, y) = x^2$) con autovalore 0, uguale direzione dell'autospazio (l'asse y), ma si vede subito che la prima ha minimo in zero e l'altra no. Così come accade in una variabile quando $f'' = 0$, in tal caso diventa rilevante il segno dei termini di ordine superiore a 2, e mentre y^4 non cambia segno, y^3 lo fa. È inutile dire che, purtroppo, lo studio generale del segno dei termini di ordine maggiore al secondo, immediato per i termini di ordine dispari che cambiano tutti segno se si cambia di segno il punto considerato, è invece di difficiltà proibitiva per i termini di ordine pari!!! Dunque, in presenza di un punto critico degenere (autovalore nullo) con autovalori non nulli concordi va presa in seria considerazione la possibilità di prendere un computer e di campionare la funzione in vicinanza del punto critico, soprattutto nelle direzioni dei vettori del nucleo della matrice, che è anche l'autospazio corrispondente all'autovalore 0, sperando che il passo adottato sia piccolo abbastanza da fornire risposte affidabili o, se è agevole, stimando il gradiente e scegliendo il passo di conseguenza. Un'altra via meno brutale è quella di studiare la restrizione della funzione all'iperpiano passante per il punto critico e parallelo all'autospazio relativo all'autovalore nullo. Ciò, almeno se l'autospazio è di dimensione uno, potrebbe risultare

di difficoltà abbordabile in quanto richiede uno studio di funzione in una variabile che ... ma ciò davvero non c'entra più nulla con la diagonalizzazione!!!

4.3 Riduzione a forma canonica delle quadriche

Una delle più interessanti applicazioni della teoria della diagonalizzazione è la classificazione delle superficie algebriche di secondo grado. Si consideri l'equazione

$$\sum_{i,j=1}^{n} \alpha_{ij} x_i x_j \ + \ \sum_{h=1}^{n} \beta_h x_h \ + \ \gamma \ = \ 0$$

Si definisce *quadrica*, o superficie algebrica di secondo grado, il luogo dei punti di $\mathbb{R}^n$ le coordinate dei quali verificano l'equazione. Per $n = 2$ i luoghi geometrici delle soluzioni di tali equazioni rappresentano, per opportune scelte dei coefficienti, le curve più antiche della storia:

$$x^2 + y^2 = 1, \quad x^2 + 4y^2 = 1, \quad y = x^2, \quad x^2 - y^2 = 1$$

Sono le *coniche*: circonferenze, ellissi, parabole, iperboli, in origine definite mediante sezioni di un cono con un piano (donde il nome) e come tali studiate da Archimede ed Apollonio, ma a partire da Fermat studiate anche col "nuovo" metodo analitico, come luoghi di soluzioni di equazioni algebriche.
Cosa accade se si considerano tre o più variabili? Rispondiamo con una domanda:"Cosa accade se la matrice (α_{ij}) è diagonale"? Il motivo della domanda è evidente: per ragioni identiche a quelle presentate ovunque sino ad ora, la matrice può essere riscritta come simmetrica e dunque può essere diagonalizzata con un opportuno cambio di base.
L'equazione (e la superficie) diventa

$$\sum_{i=1}^{n} (\lambda_i \tilde{x_i}^2 \ + \ \tilde{\beta}_i \tilde{x_i}) \ + \ \tilde{\gamma} \ = \ 0$$

espressa nelle nuove variabili $\tilde{x_i}$, componenti del vettore $x \in \mathbb{R}^n$ rispetto alla nuova base di autovettori che diagonalizza la matrice (α_{ij}).
Si può allora semplificarne l'espressione con la stessa tecnica usata per risolvere le equazioni di secondo grado: basta pensare ai termini contenenti quadrati e a quelli di primo grado come a quadrati e doppi prodotti in opportuni quadrati di binomi, sommando e sottraendo i termini noti necessari a completarli.
Basta dunque utilizzare, per ogni $i = 1..n$ per cui $\lambda_i \neq 0$, l'identità

$$ax^2 + bx \ = \ a \left[x^2 \ + \ \frac{b}{a} x \right] \ = \ a \left[(x + \frac{b}{2a})^2 - \frac{b^2}{4a^2} \right]$$

per ottenere

$$\sum_{i:\lambda_i \neq 0} \lambda_i (\tilde{x}_i - X_i)^2 \ + \ \sum_{i:\lambda_i = 0} \tilde{\beta}_i \tilde{x}_i \ + \ \Gamma \ = \ 0$$

ove si è posto $X_i = -\frac{\tilde{\beta}_i}{2\lambda_i}$ e si è indicata con Γ la somma di tutte le costanti che risultano da tutte le operazioni. Notiamo espressamente che le variabili che a primo membro appaiono nei termini di primo grado sono esattamente tutte quelle corrispondenti ad autovalori nulli nella forma quadratica dei termini di secondo grado, che non vengono "inglobate" in quadrati di binomi.
Per semplificarla ulteriormente, ed ottenere la cosiddetta *forma canonica*, immaginiamo per prima cosa di cambiare di nome le variabili, indicando con x_i quelle nuove, invece di quelle vecchie; poi ricambiamoli in modo da avere i quadrati di binomi ai primi posti, poi quelli di primo grado, e infine il termine noto. L'equazione diventa del tipo:

$$\sum_{i=1}^{k} \lambda_i (x_i - X_i)^2 \ + \ \sum_{j=k+1}^{m} b_j x_j \ + \ \Gamma \ = \ 0$$

dove x_i sono le componenti del punto rispetto alla base riordinata $u_1, u_2, ..., u_n$, $\lambda_1, \lambda_2, ..., \lambda_k$ sono i k autovalori non nulli, spostati ai primi posti ed eventualmente coincidenti fra loro, $b_{k+1}, \ldots, b_m$ sono i coefficienti dei termini di primo grado non nulli, spostati subito dopo quelli di secondo grado, e dove le variabili $x_{m+1}, \ldots, x_n$ non appaiono esplicitamente nell'equazione.
Un'ultima importante semplificazione (non necessaria se nella forma precedente non appaiono termini di primo grado perché tutti inglobati nei quadrati, cioè se gli autovalori sono tutti non nulli) si può realizzare con un ulteriore cambio di base, lasciando inalterati i primi k vettori, le componenti rispetto ai quali appaiono nei termini quadratici, e gli ultimi $n-m$, corrispondenti alle variabili non presenti nell'equazione, e cioè

$$u'_i \ = \ u_i \qquad \forall i = 1..k \quad o \quad i = m+1..n$$

rimpiazzando invece u_{k+1} con il versore

$$u'_{k+1} = \left| \sum_{j=k+1}^{m} b_j u_j \right|^{-1} \sum_{j=k+1}^{m} b_j u_j \ = \ \left[\sqrt{\sum_{k+1}^{m} b_j^2} \right]^{-1} \sum_{j=k+1}^{m} b_j u_j$$

(ricordando il teorema di Pitagora e la mutua ortogonalità di $u_{k+1}, .., u_m$) ancora ortogonale a $u_1, u_2, ..., u_k$ in quanto combinazione di vettori ad essi ortogonali, e completando

$$\{u'_1, u'_2, ..., u'_k, u'_{k+1}, u_{m+1}, .., u_n\}$$

ad una base ortonormale a $\mathbb{R}^n$, usando prima il teorema del completamento ad una base e poi il procedimento di Gram-Schmidt per renderla ortonormale. L'utilità della manovra si ravvisa immediatamente osservando che la nuova componente x'_{k+1} del generico vettore x vale

$$x'_{k+1} \ = \ x\, u'_{k+1} \ = \ x \left[\sqrt{\sum_{k+1}^{m} b_j^2} \right]^{-1} \sum_{j=k+1}^{m} b_j u_j \ =$$

$$= \left[\sqrt{\sum_{k+1}^{m} b_j^2} \right]^{-1} \sum_{j=k+1}^{m} b_j x\, u_j \ = \ \left[\sqrt{\sum_{k+1}^{m} b_j^2} \right]^{-1} \sum_{j=k+1}^{m} b_j x_j$$

e dunque l'equazione diventa

$$\sum_{i=1}^{k} \lambda_i (x'_i - X'_i)^2 \ + \ \left[\sqrt{\sum_{k+1}^{m} b_j^2} \right] x'_{k+1} \ + \ \Gamma' \ = \ 0$$

La forma così ottenuta è detta *forma canonica*. A seconda che $\sum_{k+1}^{m} b_j^2$ e Γ siano 0 o meno, del numero k di autovalori non nulli, del loro segno in relazione a quello di Γ, si ottengono diversi tipi di quadriche, talvolta "spezzate" in iperpiani (come, ad esempio, $x^2 = y^2$ in $\mathbb{R}^3$), o anche nessuna soluzione (come in $x^2 = -1$).
Nella classificazione, può risultare decisivo l'impiego del teorema di Sylvester, che permette di determinare con ragionevole rapidità e senza calcolarli, il numero degli autovalori positivi e negativi. Se però si debbano determinare gli assi di simmetria, non resta che calcolare esplicitamente gli autovalori e gli autovettori (se possibile!), o approssimarli.
Nel caso $n = 3$ (senza alcuna giustificazione geometrica e ignorando le quadriche spezzate, quelle ridotte ad un punto o senza punti reali, per mancanza di spazio) per $n = m = k$ si ottengono coni per $\Gamma = 0$, ellissoidi (per autovalori tutti concordi fra loro e discordi da Γ, iperboloidi per autovalori non nulli e discordi, ad una falda (ellittici) o a due falde (iperbolici) a seconda del numero di autovalori concordi con Γ; paraboloidi se $m = n$ ma $k = n - 1$, ellittici o iperbolici a seconda se gli autovalori non nulli siano concordi o meno; cilindri se $m < n$, ellittici o iperbolici

se $k = m$, parabolici se $k < m$.
Con atteggiamento di ossequio, scriviamo l'equazione della sfera unitaria

$$x^2 + y^2 + z^2 \ = \ 1$$

che è un ellissoide con tutti gli autovalori non nulli e uguali fra loro.
Per avere qualche appiglio geometrico esaminiamo in qualche dettaglio il caso più noto, corrispondente ad $n = 2$.

4.3.1 Le coniche ... di corsa!

Se ci sono autovalori nulli (qualcuno o tutti i λ_i sono nulli), si ha:

- Se $\lambda_1 = \lambda_2 = 0$, non ci sono termini di secondo grado e dunque il luogo delle soluzioni è una retta se almeno uno dei coefficienti di primo grado è non nullo. Se anche i termini di primo grado sono tutti nulli, il luogo non ha punti se $\Gamma \neq 0$ ed è costituito da tutto lo spazio se $\Gamma = 0$ (in tal caso l'equazione del luogo è $0 = 0$, sempre verificata).

- Se $\lambda_1 = 0$ e $\lambda_2 \neq 0$ l'equazione diventa del tipo $ax + by^2 + \Gamma = 0$ che, se $a \neq 0$, è una *parabola* (canonica) ad asse orizzontale, traslata in direzione dell'asse x se $\Gamma \neq 0$. Se invece il termine di primo grado è nullo, l'equazione diventa del tipo $by^2 + \Gamma = 0$ che non ha soluzioni se a e Γ sono non nulli e concordi mentre, nella terminologia tradizionale, *si spezza* nelle due rette $y = \pm\sqrt{-\Gamma/b}$ se sono discordi. Le due rette coincidono se $\Gamma = 0$.
Analogamente si tratta il caso in cui l'unico autovalore non nullo è λ_1. In questo caso i ruoli degli assi x e y si scambiano.

Se entrambi gli autovalori sono invece non nulli e quindi non sono presenti termini di primo grado nella forma canonica, allora:

- Se λ_1 e λ_2 sono concordi e anche Γ lo è con essi, non ci sono soluzioni (reali). Se invece è discorde si ottengono le *ellissi* (ad esempio, $2x^2 + 3y^2 - 2 = 0$). Se poi $\Gamma = 0$, allora l'unico punto del luogo è l'origine.

- Se λ_1 e λ_2 sono discordi e $\Gamma \neq 0$ si ottengono le *iperboli* (ad esempio, $2x^2 - 3y^2 - 2 = 0$). Se invece $\Gamma = 0$ allora l'equazione diventa $\lambda_1 x^2 = -\lambda_2 y^2$, ovvero

$$x = \pm\sqrt{\frac{-\lambda_2}{\lambda_1}}y$$

che è una "conica" *spezzata* nelle due rette $x = \sqrt{\frac{-\lambda_2}{\lambda_1}}y$ e $x = -\sqrt{\frac{-\lambda_2}{\lambda_1}}y$.

4.4 Possibili sviluppi

È evidente l'importanza di porre in luce il significato geometrico degli autospazi (legati agli assi di simmetria) e dei valori X_i (legati alle coordinate del vertice o del centro). Sarebbe di enorme interesse anche studiare per esteso il caso delle *quadriche* in tre variabili... senza contare le applicazioni alla meccanica del corpo rigido (ellissoide e assi principali d'inerzia), e alla meccanica dei continui (tensore degli sforzi e sforzi principali)... limitandosi al minimo sindacale!
Ciò, purtroppo, eccederebbe largamente lo spazio a nostra disposizione. I riferimenti d'obbligo sono i libri di geometria analitica, alcuni dei quali scritti da Maestri della Geometria Algebrica italiana come, ad esempio, Castelnuovo. Volendo approfondire gli aspetti geometrici della teoria classica delle quadriche in $\mathbb{R}^3$ (e della geometria analitica classica, in generale), un titolo concepito appositamente per l'uso didattico (e sul quale ha studiato l'autore) è quello di Giovanni Dantoni, che però non fa uso di diagonalizzazione, ma presenta l'approccio classico al problema basato sugli invarianti.

Per approfondimenti sugli aspetti algebrici (il teorema di Sylvester, ad esempio, o i teoremi di diagonalizzazione delle matrici unitarie) si raccomanda il libro di Algebra Lineare di Serge Lang, sconsigliando invece fortemente l'altro volume "Algebra", nonostante la sua ricca (e astrattissima) sezione di algebra lineare.
La teoria dei punti *stazionari* o *critici* (massimi, minimi, selle...) in più variabili è reperibile in tutti i secondi volumi dei corsi (di una volta) di Analisi Matematica. Un'ottima scelta, fra le tante possibili, è quello di Enrico Giusti.
L'ellissoide d'inerzia si trova su tutti i libri di Meccanica Razionale. Il tensore degli sforzi (in varie forme) su tutti quelli di teoria dell'elasticità e su qualcuno di dinamica dei fluidi ... è roba che serve!