

# Lezioni di Calcolo Numerico

## Lezione 10: Autovalori e valori singolari di matrici

Alberto Tibaldi

8 giugno 2018

### Indice

<b>1</b>	<b>Concetti fondamentali sugli autovalori</b>	<b>2</b>
1.1	Potenze e esponenziale di una matrice . . . . .	3
1.2	Quoziente di Rayleigh . . . . .	4
1.3	Raggio spettrale e norma spettrale di una matrice . . . . .	6
1.4	Cerchi di Gershgorin . . . . .	7
1.5	Matrici simili . . . . .	8
1.6	Condizionamento del calcolo degli autovalori . . . . .	9
<b>A</b>	<b>Un programmino per disegnare i cerchi di Gershgorin</b>	<b>11</b>

### Accenni sulla storia degli autovalori/autovettori

L'obiettivo di questo capitolo sarà, dopo aver *rinfrascato* i principali concetti legati agli autovalori e agli autovettori, introdurre alcune tecniche numeriche atte a calcolarli. Vorrei tuttavia approfittare di questo spazio di introduzione per accennare la storia di queste idee, suggerendo ai lettori interessati di leggere ulteriori dettagli in [1].

Andando molto indietro nel tempo, Cartesio<sup>1</sup> nel 1637 propose di trasformare una generica forma quadratica<sup>2</sup>  $ax^2 + 2bxy + cy^2$  nella sua forma normale  $\alpha x^2 + \beta y^2$  grazie all'identificazione dei suoi *assi principali*. Passando per i grandi Eulero, Lagrange, Jacobi e Cauchy, si è arrivati a Sylvester e Cayley, che hanno applicato il principio degli *assi principali* al calcolo matriciale, sviluppato in quegli anni (si parla degli anni '50 del 1800). Qui viene un po' fuori il desiderio di sfruttare questa teoria per cercare di *sostituire le matrici con dei numeri*, a patto di saper rappresentare i vettori mediante questi *assi principali*.

Questa idea di sostituire cose complicate con dei numeri era già stata applicata qualche anno prima. Durante la rivoluzione industriale la termodinamica andava di moda, e uno dei matematici più *eroici* in questo contesto era stato Fourier che, con un trucco oggi noto come **trasformata di Fourier**, aveva trovato un modo di *far sparire* le derivate dalle equazioni differenziali, trasformandole in equazioni algebriche. Questa cosa a me ricorda un po' il far sparire una matrice per sostituirle un numero.

Finora non ho nominato *autovalori* o *autovettori*, ricorrendo ai nomi utilizzati dagli antichi; questo, perché questi nomi sono abbastanza recenti. Autovalore e autovettore sono traduzioni non proprio fedelissime di **eigenvalue** e **eigenvector**. Tuttavia, anche chi possiede una certa dimestichezza con l'inglese, potrebbe non avere familiarità col prefisso *eigen-*. In effetti, i termini originali sono stati **Eigenwert** e **Eigenvektor**, e sono tedeschi. Infatti, nella Germania nazista c'era molto interesse verso questi concetti, perché su di essi si basavano sia la teoria dei Campi Elettromagnetici, approfondita al fine di produrre dei RADAR coi quali rilevare la presenza di nemici, sia la Meccanica Quantistica, utile per arrivare in qualche modo alla bomba atomica.

<sup>1</sup>che dunque non ha solo sfruttato le coordinate geometriche precedentemente inventate da Nicola d'Oresme per legare Algebra e Geometria e capito che *cogito, ergo sum* ;-)

<sup>2</sup>dunque orientata arbitrariamente in un piano nel caso bidimensionale, o nello spazio nel caso tridimensionale

Oggi, gli autovalori e autovettori trovano applicazione in moltissimi campi: dalla Meccanica delle Vibrazioni, all'Automatica, all'Elettronica, all'Elettromagnetismo, alla appunto Meccanica Quantistica, e chi più ne ha più ne metta.

## 1 Concetti fondamentali sugli autovalori

Dopo la parentesi della scorsa sezione sulle matrici rettangolari torniamo a parlare di matrici quadrate  $\underline{A} \in \mathbb{R}^{n,n}$  e, in questo contesto, introduciamo una classe di vettori un po' *mafiosi*, che insomma si sentono al di sopra delle regole dei normali *cittadini degli spazi vettoriali*. In generale, quando applichiamo una matrice a un vettore colonna moltiplicandola ad esso da sinistra, il vettore viene trasformato in un altro vettore che, volendo dare un'interpretazione geometrica<sup>3</sup>, ha direzione e lunghezza diverse da quelle di partenza. Cosa fanno di così strano questi vettori? Beh, proviamolo su un esempio: proviamo a moltiplicare da sinistra la matrice

$$\underline{A} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix}$$

al vettore

$$\underline{x} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Vediamo cosa succede calcolando  $\underline{A}\underline{x}$ :

$$\underline{A}\underline{x} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \times 3 + 3 \times 2 \\ 2 \times 3 + 1 \times 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 8 \end{bmatrix} = 4 \times \begin{bmatrix} 3 \\ 2 \end{bmatrix}.$$

Ohibò. Moltiplicando  $\underline{A}$  per  $\underline{x}$ , viene fuori un vettore che ha la stessa direzione di  $\underline{x}$ , ma che è moltiplicato per un numero, 4. Quindi, il vettore non ha subito alcuna rotazione nello spazio, ma solo un *allungamento* di 4 volte. A questo punto, potremmo chiederci:

«Ma siamo sicuri che il *mafioso* in questione sia il vettore, e non la matrice?»

Beh, abbastanza: prendiamo la stessa matrice e moltiplichiamola per un altro vettore  $\underline{v}$ :

$$\underline{A}\underline{v} = \begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \times 2 + 3 \times 3 \\ 2 \times 2 + 1 \times 3 \end{bmatrix} = \begin{bmatrix} 13 \\ 7 \end{bmatrix},$$

che non è multiplo di  $\underline{v}$ : possiamo osservare sia un allungamento, sia una rotazione! Quindi, per questa matrice  $\underline{A}$ ,  $\underline{x}$  soddisfa questa strana relazione, ma  $\underline{v}$  non lo fa. Quando vale una relazione nella forma

$$\underline{A}\underline{x} = \lambda \underline{x}, \tag{1}$$

dove  $\lambda$  è un numero, allora  $\underline{x}$  viene detto **autovettore** di  $\underline{A}$ , e  $\lambda$  viene detto **autovalore relativo a esso**. Questa relazione si può scrivere, portando tutti i termini a sinistra e raccogliendo, come il seguente sistema lineare omogeneo:

$$(\underline{A} - \lambda \underline{I})\underline{x} = \underline{0}. \tag{2}$$

Dunque,  $\underline{x}$  si può anche interpretare come una soluzione **non banale** di questo sistema<sup>4</sup>. Per il teorema di Rouché-Capelli, perché esista una soluzione non banale del sistema, si richiede che  $\underline{x}$  appartenga allo spazio nullo della matrice  $(\underline{A} - \lambda \underline{I})$ . Perché questa abbia uno spazio nullo, il suo determinante deve essere pari a 0. In effetti, in Algebra Lineare, al fine di trovare un autovalore, uno deve imporre la condizione

$$\det\{\underline{A} - \lambda \underline{I}\} = 0$$

<sup>3</sup>al solito, valida per i casi a due e/o a tre dimensioni

<sup>4</sup>infatti nell'esempio appena visto,  $\underline{x}$  non era il vettore contenente solo zeri ;-)

e questo darebbe luogo all'equazione caratteristica, che poi deve essere risolta. L'equazione caratteristica è un'equazione algebrica avente come incognite i  $\lambda$ , ovvero gli autovalori della matrice. Dal momento che si può vedere che i  $\lambda$  sono sulla diagonale della matrice, e nascendo questa equazione da un determinante<sup>5</sup>, l'equazione caratteristica avrà grado pari a  $n$ , dove  $n$  è il numero di righe/colonne della matrice. Il teorema fondamentale dell'Algebra garantisce che l'equazione caratteristica ha  $n$  soluzioni, che possono essere reali o complesse; in presenza di un autovalore complesso, anche il suo complesso coniugato sarà un autovalore.

Data una matrice  $A$ , MATLAB<sup>®</sup> è in grado di calcolarne autovalori e autovettori, mediante il comando `eig` (da *eigenvalue*, come spiegato prima). In particolare,

- se si utilizza il comando `eig` chiedendo un solo argomento di uscita, quindi nella forma

$$E = \text{eig}(A);$$

la variabile  $E$  sarà un **vettore**, contenente tutti gli autovalori della matrice  $A$ ;

- se si utilizza il comando `eig` chiedendo due argomenti in uscita, quindi nella forma

$$[V, E] = \text{eig}(A);$$

le variabili  $V$  e  $E$  saranno **due matrici** aventi la stessa dimensione di  $A$ ; in particolare,  $V$  sarà la matrice avente, come ciascuna colonna, l'autovettore di  $A$  associato all'autovalore contenuto nella corrispondente posizione sulla diagonale di  $E$ ; quest'ultima, è una matrice diagonale.

## 1.1 Potenze e esponenziale di una matrice

Dato  $\lambda_i$  il  $i$ -esimo autovalore della matrice  $\underline{A}$ ,  $\underline{x}_i$  sarà l'autovettore a esso associato. Una matrice  $\underline{A} \in \mathbb{R}^{n,n}$  ha quindi  $n$  autovalori (non necessariamente distinti tra loro),  $\{\lambda_i\}$ , ciascuno dei quali è associato a un autovettore. Volendo, possiamo definire una matrice  $\underline{X}$ , le cui colonne sono i vari autovettori  $\{\underline{x}_i\}$ :

$$\underline{X} = [\underline{x}_1 \quad \underline{x}_2 \quad \underline{x}_3 \quad \dots \quad \underline{x}_n].$$

Immaginiamo a questo punto di moltiplicare da sinistra la matrice  $\underline{A}$  per ciascuno di questi vettori; otterremo, in virtù della relazione (1) applicata a ciascuno degli autovettori, che

$$\underline{A}\underline{X} = [\underline{A}\underline{x}_1 \quad \underline{A}\underline{x}_2 \quad \underline{A}\underline{x}_3 \quad \dots \quad \underline{A}\underline{x}_n] = [\lambda_1 \underline{x}_1 \quad \lambda_2 \underline{x}_2 \quad \lambda_3 \underline{x}_3 \quad \dots \quad \lambda_n \underline{x}_n].$$

A questo punto è utile definire la matrice diagonale  $\underline{D}$

$$\underline{D} = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \lambda_n \end{bmatrix},$$

avente come ciascun elemento della diagonale un autovalore. Questa aiuta a scrivere la precedente espressione nella forma compatta

$$\underline{A}\underline{X} = \underline{X}\underline{D},$$

che altri non è che la proprietà (1), dove invece di considerare un autovettore  $\underline{x}_i$  per volta, li consideriamo tutti assieme, messi in un'unica matrice  $\underline{X}$ . Quest'ultima proprietà può essere riscritta moltiplicando ambo i membri da destra per  $\underline{X}^{-1}$ , ottenendo

$$\underline{A} = \underline{X}\underline{D}\underline{X}^{-1}. \tag{3}$$

Questa proprietà dimostra come sia possibile **diagonalizzare** la matrice  $\underline{A}$ , ovvero scriverla come prodotto di tre matrici, dove al centro di questa specie di *panino* abbiamo una matrice diagonale

<sup>5</sup>il cui calcolo richiede il prodotto degli elementi della diagonale, arrivando quindi a  $\lambda^n$

a fare da *guarnizione*, e  $\underline{X}, \underline{X}^{-1}$  a fare da *fette di pane*. Questa rappresentazione è particolarmente interessante perché ci permette di enunciare la prima situazione in cui la decomposizione agli autovalori è fondamentale. Immaginiamo di voler calcolare il quadrato di una matrice,  $\underline{A}^2$ . Anziché calcolare esplicitamente il prodotto, è possibile sostituire (3), ottenendo

$$\underline{A}^2 = (\underline{X}\underline{D}\underline{X}^{-1})(\underline{X}\underline{D}\underline{X}^{-1}) = \underline{X}\underline{D}\underline{D}\underline{X}^{-1},$$

dove però calcolare  $\underline{D}\underline{D} = \underline{D}^2$  è assolutamente banale: è sufficiente calcolare il quadrato di ciascun elemento sulla diagonale, ovvero di ciascun autovalore! Questa cosa, ovviamente, si può generalizzare con grande facilità, permettendoci di ottenere

$$\underline{A}^k = \underline{X}\underline{D}^k\underline{X}^{-1},$$

dove  $d_{ii}^k = \lambda_i^k$ . Ma questa cosa può valere anche per l'inversa della matrice! Infatti,

$$\underline{A}^{-1} = \underline{X}\underline{D}^{-1}\underline{X}^{-1},$$

dove quindi è sufficiente calcolare ciascuna componente come il reciproco di ciascun autovalore. Questa idea può essere sfruttata al fine di definire e calcolare l'esponenziale di una matrice. Sappiamo infatti che lo sviluppo di Taylor di una funzione esponenziale è:

$$e^x \simeq 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!}.$$

Allora, possiamo definire, analogamente,

$$e^{\underline{A}} \simeq \underline{I} + \underline{A} + \frac{1}{2}\underline{A}^2 + \frac{1}{3!}\underline{A}^3 + \dots + \frac{1}{n!}\underline{A}^n,$$

dove ciascuno degli  $\underline{A}^k$  si può scrivere con la sua forma diagonalizzata, permettendoci di ottenere

$$e^{\underline{A}} \simeq \underline{X}\left[\underline{I} + \underline{D} + \frac{1}{2}\underline{D}^2 + \frac{1}{3!}\underline{D}^3 + \dots + \frac{1}{n!}\underline{D}^n\right]\underline{X}^{-1},$$

ovvero,

$$e^{\underline{A}} = \underline{X}e^{\underline{D}}\underline{X}^{-1},$$

dove però è chiaro come si fa a calcolare l'esponenziale della matrice centrale: è sufficiente calcolare tutti gli esponenziali dei vari elementi sulla diagonale!

## 1.2 Quoziente di Rayleigh

Introduciamo a questo punto il concetto di quoziente di Rayleigh. Partendo da una matrice  $\underline{A}$ , da un suo autovalore  $\lambda$  e dall'autovettore a esso associato,  $\underline{x}$ , vale (1):

$$\underline{A}\underline{x} = \lambda\underline{x}.$$

A questa espressione, moltiplichiamo ambo i membri da sinistra per il vettore  $\underline{x}^H$ , detto **Hermitiano** del vettore  $\underline{x}$ . Questo è semplicemente il vettore trasposto di  $\underline{x}$ , dove al posto di usare ciascun elemento, se ne prende il **complesso coniugato**; nel caso in cui le componenti di  $\underline{x}$  fossero reali, allora questa operazione è semplicemente la trasposizione. Si ottiene:

$$\underline{x}^H \underline{A} \underline{x} = \lambda \underline{x}^H \underline{x}.$$

Possiamo osservare che sia il membro sinistro, sia il membro destro, sono dei numeri; di conseguenza, è lecito isolare  $\lambda$  ottenendo la seguente espressione:

$$\lambda = \frac{\underline{x}^H \underline{A} \underline{x}}{\underline{x}^H \underline{x}}.$$

Il rapporto

$$r_{\underline{A}} = \frac{\underline{x}^H \underline{A} \underline{x}}{\underline{x}^H \underline{x}} \quad (4)$$

viene detto **quoziente di Rayleigh** e, se  $\underline{x}$  è un autovettore della matrice  $\underline{A}$ , questo coincide con l'autovalore  $\lambda$  a esso associato. Questo strumento è molto utile dal momento che ci permette, una volta noto un certo autovettore, di poter ricavare immediatamente l'autovalore a esso associato, con un conticino molto semplice. Spesso, negli esercizi di Algebra Lineare, quello che si fa è infatti il contrario, ossia risolvere l'equazione caratteristica e ricavare l'elenco degli autovalori, e poi, risolvendo per ciascuno di essi il corrispondente sistema lineare omogeneo, trovare gli autovettori. Alcuni dei metodi che studieremo si basano sull'effettuare varie iterazioni di un certo algoritmo, dove il risultato di ciascuna iterazione è una migliore stima dell'autovettore; noto l'autovettore, potremo trovare quindi immediatamente l'autovalore, per procedere con la successiva iterazione.

È giunto il momento di scavare nel nostro passato, e chiudere alcuni dei conti lasciati in sospeso.

«Abbiamo già visto queste espressioni? Numeratore, denominatore... Sembrano familiari!»

Prima di tutto, concentriamoci sul denominatore, che è quello più semplice: è semplicemente la norma 2 del vettore  $\underline{x}$ !

$$\|\underline{x}\|_2^2 = \underline{x}^T \underline{x},$$

cosa che vale anche quando si prende l'Hermitiano.

«E il numeratore?»

Questo è un po' più difficile ma, se ci sforziamo, possiamo ricordarci che è il termine che appariva nella definizione di matrice simmetrica definita positiva! E questo, finalmente, ci permette di dare un criterio operativo: **una matrice è simmetrica definita positiva se e solo se tutti i suoi autovalori sono positivi**. Prima di tutto, un passo preliminare riguarda la dimostrazione del fatto che se una matrice è simmetrica allora i suoi autovettori sono ortogonali. Infatti, dire che una matrice è simmetrica significa dire che è uguale alla sua trasposta:

$$\underline{A} = \underline{A}^T.$$

Sostituiamo in questa relazione l'espressione del panino (3):

$$\underline{X} \underline{D} \underline{X}^{-1} = (\underline{X} \underline{D} \underline{X}^{-1})^T = (\underline{X}^{-1})^T \underline{D}^T \underline{X}^T.$$

Dal momento che  $\underline{D}$  è diagonale, questa eguaglianza implica che

$$\underline{X} = (\underline{X}^{-1})^T,$$

e, equivalentemente, che

$$\underline{X}^T = \underline{X}^{-1},$$

cosa che si può ottenere sia trasponendo entrambe le matrici della precedente espressione, sia considerando l'eguaglianza delle matrici a destra del panino. In ogni caso, quanto abbiamo appena scritto dimostra che la matrice degli autovettori,  $\underline{X}$ , è ortogonale<sup>6</sup>. Dal momento che  $\underline{X}$  è costruita incolonnando i vari autovettori, possiamo dire che **gli autovettori di una matrice simmetrica sono tra loro ortogonali**.

Al fine di proseguire con la nostra dimostrazione, prendiamo un generico vettore  $\underline{u}$ , che dunque non è necessariamente un autovettore della matrice. Tuttavia,  $\underline{u}$  può essere certamente scritto come una combinazione lineare di autovettori; considerando<sup>7</sup> per esempio una situazione tale per cui  $\underline{u}$  si può scrivere semplicemente come combinazione lineare di due autovettori, si avrebbe:

<sup>6</sup>dal momento che la sua inversa è uguale alla sua trasposta ;-)

<sup>7</sup>per fare un esempio semplice, ma la prova con  $n$  autovettori sarebbe del tutto simile

$$\underline{u} = \alpha \underline{x}_1 + \beta \underline{x}_2.$$

A questo punto se calcoliamo  $\underline{A}\underline{u}$  otteniamo, sostituendo la riscrittura in termini di combinazione lineare di autovettori,

$$\underline{A}\underline{u} = \underline{A}(\alpha \underline{x}_1 + \beta \underline{x}_2) = \alpha \underline{A}\underline{x}_1 + \beta \underline{A}\underline{x}_2 = \alpha \lambda_1 \underline{x}_1 + \beta \lambda_2 \underline{x}_2.$$

A questo punto, moltiplichiamo da sinistra il vettore  $\underline{u}^T$ :

$$\underline{u}^T \underline{A}\underline{u} = (\alpha \underline{x}_1^T + \beta \lambda_2 \underline{x}_2^T)(\alpha \lambda_1 \underline{x}_1 + \beta \lambda_2 \underline{x}_2) = \alpha^2 \lambda_1 \underline{x}_1^T \underline{x}_1 + \alpha \beta \lambda_2 \underline{x}_1^T \underline{x}_2 + \beta \alpha \lambda_1 \underline{x}_2^T \underline{x}_1 + \beta^2 \lambda_2 \underline{x}_2^T \underline{x}_2,$$

dove però i prodotti tra autovettori diversi sono nulli, dal momento che abbiamo appena dimostrato che la matrice degli autovettori è ortogonale. Riconosciamo che gli altri termini contengono i quadrati delle norme 2 degli autovettori; possiamo quindi scrivere che

$$\underline{u}^T \underline{A}\underline{u} = \alpha^2 \|\underline{x}_1\|_2^2 \lambda_1 + \beta^2 \|\underline{x}_2\|_2^2 \lambda_2$$

e, se gli autovalori  $\lambda_1$  e  $\lambda_2$  sono entrambi positivi, è evidente che dovrà essere positivo anche il membro sinistro<sup>8</sup>, dimostrando che una matrice simmetrica avente autovalori positivi è anche definita positiva.

Credo che per dimostrare nell'altro senso la proprietà si possano sempre usare questi risultati. In ogni caso, il mio obiettivo era, più che dimostrarvi queste proprietà, farvi vedere un po' di questi calcoletti, che possono sempre tornare utili.

### 1.3 Raggio spettrale e norma spettrale di una matrice

Introduciamo a questo punto un altro concetto: quello di **raggio spettrale**. Data una matrice  $\underline{A}$ , il suo raggio spettrale è il **modulo dell'autovalore di massimo modulo**:

$$\rho(\underline{A}) = \max_{i=1, \dots, n} \{|\lambda_i| : \lambda_i \text{ autovalore di } \underline{A}\}. \quad (5)$$

Calcolare questo numero mediante MATLAB<sup>®</sup> non è molto difficile:

```
rho = max(abs(eig(A)));
```

Questa definizione ci permette di chiudere un altro argomento lasciato in sospeso. Infatti, non ho mai detto come si fa a calcolare la norma 2 di una **matrice**! Beh, questa si definisce come

$$\|\underline{A}\|_2 = \sqrt{\rho(\underline{A}^T \underline{A})}. \quad (6)$$

Esistono situazioni in cui questa espressione si semplifica, per esempio in caso di matrici simmetriche. Infatti, se  $\underline{A}$  è una matrice simmetrica, si ha che

$$\underline{A} = \underline{A}^T,$$

che implica

$$\underline{A}^T \underline{A} = \underline{A}^2.$$

In questa situazione, quindi,

$$\rho(\underline{A}^T \underline{A}) = \rho(\underline{A}^2),$$

ma, poiché gli autovalori di  $\underline{A}^2$  sono uguali agli autovalori di  $\underline{A}$ , elevati al quadrato<sup>9</sup>, ed essendo il raggio spettrale semplicemente il modulo di un autovalore,

<sup>8</sup>infatti, tutti gli altri termini contenuti nell'espressione sono al quadrato, e quindi non influenzano il segno dei vari termini

<sup>9</sup>lo abbiamo dimostrato usando la proprietà del panino (3)

$$\rho(\underline{A}^2) = [\rho(\underline{A})]^2,$$

e, quindi, se la matrice  $\underline{A}$  è simmetrica,

$$\|\underline{A}\|_2 = \sqrt{\rho(\underline{A}^T \underline{A})} = \sqrt{\rho(\underline{A}^2)} = \sqrt{[\rho(\underline{A})]^2} = \rho(\underline{A}).$$

Tutto questo discorso ci ha permesso di definire la norma 2, o **norma spettrale** di una matrice; tuttavia, essa si può anche calcolare con MATLAB<sup>®</sup> usando il comando `norm(A,2)`, o anche semplicemente `norm(A)`: quando non si specifica quale norma si desidera, MATLAB<sup>®</sup> ritorna la norma 2.

## 1.4 Cerchi di Gershgorin

Il nostro obiettivo sarà calcolare, mediante algoritmi, gli autovalori di una matrice. Tuttavia, alcuni di questi algoritmi richiederanno una stima della posizione di questi autovalori, che poi verrà raffinata mediante un procedimento iterativo, quindi un procedimento *a più passi*. Uno strumento matematico che permette di determinare delle stime sono i **cerchi di Gershgorin**. Ne esistono di due tipi: **cerchi riga** e **cerchi colonna**. Partiamo dai primi: si può definire l'area  $C_i^{(r)}$  identificata dall'espressione

$$C_i^{(r)} = \left\{ z \in \mathbb{C} : |z - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}| \right\}, \quad i = 1, \dots, n. \quad (7)$$

Qua abbiamo bisogno di discutere un po'. Prima di tutto, notiamo che  $z$  è una variabile complessa, avente quindi parte immaginaria non necessariamente nulla; quando si parla di autovalori, spesso capita di avere a che fare con dei numeri complessi, dal momento che essi, come già detto, sono soluzioni di un'equazione algebrica le cui radici non sono necessariamente reali. Come dovrebbe essere noto da Analisi Matematica I, se un numero reale si può rappresentare geometricamente come un punto su una retta, allora un numero complesso si può rappresentare geometricamente come un punto su un piano, le cui ascisse e ordinate indicano la parte reale e immaginaria. L'espressione (8) contiene un'espressione tipo

$$|z - z_c| \leq r.$$

Questa identifica l'area di un cerchio nel piano complesso centrato in  $z_c = a_{ii}$  ( $i$ -esimo elemento diagonale della matrice  $\underline{A}$ ) e avente raggio  $r$  pari alla somma dei moduli degli elementi sulla  $i$ -esima riga,

$$r = \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|.$$

Si tratta di un cerchio perché è il luogo dei punti  $z$  nel piano complesso avente distanza dal punto  $z_c$  pari al raggio  $r$ ; visto che prendiamo il minore-uguale  $\leq$  non consideriamo solo la circonferenza, ma tutto l'insieme di punti.

Per i cerchi colonna, non c'è molto di più da dire:

$$C_j^{(c)} = \left\{ z \in \mathbb{C} : |z - a_{jj}| \leq \sum_{\substack{i=1 \\ i \neq j}}^n |a_{ij}| \right\}, \quad i = j, \dots, n. \quad (8)$$

Ossia, il centro di questi cerchi è sempre l'elemento diagonale  $a_{jj}$ , ma questa volta i raggi di ciascuno dei cerchi si calcolano sommando gli elementi lungo le colonne, invece che lungo le righe. Un esempio di questi cerchi è riportato in Fig. 1

Per una matrice  $\underline{A} \in \mathbb{R}^{n,n}$ , avremo  $n$  cerchi riga e  $n$  cerchi colonna.

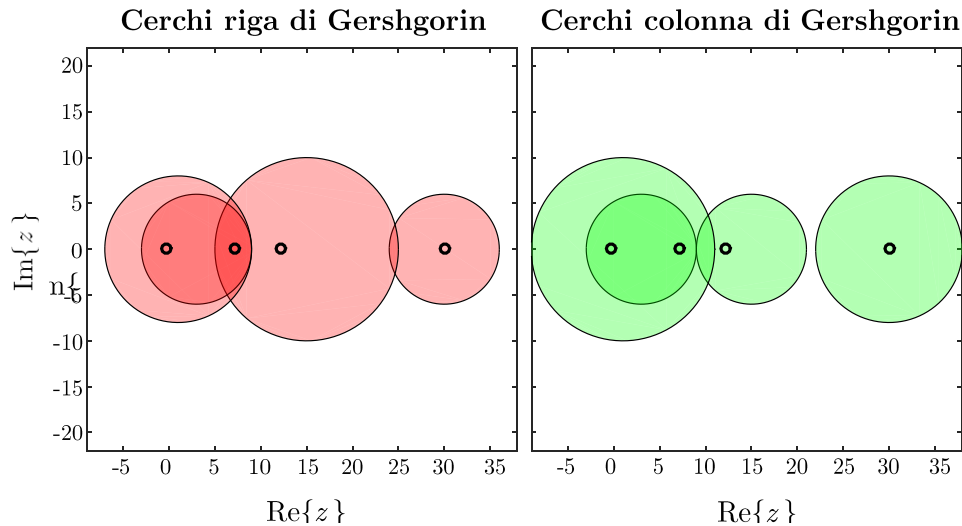


Figura 1: Esempio di cerchi riga (sinistra) e colonna (destra) di Gershgorin, ottenuti con il programma riportato in Appendice A.

Il motivo per cui ci interessiamo di questi cerchi è il loro legame con gli autovalori della matrice dai quali sono stati ricavati. Definiamo infatti  $\mathcal{R}$  e  $\mathcal{C}$  le unioni dei cerchi di Gershgorin:

$$\mathcal{R} = \bigcup_{i=1}^n C_i^{(r)}, \quad \mathcal{C} = \bigcup_{i=1}^n C_i^{(c)}.$$

Un teorema ci garantisce che **tutti gli autovalori della matrice  $\underline{A}$  appartengono sia a  $\mathcal{R}$ , sia a  $\mathcal{C}$** . Ma non solo: è anche possibile dire che **tutti gli autovalori della matrice  $\underline{A}$  appartengono all'intersezione delle regioni  $\mathcal{C}$  e  $\mathcal{R}$** ; è dunque possibile delimitare ulteriormente la superficie in cui possiamo trovare questi autovalori! In Fig. 1, queste cose si possono vedere *a occhio*, dal momento che sono stati riportati gli autovalori mediante cerchietti neri.

Questi cerchi forniscono altre indicazioni: immaginiamo di non *unire* tutte le regioni, ma solamente  $k$  e le restanti  $n - k$  in due diversi gruppi, in modo tale che questi due gruppi siano tra loro disgiunti. Nel caso di Fig. 1, cerchi colonna, potremmo per esempio usare come regione  $\mathcal{R}_1$  i tre cerchi sulla sinistra, le cui intersezioni non sono nulle, e come regione  $\mathcal{R}_2$  l'unico cerchio sulla destra. Beh, questa cosa è utile perché un teorema ci garantisce che

- dal momento che la regione  $\mathcal{R}_1$  è costituita da 3 cerchi tra loro intersecati, essa conterrà 3 autovalori;
- dal momento che la regione  $\mathcal{R}_2$  è costituita da 1 solo cerchio, essa conterrà 1 solo autovalore.

Insomma, il numero di autovalori contenuti in regioni diverse in quanto non sovrapposte, è pari al numero di cerchi costituiti da queste regioni. Questo vale sia per i cerchi riga, sia per i cerchi colonna.

## 1.5 Matrici simili

Una definizione di una relazione tra due matrici, assistita dai concetti di autovalori e autovettori, è quella di **matrici simili**. Due matrici  $\underline{A}$  e  $\underline{B}$ , entrambe aventi  $n$  righe e  $n$  colonne, vengono dette simili se esiste una certa matrice  $\underline{S}$  avente  $n$  righe e  $n$  colonne, non singolare, tale per cui

$$\underline{S}^{-1} \underline{A} \underline{S} = \underline{B}.$$

Che questa definizione sia imparentata con autovalori e autovettori è abbastanza evidente, in quanto questa definizione ricorda terribilmente la *relazione panino* (3). Inoltre, si può dimostrare che due matrici simili  $\underline{A}$  e  $\underline{B}$  hanno gli stessi autovalori. Infatti, se



$$\underline{\underline{B}} = \underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{S}},$$

e

$$\underline{\underline{A}} x = \lambda x,$$

è possibile scrivere, moltiplicando da sinistra per  $\underline{\underline{S}}^{-1}$ ,

$$\underline{\underline{S}}^{-1} \underline{\underline{A}} x = \lambda \underline{\underline{S}}^{-1} x.$$

Posso a questo punto applicare un trucchetto: posso far apparire il prodotto  $\underline{\underline{S}} \underline{\underline{S}}^{-1}$  nell'espressione, cosa lecita (è un po' come moltiplicare e dividere per un numero):

$$\underline{\underline{S}}^{-1} \underline{\underline{A}} x = \underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{I}} x = \underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{S}} \underline{\underline{S}}^{-1} x = \underbrace{[\underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{S}}]}_{\underline{\underline{B}}} \underline{\underline{S}}^{-1} x.$$

Qui, abbiamo potuto identificare la matrice  $\underline{\underline{B}}$ . Quindi, possiamo riscrivere questa espressione come

$$\underline{\underline{B}} \underbrace{\underline{\underline{S}}^{-1} x}_z = \lambda \underbrace{\underline{\underline{S}}^{-1} x}_z,$$

dove stiamo suggerendo la definizione di un vettore

$$z = \underline{\underline{S}}^{-1} x.$$

Ma allora, è possibile scrivere l'ultima espressione come

$$\underline{\underline{B}} z = \lambda z,$$

che ci dice che  $\lambda$ , autovalore della matrice  $\underline{\underline{A}}$  per ipotesi, è anche un autovalore della matrice  $\underline{\underline{B}}$ , ma con un autovettore diverso: se l'autovettore corrispondente, per  $\underline{\underline{A}}$ , era  $x$ , ora invece è questo  $z$ .

Data dunque una matrice  $\underline{\underline{A}}$ , questa si dice **diagonalizzabile** se è simile (ovvero, se condivide gli stessi autovalori) a una matrice diagonale, contenente gli autovalori sulla diagonale:

$$\underline{\underline{S}}^{-1} \underline{\underline{A}} \underline{\underline{S}} = \underline{\underline{D}}.$$

Che, moltiplicando da sinistra per  $\underline{\underline{S}}$ , permette di ottenere

$$\underline{\underline{A}} \underline{\underline{S}} = \underline{\underline{S}} \underline{\underline{D}},$$

che è esattamente la *relazione del panino* (3), solo con nomi diversi.

Detto in un altro modo, una matrice  $\underline{\underline{A}}$  di ordine  $n$  è diagonalizzabile se e solo se essa possiede  $n$  autovettori linearmente indipendenti. Infatti, se così non fosse, la matrice  $\underline{\underline{S}}$  avrebbe rango non massimo, non sarebbe invertibile, e quindi non sarebbe possibile applicare la *relazione del panino*. Esiste tuttavia un risultato che garantisce che una matrice è diagonalizzabile se i suoi autovalori sono distinti; questa è una relazione solo in un verso (non c'è il *solo se*), nel senso che non è detto che, se gli autovalori non sono distinti<sup>10</sup>, la matrice non sia diagonalizzabile.

## 1.6 Condizionamento del calcolo degli autovalori

Come esiste il problema del condizionamento nell'ambito della soluzione dei sistemi lineari, così esiste quello relativo al calcolo degli autovalori. Ricordiamo velocemente di cosa si parla: se gli elementi della matrice di partenza  $\underline{\underline{A}}$  sono soggetti a una perturbazione, generata per esempio dagli errori di arrotondamento che commettiamo al momento di memorizzarla, di quanto variano gli autovalori, rispetto al caso che consideriamo *ideale*? La risposta è: possono anche variare un casino. Considerando per esempio la matrice

<sup>10</sup>ovvero se alcuni di essi hanno molteplicità maggiore di 1

$$\underline{A} = \begin{bmatrix} 101 & -90 \\ 110 & -98 \end{bmatrix},$$

e immaginiamo di perturbare alcuni suoi elementi:

$$\tilde{\underline{A}} = \begin{bmatrix} 101 - \varepsilon & -90 - \varepsilon \\ 110 & -98 \end{bmatrix}.$$

Vediamo cosa succede nel caso imperturbato e in quello perturbato, con una perturbazione  $\varepsilon = 0.001$ . Lanciando lo script

```
clear
close all
clc

epsilon=0.001; % perturbazione

% Inserisco la matrice esatta
A = [101 -90;
     110 -98];

% Inserisco la matrice perturbata
Atilde = [101-epsilon -90-epsilon;
          110          -98];

E=eig(A) % autovalori della matrice esatta
Etilde=eig(Atilde) % autovalori della matrice perturbata
norm(A-Atilde,2) % errore di perturbazione sulla matrice
errrel = abs(E-Etilde)./abs(E) % errore relativo sugli autovalori
```

si arriva a dimostrare che, con un errore in norma 2 tra le matrici pari allo 0.1%, si hanno errori del 30% sugli autovalori! Un autovalore, infatti, passa dall'essere pari a 2, all'essere 1.7: terrificante! E questo, con una banalissima matrice  $2 \times 2$  !!!

È evidente che quindi anche il calcolo degli autovalori è soggetto al problema del condizionamento. Tuttavia, non è ancora chiaro dove questo nasca. A chiarirci le idee ci viene incontro il **teorema di Bauer-Fike**. Questo afferma che, data  $\underline{A}$  una matrice diagonalizzabile e  $\underline{S}$  invertibile tale per cui

$$\underline{S}^{-1} \underline{A} \underline{S} = \underline{D},$$

dove  $\underline{D}$  è la matrice diagonale avente gli autovalori sulla diagonale principale, allora

$$\min_{1 \leq i \leq n} |\tilde{\lambda} - \lambda_i| \leq \kappa(\underline{S}) \left\| \underline{A} - \tilde{\underline{A}} \right\|, \quad (9)$$

dove  $\tilde{\underline{A}}$  è il risultato di una perturbazione sulla matrice  $\underline{A}$ ,  $\tilde{\lambda}$  è un certo autovalore di  $\tilde{\underline{A}}$ , e

$$\kappa(\underline{S}) = \left\| \underline{S} \right\| \left\| \underline{S}^{-1} \right\|,$$

con i risultati validi per le norme 1, 2,  $\infty$ .

Cosa significa tutto questo? Prima di tutto è importante capire cosa voglia dire il membro sinistro, quel min. A una perturbazione sulla matrice corrisponde una perturbazione amplificata di  $\kappa(\underline{S})$  sugli autovalori; non è tuttavia scontato riuscire a *riconoscere* quale autovalore è diventato quale altro: se la perturbazione è grossa, potrebbe essere che non si riesca più a effettuare questo riconoscimento<sup>11</sup>. Il caso di prima dove passavamo da 2 a 1.7, per quanto traumatico, non era ingestibile. Visto che non sappiamo bene come un autovalore possa trasformarsi in un altro, tutto ciò che questo teorema può fare è usare questa minima differenza, sperando che l'autovalore trasformando nasca da quello *calcolato idealmente* a esso corrispondente.

Parlando del membro destro, sulla norma della differenza delle matrici, non c'è molto da dire: è una stima della perturbazione. Invece,  $\kappa(\underline{S})$  è molto interessante, perché assomiglia molto a quello dei sistemi lineari, ma applicato non a  $\underline{A}$ , ma a  $\underline{S}$ , ossia alla **matrice degli autovettori**! Se  $\kappa(\underline{S})$  è prossimo a 1, non avremo grosse perturbazioni, ma se fosse grande, avremmo un'enorme amplificazione della perturbazione sui dati.

<sup>11</sup>un check aggiuntivo si potrebbe ottenere guardando gli autovettori, ma non discutiamo queste cose che non servono a comprendere il discorso...

Riassumendo: se la matrice degli autovettori  $\underline{S}$  è mal condizionata, ossia composta da autovettori che *non sono molto linearmente indipendenti*, allora il calcolo degli autovettori sarà mal condizionato. Voglio però rimarcare che il condizionamento del calcolo degli autovettori e della soluzione di un sistema lineare sono **due problemi del tutto disgiunti**. Un esempio eclatante è dato dalla matrice di Hilbert: come noto, la soluzione di sistemi lineari aventi la matrice di Hilbert come matrice di sistema è un problema mal condizionato. Il calcolo dei suoi autovalori, invece, è assolutamente ben condizionato! Infatti, la matrice di Hilbert è simmetrica; prima, però, abbiamo dimostrato che una matrice simmetrica ha autovettori ortogonali. E ortogonali, significa che sono **molto indipendenti linearmente!**

In generale, per una matrice  $\underline{A}$  simmetrica, dal momento che la sua matrice degli autovettori  $\underline{S}$  è ortogonale, si ha che il numero di condizionamento in norma 2 è pari a 1; infatti:

$$\kappa_2(\underline{S}) = \|\underline{S}\|_2 \|\underline{S}^T\|_2 = \sqrt{\rho(\underline{S}^T \underline{S})} \sqrt{\rho(\underline{S} \underline{S}^T)} = \sqrt{\rho(\underline{I})} \sqrt{\rho(\underline{I})} = 1.$$

Oltre a calcolare il numero di condizionamento relativo alla soluzione di sistemi lineari, MATLAB<sup>®</sup> può calcolare anche il numero di condizionamento relativo alla determinazione dei suoi autovalori, usando il comando `condeig`.

Prima di concludere, può valere la pena di proporre la seguente riflessione sul teorema di Bauer-Fike e in particolare sull'espressione (9). Da un punto di vista concettuale, questa è assolutamente diversa da quanto si è visto nell'ambito dei sistemi lineari, in quanto il problema del condizionamento, in quel contesto, era stato applicato sull'errore **relativo** sui dati di ingresso, mentre qui abbiamo a che fare con errori **assoluti**. Tuttavia, al fine di *completare* questo ragionamento e ricercare un'analogia con il condizionamento di sistemi lineari, si tenga presente che vale la seguente relazione:

$$\min_{1 \leq i \leq n} \frac{|\tilde{\lambda} - \lambda_i|}{|\lambda_i|} \leq \kappa(\underline{S}) \left\| \underline{A}^{-1} (\underline{A} - \tilde{\underline{A}}) \right\|,$$

dove al membro sinistro calcoliamo il minimo relativo all'autovalore  $\lambda_i$ , mentre a membro destro introduciamo questa  $\underline{A}^{-1}$ , che ci permette di definire, intuitivamente, una sorta di *errore relativo di matrici*.

## A Un programmino per disegnare i cerchi di Gershgorin

Vengono ora riportati lo script per la generazione di Fig. 1, con la function utilizzata.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Questo programma puo` essere usato per produrre i cerchi di Gershgorin
% per un esempio di matrice e per visualizzare i suoi autovalori, in modo
% da cercare di chiarire questi concetti.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear
close all
clc

% Esempio di matrice
A=[30 1 2 3;
  4 15 -4 -2;
  -1 0 3 5;
  -3 5 0 1];

fLez10_Gershgorin(A)

% Calcolo gli autovalori della matrice
E=eig(A);

% Disegno gli autovalori della matrice, in parte reale e immaginaria,
% mediante dei cerchietti neri
subplot(1,2,1)
for ind=1:length(E)
  plot(real(E(ind)),imag(E(ind)), 'ko', 'LineWidth', 2)
end
ylim([-22,22])

% Disegno gli autovalori della matrice, in parte reale e immaginaria,
% mediante dei cerchietti neri
subplot(1,2,2)

```

```

for ind=1:length(E)
    plot(real(E(ind)),imag(E(ind)), 'ko', 'LineWidth', 2)
end
ylim([-22,22])

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Questa funzione e` stata tratta dalle soluzioni all'esercitazione 4, e
% viene commentata e modificata lievemente al fine di abbellire i grafici.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function fLez10_Gershgorin(A)

n = size(A,1); % trovo la dimensione della matrice

% Disegno dei cerchi riga di Gershgorin
figure(1)
set(gcf, 'Position', [277 440 1091 420])
subplot(1,2,1)
box on
centro = diag(A); % centri dei cerchi di Gershgorin
raggio_riga = sum(abs(A),2)-diag(abs(A));
theta = linspace(0,2*pi); % angolo per il disegno dei cerchi
for i=1:n % per ciascuna delle righe della matrice disegno un cerchio
    % Senza entrare in dettaglio, si sfrutta la parametrizzazione della
    % circonferenza mediante le funzioni seno e coseno.
    x = centro(i)+raggio_riga(i)*cos(theta);
    y = raggio_riga(i)*sin(theta);
    patch(x,y,'r') % il comando patch permette di disegnare i cerchi
    axis equal
    hold on
end
alpha(0.3) % questo comando cambia la trasparenza dei cerchi

% Disegno dei cerchi colonna di Gershgorin
subplot(1,2,2)
box on
centro = diag(A);
raggio_colonna = sum(abs(A),1)-diag(abs(A));
theta = linspace(0,2*pi);
for i=1:n
    x = centro(i)+raggio_colonna(i)*cos(theta);
    y = raggio_colonna(i)*sin(theta);
    patch(x,y,'g')
    axis equal
    hold on
end
alpha(0.3)

```

## Riferimenti bibliografici

- [1] L. A. Steen, "Highlights in the history of spectral theory," *The American Mathematical Monthly*, vol. 80, no. 4, pp. 359-381, Apr. 1973.