

APPUNTI DI CALCOLO NUMERICO

Sistemi lineari

Richiami di algebra lineare

Consideriamo matrici quadrate.

Una matrice A si dice diagonale se $a_{ij} = 0$ se $i \neq j$.

Una matrice A si dice triangolare superiore se $a_{ij} = 0$ se $i > j$.

Una matrice A si dice triangolare inferiore se $a_{ij} = 0$ se $i < j$.

Una matrice D si dice diagonale invertibile se $d_{ij} \neq 0$ se $i = j$ e la sua inversa D^{-1} si ottiene invertendo ciascun termine sulla diagonale.

Moltiplicando una matrice diagonale D per una matrice A qualsiasi si ottiene un riscalamento: in particolare il prodotto DA riscalda le righe, mentre il prodotto AD riscalda le colonne.

Norma di vettori e di matrici

È una funzione $\|\cdot\|: V \mapsto \mathfrak{R}$ con $V = \mathfrak{R}^n$ per i vettori e $V = \mathfrak{R}^{n \times n}$ per le matrici tale che:

1. $\|x\| \geq 0 \quad \forall x \in V, \quad \|x\| = 0 \Leftrightarrow x = 0$;
2. $\|\alpha x\| = |\alpha| \|x\| \quad \forall x \in V \quad \forall \alpha \in \mathfrak{R}$;
3. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V$ (disuguaglianza triangolare)
4. per le sole matrici: $\|AB\| \leq \|A\| \|B\| \quad \forall A, B \in \mathfrak{R}^{n \times n}$

Vettori

Norma 1:

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

Norma 2 o Norma euclidea:

$$\|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{1/2}$$

Norma ∞ o Norma del massimo

$$\|x\|_\infty = \max_i |x_i|$$

Tutte le norme sono equivalenti.

Esempio:

$$x = \begin{pmatrix} -3 \\ 4 \\ -5 \end{pmatrix}$$

$$\|x\|_1 = 12$$

$$\|x\|_2 = \sqrt{50}$$

$$\|x\|_\infty = 5$$

Matrici

Norma 1:

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

Norma 2 o Norma spettrale:

$$\|A\|_2 = \sqrt{\rho(A^T A)} \text{ dove } \rho(B) = \max_i |\lambda_i(B)| \text{ è il raggio spettrale della matrice } B$$

Norma ∞ :

$$\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$$

Norma di Frobenius:

$$\|A\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \right)^{1/2}$$

Tutte le norme sono equivalenti.

Esempio:

$$A = \begin{pmatrix} -3 & 4 \\ -5 & -6 \end{pmatrix}$$

$$\|A\|_1 = 10$$

$$\|A\|_\infty = 11$$

$$\|A\|_F = \sqrt{86}$$

Una norma è detta naturale o indotta se vale la seguente relazione:

$$\|\cdot\|_M = \|A\|_M = \max_{\|x\|_V=1} \|Ax\|_V = \sup_{\|x\|_V \neq 0} \frac{\|Ax\|_V}{\|x\|_V}$$

dove con $\|\cdot\|_M$ si intende la norma di matrice e con $\|\cdot\|_V$ si indica la norma di vettore.

Si dimostra che per $*$ = 1,2,∞ la norma $\|A\|_*$ è indotta dalla norma $\|x\|_*$. Si dimostra anche che la norma di Frobenius non è naturale in quanto, essendo $\|I\| = 1$ per ogni norma naturale, si ottiene $\|I\|_F = \sqrt{n} \neq 1$ dove n è il numero di righe (o di colonne) della matrice identità I .

La norma di un vettore e la norma di una matrice si dicono compatibili se vale:

$$\|Ax\|_V \leq \|A\|_M \|x\|_V \quad \forall A, x$$

Una norma naturale è sempre compatibile con la norma vettore che la induce, per costruzione.

Condizionamento di un sistema lineare

Sia $Ax = b$ un sistema lineare.

Perturbiamo il dato b (vettore dei termini noti)

$$b \rightarrow b + \delta b.$$

Il sistema diventa quindi

$$Ax = b \rightarrow A\bar{x} = b + \delta b.$$

Sia $\delta x = \bar{x} - x$ il disturbo sul risultato.

$$A(x + \delta x) = b + \delta b$$

$$Ax + A\delta x = b + \delta b$$

Dato che $Ax = b$, semplificando:

$$A\delta x = \delta b$$

Quindi, formalmente, si può scrivere

$$\delta x = A^{-1} \delta b.$$

Siamo certi che la matrice A^{-1} esista perché il problema è ben posto, quindi la sua matrice dei coefficienti A è quadrata ed è invertibile.

Usando due norme compatibili si può determinare il numero di condizionamento del sistema (o della matrice):

$$\begin{aligned}\delta x = A^{-1} \delta b &\Rightarrow \|\delta x\| = \|A^{-1} \delta b\| \leq \|A^{-1}\| \|\delta b\| \\ Ax = b &\Rightarrow \|b\| = \|Ax\| \leq \|A\| \|x\| \Rightarrow \frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|}\end{aligned}$$

moltiplicando le due ultime espressioni si ottiene

$$\frac{\|\delta x\|}{\|x\|} \leq \|A^{-1}\| \|b\| \frac{\|A\|}{\|b\|} = \|A\| \|A^{-1}\| \frac{\|\delta b\|}{\|b\|}$$

chiamiamo $K(A) = \|A\| \|A^{-1}\|$ in numero di condizionamento del sistema di matrice A .

Si evince che il condizionamento di un sistema lineare dipende solamente dalla matrice A dei coefficienti e non dal vettore dei termini noti, sia che si vadano a perturbare i dati in ingresso sia che si aggiunga un disturbo ai termini di A .

2

N.B.: con le norme naturali è verificato che $K(A) \geq 1$ perché

$$K(A) = \|A\| \|A^{-1}\| \geq \|AA^{-1}\| = \|I\| = 1$$

Metodi diretti

Un particolare tipo di problemi sono i cosiddetti problemi triangolari. Essi sono particolarmente semplici.

Un sistema di questo tipo ha la matrice dei coefficienti triangolare, quadrata e non singolare, cioè invertibile (quindi il determinante è dato dal prodotto dei termini sulla diagonale ed è non nullo).

In caso di matrice triangolare inferiore il sistema risulta nella forma

$$Ly = b$$

ed è possibile ricavare dalla prima equazione l'incognita $y_1 = \frac{b_1}{l_{11}}$. Sostituendo y_1

nella seconda equazione è possibile ricavare $y_2 = \frac{b_2 - l_{21}y_1}{l_{22}}$ e così via fino all'ultima

equazione. In generale $y_i = \frac{\left(b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right)}{l_{ii}}$. Questo metodo di risoluzione si chiama sostituzione in avanti.

In caso di matrice triangolare superiore il metodo di risoluzione è del tutto analogo, tuttavia l'equazione da cui iniziare è l'ultima, per poi sostituire la soluzione di quest'ultima nelle equazioni che la precedono, fino alla prima. Perciò questo metodo è detto sostituzione all'indietro. In generale, essendo il sistema

$$Ux = y,$$

la soluzione dell' i -esima equazione è $x_i = \frac{\left(y_i - \sum_{j=i+1}^n u_{ij} x_j \right)}{u_{ii}}$.

Nel caso in cui un sistema lineare non abbia la matrice triangolare, come è meglio procedere? Grazie al metodo di eliminazione di Gauss è possibile ricondurre un sistema qualunque (purché sia quadrato) ad un problema equivalente triangolare, e in particolare ad uno triangolare superiore (risolvibile agevolmente con il metodo sopra descritto).

In pratica si sostituisce una equazione con una combinazione lineare dell'equazione stessa e di un'altra del sistema, ripetendo in procedimento tutte le volte che sia necessario (una equazione alla volta) fino ad ottenere un sistema triangolare.

N.B.: per la buona riuscita del procedimento è necessario partire SEMPRE dalla prima colonna ed andare verso destra, e in ogni colonna procedere dall'alto verso il basso.

Supponiamo che, detta A la matrice dei coefficienti del sistema, l'elemento a_{11} sia non nullo.

Chiamiamo r_1 la prima riga. Alla seconda riga r_2 sostituiamo $r_2 + \left(-\frac{a_{21}}{a_{11}} \right) r_1$. In questo modo si annulla il primo elemento della seconda riga. Proseguendo nello stesso modo fino all'ultima riga, si trasforma la matrice A in una matrice triangolare superiore molto più semplice da trattare. In generale l' i -esima riga r_i

viene sostituita con $r_i + m_{i1} r_1$, $i = 2, \dots, n$ dove $m_{i1} = -\frac{a_{i1}}{a_{11}}$.

Terminato il processo sulla prima riga, si passa alla seconda. Si opera nel medesimo modo, facendo tuttavia riferimento alla riga r_2 .

La riga i -esima r_i viene sostituita con $r_i + r_2 m_{i2}$, $i = 3, 4, \dots, n$ dove $m_{i2} = -\frac{a_{i2}^{(2)}}{a_{22}^{(2)}}$. Con

l'apice (2) si indica che si sta lavorando nel secondo step dell'iterazione, indicando per convenzione con (1) il sistema di partenza.

Si procede in questo modo per tutte le colonne. Ovviamente anche il vettore dei termini noti deve essere trasformato.

Gli elementi $a_{ii}^{(i)}$ sono detti elementi pivot, mentre i termini m_{ij} si chiamano moltiplicatori.

Esempio:

Sia dato il sistema con matrici

$$A = \begin{pmatrix} 4 & 4 & 8 \\ 2 & 8 & 7 \\ 1 & 3 & 6 \end{pmatrix} \quad b = \begin{pmatrix} 12 \\ 9 \\ 7 \end{pmatrix}.$$

Alla seconda riga si somma $-\frac{1}{2}r_1$. La nuova riga diventa $\begin{pmatrix} 0 \\ 6 \\ 3 \end{pmatrix}^T$.

Sommando $-\frac{1}{4}r_1$ alla terza riga si ottiene $\begin{pmatrix} 0 \\ 2 \\ 4 \end{pmatrix}^T$.

E così via con tutte le righe e tutte le colonne, compresa quella dei termini noti.

Un altro metodo strettamente legato al precedente è la fattorizzazione LU.

Ogni passo dell'eliminazione gaussiana può essere scritto, in notazione matriciale, come

$$A^{(k+1)} = M_k A^{(k)}.$$

Perciò il passo n -esimo è dato da $A^{(n)} = M_{n-1} A^{(n-1)} = M_{n-1} M_{n-2} \cdots M_2 M_1 A = MA$ e, allo stesso modo, $b^{(n)} = Mb$.

Il sistema lineare di partenza è $Ax = b$. Sostituendo il sistema equivalente si ha $A^{(n)}x = b^{(n)}$ dove $A^{(n)}$ è una matrice triangolare superiore. Quindi, per ciò che si è precedentemente detto, $MAx = Mb$. Chiamando $U = MA$, si ha $Ux = Mb$.

La matrice M è triangolare inferiore, ma non è semplice da scrivere, in quanto i termini non nulli non sono ordinati. Sappiamo però che è sicuramente invertibile, perché è data dal prodotto delle matrici M_k che sono tutte invertibili, e il risultato di un prodotto di matrici invertibili è ancora invertibile. Poniamo $L = M^{-1}$.

Siccome $U = MA$, allora $LU = LMA = M^{-1}MA = IA = A$. Ossia il sistema iniziale è equivalente al seguente sistema

$$LUx = b$$

Si deduce che si è trovata una fattorizzazione della matrice A :

$$A = LU$$

in cui L è una matrice triangolare inferiore e U è una matrice triangolare superiore.

Inoltre per conoscere la matrice L non occorre conoscere M , infatti sulla diagonale ha tutti 1 e, al di fuori di essa, i termini non nulli sono pari agli m_{ij} , calcolati nel metodo di eliminazione gaussiana, cambiati di segno.

Se chiamiamo $y = Ux$, otteniamo il sistema $Ly = b$ che si riconduce al caso di sistema triangolare inferiore. Cioè y è la soluzione del sistema $Ly = b$ e, successivamente, è possibile calcolare x risolvendo il sistema $Ux = y$.

Questo metodo permette la risoluzione di problemi che presentano una sola matrice dei coefficienti e diversi vettori dei termini noti. Ad esempio i problemi fisici, in cui la matrice dei coefficienti rappresenta il modello matematico utilizzato per modellare la realtà, mentre i vettori dei termini noti sono i diversi ingressi che potrei avere. La possibilità di “riutilizzare” una parte dei calcoli quando i cambiano solamente i dati in ingresso, permette di ridurre il costo computazionale del problema.

Con costo computazionale si intende l'ordine di grandezza del numero di operazioni da effettuare calcolato in funzione della dimensione n del problema.

Il costo computazionale di un sistema risolto con fattorizzazione LU è dell'ordine di $\frac{n^3}{3}$ moltiplicazioni, mentre quello di un sistema triangolare è dell'ordine di $\frac{n^2}{2}$ moltiplicazioni. Se n è molto grande si vede che i metodi di fattorizzazione sono particolarmente indicati.

Esempio:

Si consideri la seguente matrice dei coefficienti:

$$A = \begin{pmatrix} 4 & 4 & 8 \\ 2 & 8 & 7 \\ 1 & 3 & 6 \end{pmatrix}.$$

Le matrici L e U risultano

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 1/4 & 1/3 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 4 & 4 & 8 \\ 0 & 6 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

e si può verificare che $LU = A$.

Inoltre il determinante della matrice dei coefficienti, per il teorema di Binet, risulta

$$\det(A) = \det(L)\det(U) = 48$$

È sempre possibile eseguire una eliminazione gaussiana o una fattorizzazione LU ?

I termini definiti come moltiplicatori nel metodo di eliminazione gaussiana portano, nella definizione, a denominatore un elemento pivot. Se uno di questi elementi pivot è nullo, il procedimento si blocca e non si riesce a risolvere il sistema, anche se esiste una soluzione.

Un teorema, dimostrabile, di esistenza afferma infatti che se una matrice A di ordine n ha tutti i minori principali di testa invertibili, allora esiste, ed è unica, una fattorizzazione LU di A .

Le matrici a diagonale dominante, per righe o per colonne, e le matrici simmetriche soddisfano le ipotesi del teorema precedente.

Si dicono matrici simmetriche quelle per cui vale $A = A^T$. Ad esempio:

$$\begin{pmatrix} 4 & 3 & 1 \\ 3 & 2 & 0 \\ 1 & 0 & 7 \end{pmatrix}.$$

Si dicono matrici simmetriche definite positive se sono simmetriche e soddisfano l'ulteriore relazione

$$\forall x \neq 0 \quad x^T A x > 0.$$

Questa relazione ha significato in quanto, essendo x un vettore, la quantità $x^T A x$ risulta essere uno scalare.

Si dimostra che una condizione equivalente che, se verificata, ci assicura che la matrice con cui stiamo lavorando sia simmetrica e definita positiva è che tutti i suoi autovalori siano strettamente positivi. In formule $\lambda_i(A) > 0 \quad \forall i$.

N.B.: le matrici simmetriche hanno sempre autovalori reali (teorema spettrale).

Se si richiede che la matrice sia semidefinita positiva significa che si impongono le seguenti disuguaglianze nelle precedenti formule

$$\begin{aligned} x^T A x &\geq 0, \quad \forall x \neq 0 \\ \lambda_i(A) &\geq 0, \quad \forall i. \end{aligned}$$

Si dicono matrici a predominanza diagonale per righe le matrici che presentano sulla diagonale, per ogni riga, un elemento di valore assoluto maggiore della somma dei valori assoluti dei restanti elementi:

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad \forall i.$$

Se si ragiona in modo del tutto analogo sulle colonne, si può determinare se una matrice è a predominanza diagonale per colonne.

Esempio:

La seguente matrice

$$A = \begin{pmatrix} 14 & -3 & 1 \\ 13 & 22 & 0 \\ -1 & 0 & -7 \end{pmatrix}$$

ha predominanza diagonale per righe, ma non per colonne.

Il teorema di esistenza della fattorizzazione LU afferma inoltre che esiste una fattorizzazione LU della matrice A anche se quest'ultima è singolare. In questo caso la matrice U risulterà singolare e il sistema non potrà essere risolto.

Un ultimo aspetto, per ciò che riguarda la risoluzione di sistemi lineari con metodi diretti, riguarda il pivoting.

Vogliamo risolvere un sistema la cui matrice dei coefficienti è $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Per poter applicare un qualunque metodo diretto devo prima applicare una matrice di permutazione P_{ij} .

Una matrice di permutazione semplice P_{ij} è ottenuta a partire dalla matrice identità scambiandone le righe i e j .

Il prodotto $P_{ij}A$ scambia le righe i e j di A .

Il prodotto AP_{ij} scambia le colonne i e j di A .

Il prodotto di più matrici di permutazione semplice genera una matrice di permutazione il cui effetto su A è quello di scambiare tutte le righe e le colonne associate alle matrici di permutazione semplice di partenza.

Esempio:

$$A = \begin{pmatrix} 14 & -3 & 1 \\ 13 & 22 & 0 \\ -1 & 0 & 7 \end{pmatrix}$$

$$PA = P_{23}P_{13}A = \begin{pmatrix} -1 & 0 & -7 \\ 14 & -3 & 1 \\ 13 & 22 & 0 \end{pmatrix}, \text{ dove } P = P_{23}P_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Teorema: data una matrice A qualunque di ordine n , esiste una matrice di permutazione P tale per cui è possibile ottenere una fattorizzazione LU della matrice PA :

$$PA = LU .$$

Per trovare una di queste matrici di permutazione P è possibile applicare moltissime strategie. Una di queste è il pivoting parziale: se durante l'applicazione del metodo di eliminazione di Gauss si ottiene un elemento pivot nullo, è possibile scambiare due righe tra quelle al di sotto della riga con elemento pivot nullo in modo tale da poter proseguire con l'eliminazione gaussiana.

Quanto vale il determinante di una matrice di permutazione P ? Se S è il numero di permutazioni semplici eseguite per ottenere P , allora

$$\det(P) = (-1)^S .$$

Dimostrazione: il k -esimo passo dell'eliminazione gaussiana, applicando il pivoting parziale, risulta $A^{(k+1)} = M_k P_k A^{(k)}$.

Quindi $A^{(n)} = M_{n-1} P_{n-1} M_{n-2} P_{n-2} \dots M_2 P_2 M_1 P_1 A$ e, di conseguenza, $U = M_{n-1} P_{n-1} \dots M_2 P_2 M_1 P_1 A$, ossia si trova una fattorizzazione LU della matrice PA , infatti modificando opportunamente le matrici M è possibile raggruppare le matrici P_i ed M_i e scrivere $U = MPA$. Da cui $M^{-1}U = PA$ e, ponendo $L = M^{-1}$ si ottiene il risultato cercato: $LU = PA$.

c.v.d.

Applicazione del metodo: dato un sistema $Ax = b$ si applica prima la permutazione ai due membri dell'equazione ottenendo $PAx = Pb$. Si risolve successivamente questo nuovo sistema con la solita tecnica:

1. $Ly = Pb$, sistema triangolare inferiore;
2. $Ux = y$, sistema triangolare superiore.

La strategia del pivoting può anche essere applicata per aumentare la stabilità dell'eliminazione gaussiana in quanto è utilizzabile per fare in modo di avere gli elementi pivot più grandi possibile. In questo modo i moltiplicatori diventano più piccoli e ciò aumenta la stabilità del metodo.

N.B.: se scelgo il pivot massimo, ottengo moltiplicatori tutti minori o uguali all'unità.

Se la matrice è a predominanza diagonale per colonne, il pivoting non apporta alcuna modifica ad essa, in quanto non si effettua alcuno scambio.

Se la matrice è simmetrica e definita positiva, l'eliminazione gaussiana è già stabile ed è quindi più conveniente utilizzare la fattorizzazione di Cholesky che sfrutta la simmetria della matrice.

Come costruire L ?

Esempio:

$$\text{Sia } PA = LU \text{ dove } A = \begin{pmatrix} 1 & 4 & 8 \\ 2 & 0 & 7 \\ 4 & 2 & 6 \end{pmatrix}.$$

Prima permutazione:

$$P_1 = P_{13}$$

$$P_{13}A = \begin{pmatrix} 4 & 2 & 6 \\ 2 & 0 & 7 \\ 1 & 4 & 8 \end{pmatrix}$$

$$A^{(2)} = \begin{pmatrix} 4 & 2 & 6 \\ \frac{1}{2} & -1 & 4 \\ \frac{1}{4} & \frac{7}{2} & \frac{13}{2} \end{pmatrix} \text{ dove } \frac{1}{2} \text{ e } \frac{1}{4} \text{ sono i moltiplicatori cambiati di segno.}$$

Seconda permutazione:

$$P_2 = P_{23}$$

$$P_{23}A^{(2)} = \begin{pmatrix} 4 & 2 & 6 \\ \frac{1}{4} & \frac{7}{2} & \frac{13}{2} \\ \frac{1}{2} & -1 & 4 \end{pmatrix}$$

$$A^{(3)} = \begin{pmatrix} 4 & 2 & 6 \\ \frac{1}{4} & \frac{7}{2} & \frac{13}{2} \\ \frac{1}{2} & -\frac{2}{7} & \frac{41}{7} \end{pmatrix} \text{ dove } \frac{1}{4}, \frac{1}{2} \text{ e } -\frac{2}{7} \text{ sono i moltiplicatori cambiati di segno.}$$

Quindi:

$$U = \begin{pmatrix} 4 & 2 & 6 \\ 0 & \frac{7}{2} & \frac{13}{2} \\ 0 & 0 & \frac{41}{7} \end{pmatrix} \text{ e } L = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & 1 & 0 \\ \frac{1}{2} & -\frac{2}{7} & 1 \end{pmatrix}$$

La matrice di permutazione risulta $P = P_{23}P_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

La fattorizzazione di Cholesky, citata in precedenza, si applica se A è quadrata di ordine n , simmetrica e definita positiva. Si dimostra che esiste una e una sola matrice triangolare superiore R con gli elementi sulla diagonale positivi tale che

$$A = R^T R$$

dove R^T è triangolare inferiore. Il costo computazionale di questa fattorizzazione è dell'ordine di $\frac{n^3}{6}$ moltiplicazioni, ed è quindi più conveniente della fattorizzazione LU .

Infine un altro tipo di fattorizzazione, tuttavia con costo computazionale decisamente più elevato della fattorizzazione LU , è la fattorizzazione QR. Questo metodo permette di scrivere la matrice A come il prodotto di una matrice R triangolare superiore e di una matrice Q ortogonale.

Si dice ortogonale una matrice per cui vale $A^{-1} = A^T$.

Metodi iterativi

I metodi diretti sono molto efficienti se si trattano matrici di piccole dimensioni e dense, cioè con pochi elementi nulli.

Se la matrice di un sistema è sparsa ed è di ordine elevato, allora i metodi diretti sono pressoché inapplicabili. Infatti il fenomeno del fill in rischia di far sparire buona parte degli elementi nulli. E quindi il sistema si complica grandemente.

I metodi iterativi non modificano la matrice di partenza A , ma generano una successione infinita di vettori $x^{(k)}$ che converge alla soluzione del sistema lineare per $k \rightarrow \infty$, sotto opportune ipotesi.

Una successione di vettori $\{x^{(k)}\}_{k \geq 0} \in \mathfrak{R}^n$ converge ad un vettore $x \in \mathfrak{R}^n$ se esiste una qualunque norma per cui

$$\lim_{k \rightarrow \infty} \|x^{(k)} - x\| = 0$$

e in tal caso si pone $\lim_{k \rightarrow \infty} x^{(k)} = x$.

Si noti che ci si riconduce allo studio della convergenza di una successione di numeri, in quanto la funzione norma restituisce un numero.

Questo ragionamento vale per qualunque norma venga considerata perché tutte le norme sono equivalenti: se una successione converge in una norma, converge anche con tutte le altre; analogamente, se una successione non converge in una norma, allora non converge con nessun'altra.

Si può tradurre la condizione di convergenza anche in componenti: ogni componente i -esima converge alla componente i del limite della successione. In formule

$$\lim_{k \rightarrow \infty} x_i^{(k)} = x_i, \quad \forall i = 1, 2, \dots, n.$$

La stessa cosa si può dire per le successioni di matrici.

Teorema: sia A una matrice quadrata e $\rho(A)$ il suo raggio spettrale. Allora

$$\lim_{k \rightarrow \infty} A^k = 0 \Leftrightarrow \rho(A) < 1$$

dove A^k è una particolare successione di matrici i cui termini sono matrici elevate a k .

Osservazione: per ogni norma di matrice compatibile con una norma di vettori, in particolar modo per le norme naturali, si ha $\rho(A) \leq \|A\|$.

Che cos'è un metodo iterativo?

Data una stima iniziale $x^{(0)}$ della soluzione del sistema lineare $Ax = b$ ben posto, si costruisce una successione di vettori $\{x^{(k)}\}_{k \geq 0}$ che converga a x risolvendo, ad ogni passo, dei sistemi lineari molto più semplici di quello di partenza.

Come si costruisce la successione?

Detta A la matrice del sistema, la possiamo dividere in almeno due componenti con un processo di splitting. Supponiamo di dividerla in due fattori: $A = M + N$. Allora:

$$Ax = (M + N)x = b \Rightarrow Mx = -Nx + b$$

ipotizzando di essere all'iterazione $k+1$, conoscendo perciò $x^{(k)}$ e desiderando calcolare $x^{(k+1)}$, possiamo porre

$$Mx^{(k+1)} = -Nx^{(k)} + b.$$

Affinché questo sistema sia ben posto la matrice M deve essere invertibile, ossia deve essere vero che $\det(M) \neq 0$.

Questo approccio è utile se M è molto semplice, ad esempio diagonale o triangolare.

Abbiamo quindi

$$x^{(k+1)} = -M^{-1}Nx^{(k)} + M^{-1}b,$$

chiamiamo $B = -M^{-1}N$ e $c = M^{-1}b$

$$x^{(k+1)} = Bx^{(k)} + c.$$

A partire da questa relazione possiamo costruire la successione $\{x^{(k)}\}_{k \geq 0}$:

$$x^{(1)} = Bx^{(0)} + c$$

$$x^{(2)} = Bx^{(1)} + c$$

$$x^{(3)} = Bx^{(2)} + c$$

...

dove la matrice B è chiamata matrice di iterazione ed è l'unica responsabile della convergenza del metodo. In pratica, un metodo iterativo è convergente a seconda di come è fatta la matrice di iterazione.

Proprietà di consistenza: per i sistemi lineari, se si costruisce con un metodo iterativo una successione convergente, allora il limite della successione è una soluzione del sistema lineare considerato.

Ovviamente la successione dipende dalla stima $x^{(0)}$ della soluzione. Per i sistemi lineari si dimostra che la successione converge, o meno, indipendentemente dalla scelta della stima iniziale.

Un metodo iterativo si dice convergente se la successione generata converge indipendentemente da $x^{(0)}$.

Come stabilire se $\{x^{(k)}\}_{k \geq 0}$ converge?

Definiamo errore commesso al passo k la differenza

$$e^{(k)} = x^{(k)} - x$$

e se $x^{(k)} \rightarrow x$, allora $e^{(k)} \rightarrow 0$. Queste due condizioni sono equivalenti.

Sostituendo

$$\begin{aligned} e^{(k)} &= x^{(k)} - x = Bx^{(k-1)} + c - x = Bx^{(k-1)} + c - (Bx + c) = B(x^{(k-1)} - x) = Be^{(k-1)} \Rightarrow \\ \Rightarrow e^{(k)} &= Be^{(k-1)} = B(Be^{(k-2)}) = B^2 e^{(k-2)} = \dots = B^k e^{(0)} \end{aligned}$$

e, riassumendo, si ottiene $e^{(k)} = B^k e^{(0)}$. La successione converge se $B^k e^{(0)} \rightarrow 0$: fissato un $x^{(0)}$

$$\lim_{k \rightarrow \infty} x^{(k)} = x \Leftrightarrow \lim_{k \rightarrow \infty} e^{(k)} = 0 \Leftrightarrow \lim_{k \rightarrow \infty} B^k e^{(0)} = 0 \Leftrightarrow \lim_{k \rightarrow \infty} B^k = 0$$

e cioè la successione converge se e solo se $\rho(B) < 1$, indipendentemente dalla scelta di $x^{(0)}$.

Teorema: un metodo iterativo definito dalla matrice B converge $\forall x^{(0)} \in \mathfrak{R}^n$ se e solo se $\rho(B) < 1$ (condizione necessaria e sufficiente).

Questo teorema non è facilmente dimostrabile: bisogna infatti conoscere $B = -M^{-1}N$ e calcolarne il raggio spettrale, e questo non è generalmente semplice.

Una condizione sufficiente di convergenza è: sia $\|B\|$ una norma di matrice naturale. Se $\|B\| < 1$, allora il metodo iterativo converge perché $\rho(B) \leq \|B\|$. Se $\|B\| > 1$ non è possibile affermare nulla sul raggio spettrale.

Il raggio spettrale dà anche informazioni sulla velocità di convergenza della successione, cioè su quanto velocemente l'errore tende a zero. Più il raggio spettrale è prossimo a zero, maggiore è la velocità di convergenza.

N.B.: $0 \leq \rho(B) < 1$.

Uno dei due metodi iterativi più semplici è il metodo di Jacobi o metodo delle sostituzioni simultanee.

Dividiamo la matrice A in tre parti:

- E = la parte di A sotto la diagonale;
- D = la diagonale di A ;
- F = la parte di A sopra la diagonale.

Poniamo $M = D$ e $N = E + F$. Quindi $B_j = -D^{-1}(E + F)$ è la matrice di iterazione associata a questo metodo.

Il metodo di Jacobi è applicabile solamente se la matrice D è invertibile, ossia se la diagonale di A non contiene elementi nulli.

Scriviamo l'algoritmo componente per componente:

$$Mx^{(k+1)} = -Nx^{(k)} + b \Leftrightarrow Dx^{(k+1)} = -(E + F)x^{(k)} + b$$

$$x_i^{(k+1)} = \frac{\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} \right)}{a_{ii}}$$

in forma compatta $x_i^{(k+1)} = D^{-1}(b - Ex^{(k)} - Fx^{(k)})$.

Esempio:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \quad b = \begin{pmatrix} 2 \\ 5 \\ 9 \end{pmatrix}, \quad x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 9 \end{pmatrix} \text{ e } (E + F) = \begin{pmatrix} 0 & 2 & 3 \\ 4 & 0 & 6 \\ 7 & 8 & 0 \end{pmatrix}$$

Calcoliamo $x^{(1)}$:

$$Dx^{(1)} = -(E + F)x^{(0)} + b$$

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 2 \\ 4x_1 + 5x_2 + 6x_3 = 5 \\ 7x_1 + 8x_2 + 9x_3 = 9 \end{cases}$$

$$\begin{cases} x_1^{(k+1)} = \frac{2 - 2x_2^{(k)} - 3x_3^{(k)}}{1} \\ x_2^{(k+1)} = \frac{5 - 4x_1^{(k)} - 6x_3^{(k)}}{5} \\ x_3^{(k+1)} = \frac{9 - 7x_1^{(k)} - 8x_2^{(k)}}{9} \end{cases}$$

$$\begin{cases} x_1^{(1)} = 2 \\ x_2^{(1)} = 1 \\ x_3^{(1)} = 1 \end{cases} \Rightarrow x^{(1)} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

Calcoliamo $x^{(2)}$:

$$\begin{cases} x_1^{(2)} = \frac{2 - 2 - 3}{1} = -3 \\ x_2^{(2)} = \frac{5 - 8 - 6}{5} = -\frac{9}{5} \\ x_3^{(2)} = \frac{9 - 14 - 8}{9} = -\frac{13}{9} \end{cases} \Rightarrow x^{(2)} = \begin{pmatrix} -3 \\ -\frac{9}{5} \\ -\frac{13}{9} \end{pmatrix}$$

Questo metodo non converge.

Se almeno un elemento sulla diagonale di A è nullo, il metodo di Jacobi non può essere applicato. È comunque sempre possibile riordinare le equazioni del sistema in modo tale da avere sulla diagonale tutti elementi non nulli ($a_{ii} \neq 0 \forall i$), se la matrice A è non singolare.

L'altro metodo iterativo particolarmente semplice è quello di Gauss-Seidel. Partizionando la matrice A nello stesso modo del metodo di Jacobi, si definiscono diversamente le matrici M e N . In questo caso

$$M = E + D, \quad N = F$$

e quindi la matrice di iterazione risulta $B_{GS} = -(E + D)^{-1}F$.

In componenti $\forall i = 1, 2, \dots, n$ vale:

$$x_i^{(k+1)} = \frac{\left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij} x_j^{(k)} \right)}{a_{ii}}.$$

A mano a mano che si calcolano le nuove approssimazioni delle componenti, le si utilizzano per calcolare le approssimazioni delle componenti restanti.

In forma compatta $x^{(k+1)} = D^{-1}(b - Ex^{(k+1)} - Fx^{(k)})$.

I metodi fin qui descritti sono metodi iterativi che convergono per $k \rightarrow \infty$. È chiaro che non si arriverà mai fino all'infinito per risolvere un sistema lineare. Ci si pone quindi il problema di capire quando è possibile fermarsi. È il cosiddetto test di arresto.

È necessario stabilire un criterio di arresto delle iterazioni.

Per fare questo si ha bisogno di una stima dell'errore e bloccare il processo non appena si soddisfa una delle due seguenti condizioni.

1. controllo sulla distanza relativa fra due iterate successive: definita una certa tolleranza si blocca il processo quando $\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k+1)}\|} \leq \text{tolleranza}$. In

modo analogo, ma sconsigliato, si può operare con la distanza assoluta, fermando le iterazioni quando $\|x^{(k+1)} - x^{(k)}\| \leq \text{tolleranza}$.

2. controllo sul residuo dell'equazione: si definisce residuo l'espressione $r^{(k+1)} = b - Ax^{(k+1)}$. Si smette di iterare quando, definita una certa tolleranza, accade che $\frac{\|b - Ax^{(k+1)}\|}{\|b\|} \leq \text{tolleranza}$.

La tolleranza viene scelta in base a ragionamenti su tempo a disposizione e precisione del risultato desiderata. Non è possibile scegliere come tolleranza valori troppo vicini alla precisione di macchina, ad esempio se la precisione di macchina è 10^{-16} , non si possono imporre tolleranze minori di 10^{-12} .

N.B.: in generale un residuo piccolo, sotto la soglia di tolleranza, non implica automaticamente un errore piccolo.

Infatti:

$$\|e^{(k)}\| = \|x^{(k)} - x\| = \|x^{(k)} - A^{-1}b\| = \|A^{-1}(Ax^{(k)} - b)\| = \|-A^{-1}(Ax^{(k)} - b)\| = \|-A^{-1}r^{(k)}\| \leq \|A^{-1}\| \|r^{(k)}\|$$

e dato che $\frac{1}{\|x\|} \leq \frac{\|A\|}{\|b\|}$ si ha che:

$$\frac{\|e^{(k)}\|}{\|x\|} \leq \|A^{-1}\| \|r^{(k)}\| \frac{\|A\|}{\|b\|} = K(A) \frac{\|r^{(k)}\|}{\|b\|}$$

dove $\frac{\|e^{(k)}\|}{\|x\|}$ è l'errore relativo, $\frac{\|r^{(k)}\|}{\|b\|}$ è il residuo relativo e $K(A)$ è il numero di condizionamento di A .

Quindi $\frac{\|r^{(k)}\|}{\|b\|} \leq \text{tolleranza} \Rightarrow \frac{\|e^{(k)}\|}{\|x\|} \leq K(A) \frac{\|r^{(k)}\|}{\|b\|} \leq K(A) \cdot \text{tolleranza}$.

Determiniamo ora le condizioni di convergenza per i metodi di Jacobi e di Gauss-Seidel. Sono sostanzialmente dei teoremi che risultano più spendibili della definizione generale di convergenza.

In caso di matrici a predominanza diagonale, per righe o per colonne, i metodi di Jacobi e di Gauss-Seidel sono convergenti. La matrice da considerare è A , cioè quella del sistema. Questa è una condizione sufficiente, infatti se non c'è predominanza diagonale non si può dire nulla sulla convergenza.

In caso di matrici simmetriche definite positive il metodo di Gauss-Seidel è convergente.

In generale se uno dei due metodi è convergente, non è detto che anche l'altro converga. L'unica cosa che si può dire è che, se entrambi i metodi convergono, allora il metodo di Gauss-Seidel converge più velocemente di quello di Jacobi.

Questi risultati sono riassunti dai seguenti teoremi.

Teorema di Stein-Rosenberg: sia $A \in \mathfrak{R}^{n \times n}$ con $a_{ij} \leq 0 \quad \forall i \neq j$ e $a_{ii} > 0$. Allora si verifica uno e uno solo dei seguenti risultati:

1. $0 < \rho(B_{GS}) < \rho(B_J) < 1$, entrambi convergono e Gauss-Seidel è migliore di Jacobi;
2. $1 < \rho(B_J) < \rho(B_{GS})$, entrambi divergono;
3. $\rho(B_{GS}) = \rho(B_J) = 0$, entrambi convergono alla medesima velocità;
4. $\rho(B_{GS}) = \rho(B_J) = 1$, entrambi divergono.

Teorema: sia A una matrice tridiagonale con elementi sulla diagonale non nulli, allora vale $\rho(B_{GS}) = \rho^2(B_J)$ e cioè i due metodi convergono, o divergono, simultaneamente.

Questi due teoremi si usano in pratica per capire se il metodo di Gauss-Seidel converge, o meno, in caso di matrice A non semplice. Si ragiona sul metodo di Jacobi e poi, applicando i teoremi, si deduce il risultato su Gauss-Seidel.