RISOLUZIONE DI SISTEMI LINEARI: METODI ITERATIVI (Il numero di condizionamento: stabilità della risoluzione di un sistema lineare)

In questa lezione verrà affrontato il problema della stabilità nella risoluzione di un sistema lineare. In particolare, ci si concentrerà sull'analisi dell'errore che può insorgere nella soluzione a seguito di variazioni nei parametri del problema, ovvero nei coefficienti della matrice del sistema e nel termine noto. Successivamente, verrà introdotto il concetto di condizionamento di un problema numerico e, contestualmente, si accennerà alla nozione di stabilità, ponendo particolare attenzione alla relazione tra la stabilità e il numero di condizionamento del problema.

L'obiettivo è quindi condurre un'analisi della stabilità di un metodo numerico, studiando in particolare la sensibilità della soluzione del sistema lineare rispetto a perturbazioni nei dati del problema. Vogliamo quindi studiare la sensibilità della soluzione del sistema lineare

$$A\boldsymbol{x} = b$$

rispetto alle perturbazioni dei dati A e $\boldsymbol{b}$. A tal proposito consideriamo una matrice perturbata

$$A^* = A + \delta A,$$

dove δA è una matrice di perturbazione, ovvero a una matrice i cui coefficienti δA_{ij} daranno una misura della perturbazione dei coefficienti A_{ij} della matrice A. Analogamente definiamo un vettore termine noto perturbato

$$\boldsymbol{b}^* = \boldsymbol{b} + \delta\boldsymbol{b},$$

dove le componenti del vettore perturbazione $\delta\boldsymbol{b}$ andranno a perturbare le corrispondenti componenti del vettore termine noto $\boldsymbol{b}$.

Esistono diversi modi e situazioni in cui il problema può risultare effettivamente perturbato. Ad esempio, anche nel caso in cui le matrici A e il vettore $\boldsymbol{b}$ siano conosciuti con esattezza, è evidente che la loro rappresentazione in un calcolatore (ossia in termini di quelli che vengono chiamati numeri di macchina) può comportare l'introduzione di errori di arrotondamento. Di conseguenza, ci troveremo inevitabilmente ad affrontare coefficienti A_{ij} e b_i perturbati rispetto ai loro valori reali.

Un numero macchina è un numero rappresentato all'interno di una macchina di calcolo (o calcolatore) e costituisce un'approssimazione del numero esatto con cui teoricamente si dovrebbero effettuare le operazioni. L'approssimazione può essere realizzata tramite metodi di arrotondamento o di troncamento. Durante la programmazione di un algoritmo, è fondamentale distinguere tra due insiemi numerici: quello teorico o ideale con cui si opera (ad esempio, l'insieme dei numeri reali $\mathbb{R}$) e l'insieme dei numeri macchina $\mathbb{F}$, che tiene conto dell'ambiente di calcolo effettivo in cui si opera. Il simbolo $\mathbb{F}$ deriva dall'espressione inglese "Floating point", che significa virgola mobile. Le operazioni che è possibile eseguire con i numeri macchina e le loro proprietà definiscono un ambiente operativo con caratteristiche differenti rispetto agli insiemi numerici fondamentali e alle operazioni in essi definite. Questo ambiente, noto come aritmetica finita di macchina, dipende dalla rappresentazione finita dei numeri e dai conseguenti errori introdotti dall'uso di numeri approssimati. Vedremo in dettaglio le operazioni con i numeri macchina in un nucleo dedicato. Per ora ci basti osservare che, la propagazione

degli errori può talvolta portare a risultati inattendibili o a un'instabilità non controllabile della soluzione.
Naturalmente, esistono molte altre situazioni in cui il problema è perturbato fin dall'origine, ossia quando i coefficienti di A e $\boldsymbol{b}$ non sono noti esattamente, poiché, ad esempio, derivano da misure sperimentali, le quali sono a loro volta soggette a errori. In questi casi, ci troveremo con una matrice e un vettore di termini noti entrambi perturbati.
In definitiva, possiamo aspettarci che in corrispondenza di questi dati perturbati, la soluzione $\boldsymbol{x}$ (del sistema lineare $A\boldsymbol{x} = b$) cambi. Quindi anziché calcolare la soluzione esatta $\boldsymbol{x}$ del sistema lineare, calcoleremo la soluzione perturbata $\boldsymbol{x}^*$, che rappresenteremo come

$$\boldsymbol{x}^* = \boldsymbol{x} + \delta\boldsymbol{x},$$

dove $\delta\boldsymbol{x}$ (incognito) rappresenta la misura della perturbazione del vettore soluzione.
Possiamo quindi pensare che il sistema che di fatto si risolve è

$$(A + \delta A)(\boldsymbol{x} + \delta\boldsymbol{x}) = \boldsymbol{b} + \delta\boldsymbol{b},$$

ovvero, un nuovo sistema la cui matrice è $A + \delta A$ e il termine noto è $\boldsymbol{b} + \delta\boldsymbol{b}$. Di conseguenza, la soluzione del nuovo sistema sarà $\boldsymbol{x} + \delta\boldsymbol{x}$.
Il problema che ci poniamo è quello di stimare $\delta\boldsymbol{x}$ in funzione di δA e $\delta\boldsymbol{b}$, ovvero trovare una maggiorazione dell'errore commesso sulla soluzione in relazione all'errore sui dati. Per misurare questi errori, ricorreremo alle norme di vettori e di matrici.
Supponiamo quindi di avere una norma vettoriale $\|\boldsymbol{a}\|$, e la corrispondente norma matriciale indotta dalla norma del vettore, denotata come $\|A\|$. Per ogni matrice A, invertibile (ovvero per cui è possibile calcolare A^{-1}), definiamo un numero reale chiamato numero di condizionamento della matrice della matrice A

$$K(A) = \|A\|\|A^{-1}\|.$$

Facciamo una breve considerazione sulla natura del numero di condizionamento e sulle sue proprietà.
In primo luogo, il numero di condizionamento dipende dalla norma scelta; pertanto, se si cambia la norma, il numero di condizionamento varia di conseguenza. Inoltre, è importante notare che il numero di condizionamento è sempre maggiore o uguale a uno

$$K(A) \geq 1.$$

Infatti, se consideriamo la matrice identità I (che ha tutti gli elementi uguali a 1 sulla diagonale principale e zero altrove), possiamo osservare che la norma dell'identità è pari a 1. Indipendentemente dalla norma scelta (euclidea, infinita o qualsiasi altra norma), la norma della matrice identità risulta sempre essere uguale a 1. Inoltre, la matrice identità può essere espressa come il prodotto di A per A^{-1}.
Utilizzando una delle proprietà delle norme, che afferma che la norma del prodotto è minore o uguale al prodotto delle norme, otteniamo esattamente il numero di condizionamento $K(A)$

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

È importante fare attenzione al fatto che il numero di condizionamento di una matrice non ha necessariamente alcuna relazione diretta con il determinante della matrice stessa. In altre parole, un determinante grande non implica necessariamente un numero di condizionamento grande, e analogamente, un determinante piccolo non corrisponde necessariamente a un numero di condizionamento elevato. Determinante e numero di condizionamento sono, infatti, due indicatori distinti della matrice.
Consideriamo ora la norma 2 (quindi quella associata alla norma euclidea) della matrice A e definiamo il numero di condizionamento euclideo come

$$K_2(A) = \|A\|_2\|A^{-1}\|_2,$$

Osserviamo che, se la matrice A è simmetrica e definita positiva (SDP)
il numero di condizionamento euclideo $K_2(A)$ ha una forma particolarmente semplice. Si può infatti dimostrare che, se A è SDP allora

$$\|A\|_2 = \lambda_{max}(A), \qquad \|A^{-1}\|_2 = \frac{1}{\lambda_{min}(A)},$$

ovvero, la norma due di A coincide con il massimo autovalore della matrice A. Inoltre, la norma due di A^{-1} è caratterizzata con il minimo autovalore della matrice A. Da cui, deriva che

$$K_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)}.$$

Questa relazione garantisce che il numero di condizionamento rispetto alla norma 2 di una matrice A SDP è pari al rapporto tra il massimo e il minimo autovalore di A. Vedremo che in molte circostanze, questa caratterizzazione si rivelerà utile per analizzare la convergenza di metodi iterativi nella risoluzione dei sistemi lineari.
Vogliamo quindi caratterizzare gli errori sulle soluzioni in funzione degli errori sui dati. A tal fine, utilizzeremo lo strumento del numero di condizionamento della matrice. È possibile dimostrare, ma non ne daremo esplicita dimostrazione, che

$$\frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leq \frac{K(A)}{1 - K(A)\frac{\|\delta A\|}{\|A\|}} \left(\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\delta A\|}{\|A\|} \right),$$

dove $\delta \boldsymbol{x}$ è la perturbazione sulla soluzione del sistema. Tale relazione ci indica che l'errore relativo commesso sulla soluzione è minore o uguale a una certa costante

$$\frac{K(A)}{1 - K(A)\frac{\|\delta A\|}{\|A\|}}$$

moltiplicata per una quantità che racchiude l'errore relativo sul vettore del termine noto $\boldsymbol{b}$ e l'errore relativo sulla matrice A. Pertanto, l'errore sulla soluzione è limitato da una costante che dipende dal numero di condizionamento di A, moltiplicata per l'errore complessivo sui dati. Se introduciamo

$$c = K(A) \quad \text{e} \quad \epsilon = \frac{\|\delta A\|}{\|A\|},$$

allora la costante $\frac{K(A)}{1-K(A)\frac{\|\delta A\|}{\|A\|}}$, che chiameremo α è

$$\alpha(c) = \frac{c}{1 - c\epsilon}.$$

Supponendo che $\epsilon = \frac{\|\delta A\|}{\|A\|}$ sia piccola, essendo δA la perturbazione sulla matrice A, possiamo osservare che la curva $\alpha(c)$ è descritta da un grafico del tipo mostrato in Figura 1.
Infatti, il numero di condizionamento c è sempre maggiore uguale ad 1, $c \geq 1$. In particolare, risulta

$$\alpha(1) = \frac{1}{1 - \epsilon} \sim 1,$$

Inoltre, $\frac{1}{\epsilon}$ è un asintoto verticale della curva $\alpha(c)$.
Possiamo osservare che, se il numero di condizionamento è piccolo (dell'ordine dell'unità, ad esempio 1, 2, 3, 4…), il sistema si comporta in modo favorevole, poiché la costante $\alpha(c)$ risulta essere relativamente piccola. Al contrario, se il numero di condizionamento è elevato, la costante $\alpha(c)$ tende a diventare grande, indicando che il sistema è meno stabile e più sensibile agli errori nei dati.

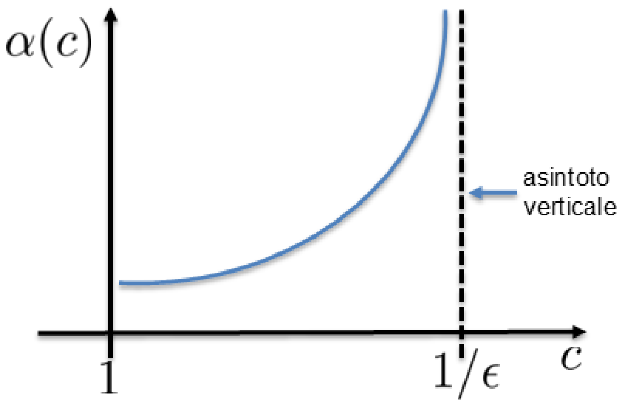

Figura 1: Andamento della funzione $\alpha(c)$ al variare del valore c.

Possiamo dunque concludere che se $K(A)$ è piccolo

$$K(A) \approx 1,$$

si dice che la matrice A è ben condizionata, poiché il numero di condizionamento è relativamente basso. In questo caso, il sistema $A\boldsymbol{x} = \boldsymbol{b}$ è ben condizionato: piccole perturbazioni nei dati determinano piccole variazioni nella soluzione. Questa è esattamente la definizione generale di buon condizionamento di un metodo numerico. Viceversa, se $K(A)$ è molto grande

$$K(A) \gg 1,$$

allora la matrice A è mal condizionata e il sistema associato $A\boldsymbol{x} = \boldsymbol{b}$, è mal condizionato: piccole perturbazioni sui dati A e $\boldsymbol{b}$ possono produrre grandi errori sulla soluzione $\boldsymbol{x}$.

In definitiva abbiamo osservato che, per matrici mal condizionate ($K(A) \gg 1$), la risoluzione del sistema lineare può comportare grandi perturbazioni sulla soluzione, anche in seguito a piccole perturbazioni nei dati. Al contrario, se la matrice è ben condizionata ($K(A) \approx 1$), questo fenomeno patologico non si verifica.

Consideriamo un semplice esempio: un sistema lineare 2×2

$$\begin{cases} 7x_1 + 10x_2 = b_1 \\ 5x_1 + 7x_2 = b_2 \end{cases}.$$

Questo è un sistema apparentemente banale, dove la matrice A è

$$A = \begin{pmatrix} 7 & 10 \\ 5 & 7 \end{pmatrix},$$

e la sua inversa può essere facilmente calcolata e risulta

$$A^{-1} = \begin{pmatrix} -7 & 10 \\ 5 & -7 \end{pmatrix}.$$

Se il termine noto ha componenti $b_1 = 1$ e $b_2 = 0.7$, la soluzione corrispondente è $x_1 = 0, x_2 = 0.1$. Se perturbassimo il solo vettore dei termini noti $\boldsymbol{b}$, lasciando inalterata la matrice A, quindi $\delta A = 0$ e $\delta\boldsymbol{b} \neq \boldsymbol{0}$, con

$$b_1 = 1.01, b_2 = 0.69,$$

la soluzione corrispondente sarà

$$x_1 = -0.17, x_2 = 0.22.$$

Osserviamo che abbiamo ottenuto un cambiamento molto drastico nella soluzione in corrispondenza di una perturbazione sui dati che, in realtà, è molto modesta (dell'ordine dell'1% su entrambe le componenti)

$$\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} \approx 1\% \quad \Rightarrow \quad \frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} > 100\%$$

In effetti l'errore relativo sul termine noto è dell'ordine del 1%, mentre l'errore relativo sulla soluzione è superiore al 100%.
Questo esempio mette in evidenza che, in effetti, possono verificarsi grosse perturbazioni sulle soluzioni. Il responsabile di questo comportamento patologico è, in realtà, il condizionamento della matrice A. Si può verificare che il numero di condizionamento di A è

$$K(A) = 289.$$

Un numero dell'ordine di grandezza di 100 funge da moltiplicatore dell'errore sui coefficienti, generando così un errore significativo sulle soluzioni a causa di questa moltiplicazione. Pertanto, questo esempio fornisce una prova euristica alla teoria che abbiamo precedentemente illustrato. Abbiamo osservato che matrici ben condizionate danno origine a sistemi lineari le cui soluzioni sono poco sensibili alle perturbazioni nei dati del problema. Al contrario, matrici mal condizionate possono generare sistemi la cui soluzione dipende in modo estremamente critico dalle perturbazioni nei dati. Di conseguenza, si parla di grande sensibilità ai dati, e il sistema associato si considera mal condizionato.