RISOLUZIONE DI SISTEMI LINEARI: METODI DIRETTI (La Fattorizzazione LU)

In questa lezione analizzeremo come il metodo di eliminazione di Gauss (MEG), presentato nella lezione precedente per la risoluzione dei sistemi lineari come metodo diretto, possa essere interpretato come un metodo di fattorizzazione della matrice A di partenza.
Ricordiamo che risolvere un sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ mediante il MEG equivale a trasformare il sistema in un altro equivalente della forma $U\boldsymbol{x} = \boldsymbol{f}$, dove U è una matrice triangolare superiore.
Ora analizziamo come il MEG corrisponda effettivamente alla fattorizzazione della matrice iniziale A come prodotto di due matrici

$$A = LU,$$

dove L è una matrice triangolare inferiore e U una matrice triangolare superiore. Più precisamente, U corrisponde alla matrice finale del sistema, ovvero la matrice $A^{(n)}$ ottenuta dopo $n-1$ trasformazioni del MEG. La matrice triangolare inferiore L presenta elementi diagonali pari a uno, elementi sopra la diagonale uguali a zero e, nella parte sotto diagonale, contiene i coefficienti $m_{21}, \dots, m_{n1}, \dots, m_{n,n-1}$ che rappresentano i moltiplicatori generati dal MEG

$$L = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ m_{21} & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ m_{n1} & \dots & m_{n,n-1} & 1 \end{pmatrix}.$$

Quindi la matrice L della fattorizzazione è una matrice che contiene, come elementi sotto diagonali i moltiplicatori m_{ik} che il MEG ha generato.
Non è immediato, ed ovvio, dimostrare che la matrice A possa effettivamente essere espressa come il prodotto di L e U, né che L sia definita in questo modo. Forniamo di seguito una traccia per comprendere il processo di costruzione della matrice L. A tale scopo, introduciamo delle matrici di trasformazione, che chiameremo $M^{(k)}$, della forma

$$M^{(k)} = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & & -m_{k+1,k} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & & -m_{n,k} & 0 & \cdots & 1 \end{pmatrix} = I - \boldsymbol{m}_k e_k^T.$$

Per ogni k, la matrice $M^{(k)}$ è una matrice triangolare inferiore con elementi diagonali pari a uno ed elementi sotto la diagonale uguali a zero, ad eccezione della colonna k-esima. La colonna k-esima contiene i moltiplicatori $m_{k+1,k}, \dots, m_{n,k}$, presi con segno opposto. Come si nota, questa matrice $M^{(k)}$ può essere espressa in forma compatta come una perturbazione della matrice identità I, ovvero la matrice con tutti uno sulla diagonale e zeri altrove, a cui viene sottratto il prodotto tra un vettore colonna $\boldsymbol{m}_k$ e un vettore riga (i.e., vettore trasposto) $\boldsymbol{e}_k^T$. Il vettore $\boldsymbol{m}_k$ è un vettore colonna con tutti zeri fino alla posizione $k-1$, mentre le sue componenti a partire dall'indice $k+1$ corrispondono ai moltiplicatori $m_{k+1,k}, \dots, m_{n,k}$

$$\boldsymbol{m}_k = \left(0, \dots, 0, m_{k+1,k}, \dots, m_{n,k}\right)^T.$$

Il vettore $\boldsymbol{e}_k^T$ è un vettore riga le cui componenti sono tutte uguali a zero, tranne la componente k-esima, che è pari a 1. Esso è anche definito come il vettore della base canonica di $\mathbb{R}^n$. Si osservi che, essendo il risultato del prodotto tra un vettore colonna $\boldsymbol{m}_k$ e un vettore riga $\boldsymbol{e}_k^T$, si ottiene una matrice di dimensioni $n \times n$, ovvero le stesse della matrice di partenza $M^{(k)}$.
Con questo formalismo, si può algebricamente verificare che il MEG equivale ad applicare alla matrice A di partenza, una successione di matrici $M^{(1)}, M^{(2)}, \dots, \dots, M^{(k)}, \dots, M^{(n-2)}, M^{(n-1)}$, ovvero

$$M^{(n-1)}M^{(n-2)} \dots M^{(1)}A = U.$$

Il prodotto di queste trasformazioni genera la matrice U. Pertanto, possiamo considerare la matrice U del sistema finale come il risultato delle successive moltiplicazioni della matrice A per le matrici $M^{(k)}$. Questa proprietà si può verificare formalmente. Supponendo, quindi, corretta questa rappresentazione

$$M^{(n-1)}M^{(n-2)} \dots M^{(1)}A = U \Longrightarrow MA = U$$

definiamo il prodotto di tutte queste matrici come M. Se calcoliamo l'inversa di M, otteniamo esattamente la matrice L, ovvero

$$L = \left(M^{(n-1)}M^{(n-2)} \dots M^{(1)}\right)^{-1} = (M)^{-1}$$

Quindi, se definiamo come L la matrice inversa del prodotto di tutte queste matrici $M^{(k)}$, si osserva che abbiamo ottenuto

$$L^{-1}A = U,$$

e dunque, pre-moltiplicando per L a sinistra e a destra dell'uguale, troviamo che A è uguale ad L per U

$$LL^{-1}A = LU \Longrightarrow A = LU, \text{ dato che } LL^{-1} = I.$$

Abbiamo quindi verificato che il MEG conduce effettivamente alla scomposizione della matrice A come prodotto di due matrici triangolari:

- Una matrice triangolare inferiore L, determinata univocamente dai moltiplicatori generati nel corso del processo del MEG,

- Una matrice triangolare superiore U, che è la stessa che si ottiene nell'ultimo stadio del MEG.

Questa fattorizzazione

$$A = LU,$$

è nota come fattorizzazione LU e conferma che il MEG può essere interpretato come un metodo di fattorizzazione della matrice A.
La fattorizzazione LU, così presentata, è una reinterpretazione del MEG. È interessante dal punto di vista matematico, ma lo è ancor di più dal punto di vista pratico. Infatti, è possibile dimostrare che con questa interpretazione è possibile risolvere il sistema $A\boldsymbol{x} = \boldsymbol{b}$ senza dover eseguire esplicitamente il MEG, ma sviluppando un algoritmo che sfrutta direttamente la caratterizzazione di A come prodotto di L per U. Tale algoritmo è detto decomposizione di Doolittle. Una sua derivazione è mostrata nell' Approfondimento di questa lezione.
Vediamo dunque quali sono i vantaggi di questa rappresentazione per risolvere un sistema lineare generico. In primo luogo, L e U non dipendono dal termine noto del sistema, ma solo dalla matrice A. Pertanto, è possibile utilizzare la stessa fattorizzazione LU per risolvere più sistemi lineari con la stessa matrice, ma con termini noti differenti. Questo consente di eseguire una sola volta il processo di trasformazione di A nel prodotto di L per U

$$A = LU.$$

Una volta noti L ed U, la risoluzione di un generico sistema

$$A\boldsymbol{x} = \boldsymbol{c},$$

dove $\boldsymbol{c}$ è un generico vettore termine noto, si riconduce alla risoluzione del sistema

$$LU\boldsymbol{x} = \boldsymbol{c},$$

e se si pone

$$\boldsymbol{y} = U\boldsymbol{x},$$

troviamo

$$L\boldsymbol{y} = \boldsymbol{c},$$
$$U\boldsymbol{x} = \boldsymbol{y},$$

dove Il sistema $L\boldsymbol{y} = \boldsymbol{c}$ è un sistema autonomo con termine noto $\boldsymbol{c}$, che viene assegnato all'inizio, e con matrice L, che è la matrice triangolare inferiore della fattorizzazione. Una volta trovato $\boldsymbol{y}$, come soluzione del sistema $L\boldsymbol{y} = \boldsymbol{c}$, lo si usa come termine noto nel nuovo sistema $U\boldsymbol{x} = \boldsymbol{y}$, da cui si calcola la soluzione $\boldsymbol{x}$.
Pertanto, se L e U sono già note, è necessario risolvere due sistemi triangolari: prima un sistema triangolare inferiore $L\boldsymbol{y} = \boldsymbol{c}$, e poi un sistema triangolare superiore $U\boldsymbol{x} = \boldsymbol{y}$. In questo modo, abbiamo due sistemi triangolari da risolvere, uno inferiore e l'altro superiore, il che richiede un numero di operazioni complessivo dell'ordine di $2n^2$ operazioni.
Abbiamo quindi visto che la matrice A, se scomposta nel prodotto LU, permette di risolvere un sistema $A\boldsymbol{x} = \boldsymbol{b}$ (o $A\boldsymbol{x} = \boldsymbol{c}$, ovvero con un termine noto arbitrario) tramite il processo di sostituzione in avanti e all'indietro. In questo modo, il numero complessivo di operazioni necessarie è dell'ordine di $2n^2$.

Approfondimento:
Vediamo adesso come possiamo procedere alla fattorizzazione della matrice A nel prodotto LU, senza ricorrere all'applicazione del Metodo di Eliminazione di Gauss (MEG). Abbiamo osservato, nella lezione principale, che la fattorizzazione LU è equivalente al MEG, tuttavia, vogliamo adesso determinare direttamente i fattori L ed U, senza passare attraverso il processo del MEG, che richiede una sequenza di $n-1$ passaggi. Per farlo, possiamo adottare un approccio puramente algebrico, imponendo relazioni algebriche sui singoli elementi della matrice.
Abbiamo quindi che

$$A = LU,$$

e questo significa che

$$(a_{ij}) = (LU)_{ij},$$

ovvero, abbiamo una relazione di identità fra gli elementi che hanno gli stessi indici.
Poiché gli elementi di A, U e L sono

$$A = (a_{ij}), \qquad U = (u_{kj}), \qquad L = (m_{ik}),$$

Se andiamo a imporre questa relazione $(a_{ij}) = (LU)_{ij}$ elemento per elemento, troviamo la relazione

$$a_{ij} = \sum_{k=1}^{n} m_{ik} u_{kj},$$

che è possibile ricavare utilizzando la regola di moltiplicazione (righe per colonne) di due matrici fra loro. La specificità di questa operazione risiede nel fatto che L è una matrice triangolare inferiore e U è una matrice triangolare superiore. Di conseguenza, molti dei prodotti che compaiono in questa fattorizzazione saranno effettivamente nulli, poiché o m_{ik} è uguale a zero, oppure u_{ik} è uguale a zero. Più precisamente, se k è maggiore di i o maggiore di j (o di entrambi), il prodotto corrispondente risulterà nullo. Pertanto, possiamo limitare la somma all'indice superiore r, che è definito come il minimo tra i e j

$$a_{ij} = \sum_{k=1}^{n} m_{ik} u_{kj} = \sum_{k=1}^{r} m_{ik} u_{kj} \qquad r = \min(i,j).$$

Questo ci permette di caratterizzare queste uguaglianze nei due casi seguenti

$$se\ j \geq p \quad a_{pj} = \sum_{k=1}^{p} m_{pk} u_{kj} = u_{pj} + \sum_{k=1}^{p-1} m_{pk} u_{kj},$$

$$se\ i > p \quad a_{ip} = \sum_{k=1}^{p} m_{ik} u_{kp} = m_{ip} u_{pp} + \sum_{k=1}^{p-1} m_{ik} u_{kp}.$$

Nel caso $j \geq p$, troviamo l'elemento a_{pj}, dove abbiamo isolato, nella seconda uguaglianza, l'elemento u_{pj} poiché esso viene moltiplicato per uno. In altre parole, stiamo considerando gli elementi il cui indice di riga è minore dell'indice di colonna, corrispondenti alla parte triangolare inferiore della matrice A. Se invece $i > p$, stiamo considerando gli altri elementi e isoliamo il termine diagonale u_{pp}. Queste relazioni possono essere derivabili in modo puramente algebrico.
Vogliamo infine esaminare come queste due formule ci permettano di calcolare gli elementi m_{ip} della matrice L e gli elementi u_{p+1}, j della matrice U in maniera sequenziale, ossia calcolando prima una riga e poi una colonna successiva. Iniziamo osservando che la matrice U ha come prima riga gli stessi elementi della matrice A, poiché il processo del MEG non modifica mai la prima riga di A. Questo ci consente di calcolare immediatamente gli elementi della prima riga di U

$$u_{1j} = a_{1j} \qquad j = 1, \dots, n, \quad \text{prima riga di U.}$$

Possiamo poi utilizzare questa formula

$$m_{ip} = \frac{1}{u_{pp}} \left(a_{ip} - \sum_{k=1}^{p-1} m_{ik} u_{kp} \right) \qquad i > p,$$

con $p = 1$ per generare la prima colonna, ovvero tutti gli elementi della prima colonna della matrice L. Dopodiché possiamo calcolare la riga di posizione $p + 1$ e quindi tutti gli elementi di tipo u_{p+1}, j in questa maniera

$$u_{p+1,j} = a_{p+1,j} - \sum_{k=1}^{p} m_{p+1,k} u_{kj} \qquad j \geq p + 1,$$

Una volta determinate tutte le righe precedenti e le colonne precedenti, è possibile generare la nuova riga e la nuova colonna in modo ricorsivo. Questo approccio sequenziale consente di calcolare tutti gli elementi delle righe della matrice U e delle colonne della matrice L, direttamente a partire dagli elementi della matrice A. Pertanto, non è necessario eseguire il MEG. Di fatto, Il MEG è stato utilizzato esclusivamente per comprendere come ottenere la caratterizzazione della matrice A nel prodotto LU. Infine, una volta calcolati algebricamente tutti gli elementi di L e U, possiamo risolvere i sistemi triangolari inferiori e superiori associati rispettivamente alle matrici L e U.