

Risoluzione di sistemi lineari sparsi e di grandi dimensioni

Un sistema lineare $A\mathbf{x} = \mathbf{b}$ con $A \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$, è sparso quando il numero di elementi della matrice A diversi da zero è αn , con $n \gg \alpha$.

Una caratteristica dei sistemi lineari derivanti dalla discretizzazione di equazioni ellittiche è quello di avere dimensioni molto elevate (se $N = M = 100$, allora $n \simeq 10^4$) e matrice dei coefficienti a struttura sparsa. I cosiddetti metodi diretti (metodo di Gauss, fattorizzazione LU , fattorizzazione QR , etc.) non possono essere usati efficientemente per risolvere sistemi di grandi sparsi dimensioni per due motivi:

1. La quantità di memoria richiesta
2. L'alto numero di operazioni richiesto (proporzionale a n^3).

Metodi iterativi per sistemi lineari

Si tratta di una classe di metodi basata sulla costruzione di una successione di vettori $\mathbf{x}^{(k)}$ che, a partire da un vettore $\mathbf{x}^{(0)}$, **approssimazione iniziale della soluzione**, sotto opportune ipotesi converge alla soluzione del sistema lineare, cioè:

$$\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{x}^* = A^{-1}\mathbf{b}.$$

Il metodo di Jacobi

Assumiamo che gli elementi diagonali della matrice a_{ii} , $i = 1, \dots, n$, siano diversi da 0.

La i -esima equazione del sistema si scrive

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i$$

e, isolando x_i risulta:

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

Queste n equazioni sono del tutto equivalenti al sistema di partenza.

Si tratta di n uguaglianze solo se le componenti x_i sono quelle del vettore soluzione.

Conoscendo l'approssimazione iniziale $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$ si può ottenere un'approssimazione, al passo 1, ponendo:

$$x_i^{(1)} = \left(b_i - \sum_{j=1, j \neq i}^n a_{ij} x_j^{(0)} \right) \frac{1}{a_{ii}} \quad i = 1, \dots, n.$$

e calcolare la successione di vettori $\{\mathbf{x}^{(k)}\}$ ponendo

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

con $k = 0, 1, 2, \dots$. Questa è l'espressione del metodo di Jacobi.

Il Metodo di Gauss-Seidel

Una variante del metodo di Jacobi si ottiene osservando che, quando si calcola $x_i^{(k+1)}$ si possono utilizzare le approssimazioni già calcolate $x_j^{(k+1)}$, con $j = 1, \dots, i - 1$, ipotizzando che queste siano più precise, ottenendo così

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

Il Metodo del Rilassamento

Si deve osservare che entrambi i metodi appena introdotti non utilizzano la componente i -esima al passo k . Per questo si introduce una nuova variante che coinvolge tale valore a partire da un parametro $\omega \neq 0$. In particolare indicata con $\mathbf{x}_{GS}^{(k+1)}$ l'approssimazione ottenuta con il metodo di Gauss-Seidel, si può ottenere una nuova approssimazione ponendo

$$\mathbf{x}^{(k+1)} = \omega \mathbf{x}_{GS}^{(k+1)} + (1 - \omega) \mathbf{x}^{(k)}.$$

In modo esplicito si ottiene il seguente schema iterativo:

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

Questa classe di metodi prende il nome di *Metodo di Rilassamento*. Si osserva facilmente che se si pone $\omega = 1$ il metodo di Rilassamento coincide con il metodo di Gauss-Seidel.

Convergenza dei Metodi Iterativi

Per la convergenza dei metodi introdotti conviene dare un'impostazione generale, nota con il nome di **Teoria dello splitting**.

Definizione. Una coppia di matrici M ed N costituiscono uno splitting per la matrice non singolare A se:

- $A = M + N$;
- $\det M \neq 0$.

Una volta scelto lo splitting, il sistema $A\mathbf{x} = \mathbf{b}$, cioè

$$(M + N)\mathbf{x} = \mathbf{b}$$

si mette nella forma

$$M\mathbf{x} = -N\mathbf{x} + \mathbf{b}$$

Poichè M è non singolare

$$\mathbf{x} = -M^{-1}N\mathbf{x} + M^{-1}\mathbf{b}$$

ovvero, ponendo

$$B = -M^{-1}N, \quad \mathbf{c} = M^{-1}\mathbf{b}.$$

si ha:

$$\mathbf{x} = B\mathbf{x} + \mathbf{c}.$$

L'equazione suggerisce un metodo iterativo per approssimare la soluzione del sistema. Scelto un vettore iniziale $\mathbf{x}^{(0)}$ si definisce la successione di vettori

$$M\mathbf{x}^{(k+1)} = -N\mathbf{x}^{(k)} + \mathbf{b}, \quad k = 0, 1, 2, \dots$$

o equivalentemente

$$\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{c}, \quad k = 0, 1, 2, \dots$$

È evidente che la matrice M deve essere scelta in modo tale che il sistema lineare sia di facile soluzione.

Definizione.

Il metodo iterativo è convergente per un sistema $A\mathbf{x} = \mathbf{b}$ se la successione dei vettori approssimazione è convergente con un qualsiasi vettore iniziale.

Una definizione equivalente coinvolge la successione dei vettori errore

$$\mathbf{e}^{(k)} = \mathbf{x}^{(k)} - \mathbf{x}, \quad k = 0, 1, 2, \dots$$

dove \mathbf{x} è la soluzione teorica del sistema.

Definizione.

Il metodo iterativo è convergente per un sistema $A\mathbf{x} = \mathbf{b}$ se la successione dei vettori errore è infinitesima per ogni vettore errore iniziale:

$$\lim_{k \rightarrow \infty} \mathbf{e}^{(k)} = \mathbf{0}.$$

Sottraendo le relazioni:

$$\mathbf{x}^{(k+1)} = B\mathbf{x}^{(k)} + \mathbf{c}, \quad \mathbf{x} = B\mathbf{x} + \mathbf{c}$$

si ottiene la cosiddetta **prima equazione degli errori**:

$$\mathbf{e}^{(k+1)} = B\mathbf{e}^{(k)}, \quad k = 0, 1, 2, \dots$$

Tale equazione è posta anche in una seconda forma che lega $\mathbf{e}^{(k)}$ all'errore iniziale $\mathbf{e}^{(0)}$, infatti, per induzione, è facile verificare la cosiddetta **seconda equazione degli errori**:

$$\mathbf{e}^{(k)} = B^k \mathbf{e}^{(0)}.$$

Dalle equazioni degli errori si deduce che la convergenza di un metodo iterativo dipende esclusivamente dallo splitting scelto per la matrice dei coefficienti. La matrice B definita in precedenza prende il nome di **matrice di iterazione del metodo**.

Definizione.

Una matrice quadrata B si dice *convergente* se

$$\lim_{k \rightarrow +\infty} B^k = 0.$$

Teorema.

Un metodo iterativo è convergente se e solo se è convergente la relativa matrice di iterazione.

Corollario.

Se esiste una norma $\| \cdot \|$ tale che

$$\|B\| < 1$$

allora il metodo iterativo è convergente.

Corollario.

Se un metodo iterativo è convergente allora il determinante della matrice di iterazione è, in modulo, minore di 1.

Ricordiamo che il raggio spettrale della matrice B , $\rho(B)$, è, per definizione il massimo modulo di un autovalore quindi abbiamo la seguente formulazione alternativa del teorema di convergenza.

Teorema.

Un metodo iterativo è convergente se e solo se è il raggio spettrale della matrice di iterazione è minore di 1.

Vediamo ora come riottenere, dal punto di vista dello splitting i metodi iterativi precedentemente introdotti. Decomponiamo la matrice A come

$$A = L + D + U$$

dove D è la matrice degli elementi diagonali, L è la parte strettamente triangolare inferiore mentre U è la parte strettamente triangolare superiore.

La formulazione del metodo di Jacobi può essere anche la seguente:

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

che, in forma matriciale si può scrivere come

$$D\mathbf{x}^{(k+1)} = -(L + U)\mathbf{x}^{(k)} + \mathbf{b}, \quad k = 0, 1, 2, \dots$$

il che equivale a scegliere

$$M = D, \quad N = U + L.$$

Quindi

$$B_J = -D^{-1}(U + L).$$

Il metodo di Gauss-Seidel può essere scritto come

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

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

che, in forma matriciale diventa

$$(D + L)\mathbf{x}^{(k+1)} = -U\mathbf{x}^{(k)} + \mathbf{b}, \quad k = 0, 1, 2, \dots$$

che corrisponde invece alla scelta

$$M = D + L \quad N = U.$$

La matrice di iterazione è quindi:

$$B_{GS} = -(D + L)^{-1}U.$$

Per il metodo di Rilassamento abbiamo invece

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

$$D\mathbf{x}^{(k+1)} = \omega(\mathbf{b} - L\mathbf{x}^{(k+1)} - U\mathbf{x}^{(k)}) + (1 - \omega)D\mathbf{x}^{(k)}$$

da cui

$$(D + \omega L)\mathbf{x}^{(k+1)} = [(1 - \omega)D - \omega U]\mathbf{x}^{(k)} + \omega\mathbf{b}$$

$$\left(\frac{D}{\omega} + L\right)\mathbf{x}^{(k+1)} = \left[\left(\frac{1}{\omega} - 1\right)D - U\right]\mathbf{x}^{(k)} + \mathbf{b}$$

ricavando così

$$M = \frac{D}{\omega} + L, \quad N = \left(1 - \frac{1}{\omega}\right)D + U.$$

Per quello che riguarda l'intervallo di appartenenza del parametro ω vale il seguente risultato.

Teorema.

Condizione necessaria per la convergenza del metodo di Rilassamento è che

$$|\omega - 1| < 1,$$

quindi se ω è reale:

$$0 < \omega < 2.$$

Dimostrazione. La matrice M è triangolare inferiore e, poichè U ha elementi diagonali nulli è

$$\det M = \det \frac{D}{\omega} = \frac{1}{\omega^n} \det D.$$

Analogamente per la matrice N :

$$\det N = \frac{(\omega - 1)^n}{\omega^n} \det D.$$

Quindi

$$|\det(M^{-1}N)| = |(1 - \omega)^n|.$$

Poiché condizione necessaria per la convergenza di un metodo iterativo è che il modulo del determinante della matrice di iterazione sia minore di 1 abbiamo:

$$|(1 - \omega)^n| < 1 \Rightarrow |1 - \omega| < 1 \Rightarrow 0 < \omega < 2. \quad \square$$

Classi di Matrici per cui i Metodi Iterativi sono Convergenti

- Matrici a Stretta Predominanza Diagonale per Righe:

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

- Matrici a Stretta Predominanza Diagonale per Colonne:

$$|a_{jj}| > \sum_{i=1, i \neq j}^n |a_{ij}|, \quad j = 1, \dots, n$$

- Matrici a Predominanza Diagonale per Righe:

$$|a_{ii}| \geq \sum_{j=1, j \neq i}^n |a_{ij}|, \quad i = 1, \dots, n$$

ed esiste $k \in \{1, 2, \dots, n\}$ tale che:

$$|a_{kk}| > \sum_{j=1, j \neq k}^n |a_{kj}|$$

- Matrici a Predominanza Diagonale per Colonne:

$$|a_{jj}| \geq \sum_{i=1, i \neq j}^n |a_{ij}|, \quad j = 1, \dots, n$$

ed esiste $k \in \{1, 2, \dots, n\}$ tale che:

$$|a_{kk}| > \sum_{i=1, i \neq k}^n |a_{ik}|.$$

Il Metodo a cinque punti

Le matrici derivanti dalla discretizzazione dell'equazione di Laplace hanno il termine

$$a_{ii} = -2(h^2 + k^2)$$

poi sulla riga i -esima ci sono al più 4 elementi diversi da zero:

$$a_{i,i-1} = a_{i,i+1} = k^2, \quad a_{i,i-N+1} = a_{i,i+N-1} = h^2$$

quindi i metodi sono convergenti.

Metodi di tipo Gradiente

Consideriamo il problema di minimizzare la funzione

$$\Phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{x}^T \mathbf{b}, \quad \mathbf{x} \in \mathbb{R}^n,$$

dove $\mathbf{b} \in \mathbb{R}^n$ e $A \in \mathbb{R}^{n \times n}$ è una matrice simmetrica e definita positiva, cioè

$$\mathbf{x}^T A\mathbf{x} > 0, \quad \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq 0.$$

La soluzione del problema è nota perchè il gradiente vale:

$$\nabla\Phi(x) = A\mathbf{x} - \mathbf{b}$$

e si annulla quando

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

Nel caso $n = 2$ con

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

la funzione $\Phi(x, y)$ può essere scritta esplicitamente:

$$\Phi(x, y) = \frac{1}{2} \left[a_{11}x^2 + (a_{12} + a_{21})xy + a_{22}y^2 \right] - b_1x - b_2y,$$

e, poichè A è simmetrica:

$$\Phi(x, y) = \frac{1}{2} \left[a_{11}x^2 + 2a_{12}xy + a_{22}y^2 \right] - b_1x - b_2y.$$

Poichè i due problemi hanno la stessa soluzione ha senso quindi applicare i metodi di minimizzazione di $\Phi(\mathbf{x})$ per risolvere il sistema lineare

$$A\mathbf{x} = \mathbf{b}.$$

Un metodo per minimizzare $\Phi(\mathbf{x})$ è il metodo di **steepest descent** (traducibile in italiano come **metodo di più ripida discesa**). Il metodo si basa sull'osservazione che nel punto $\mathbf{x}_k \in \mathbb{R}^n$ la funzione $\Phi(\mathbf{x})$ decresce più rapidamente nella direzione del gradiente con il segno cambiato:

$$-\nabla\Phi(\mathbf{x}_k) = \mathbf{b} - A\mathbf{x}_k.$$

Poniamo

$$\mathbf{r}_k = \mathbf{b} - A\mathbf{x}_k$$

il vettore **residuo** nell'approssimazione \mathbf{x}_k .

Se il residuo è diverso da zero cerchiamo di trovare il parametro $\alpha > 0$ tale che:

$$\Phi(\mathbf{x}_k + \alpha\mathbf{r}_k) < \Phi(\mathbf{x}_k).$$

ed in particolare il valore di α tale che

$$\Phi(\mathbf{x}_k + \alpha_k\mathbf{r}_k) = \min_{\alpha \in \mathbb{R}} \Phi(\mathbf{x}_k + \alpha\mathbf{r}_k).$$

Calcolando la derivata di $\Phi(\mathbf{x}_k + \alpha \mathbf{r}_k)$ rispetto ad α risulta:

$$\begin{aligned}\Phi'(\mathbf{x}_k + \alpha \mathbf{r}_k) &= \mathbf{r}_k^T \nabla \Phi(\mathbf{x}_k + \alpha \mathbf{r}_k) = \\ &= \mathbf{r}_k^T [A(\mathbf{x}_k + \alpha \mathbf{r}_k) - \mathbf{b}] = \\ &= \mathbf{r}_k^T [A\mathbf{x}_k - \mathbf{b} + \alpha A\mathbf{r}_k] = \\ &= \mathbf{r}_k^T [-\mathbf{r}_k + \alpha A\mathbf{r}_k] = \alpha \mathbf{r}_k^T A\mathbf{r}_k - \mathbf{r}_k^T \mathbf{r}_k = 0\end{aligned}$$

Quindi

$$\Phi'(\mathbf{x}_k + \alpha \mathbf{r}_k) = 0 \quad \Rightarrow \quad \alpha = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_k^T A\mathbf{r}_k}.$$

Osserviamo che α_k è sempre ben definito perchè la matrice è simmetrica e definita positiva e quindi $\mathbf{r}_k^T A\mathbf{r}_k > 0$.

Abbiamo quindi il seguente algoritmo che descrive lo steepest descent method:

```
 $k = 0$   
 $\mathbf{x}_0$  arbitrario  
 $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$   
while  $\mathbf{r}_k \neq 0$   
     $\alpha_k = (\mathbf{r}_k^T \mathbf{r}_k) / (\mathbf{r}_k^T A \mathbf{r}_k)$   
     $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{r}_k$   
     $\mathbf{r}_{k+1} = \mathbf{b} - A\mathbf{x}_{k+1}$   
     $k = k + 1$   
end
```

Il vantaggio di questo metodo è che si genera un insieme di direzioni $\mathbf{r}_0, \mathbf{r}_1, \dots$ a due a due ortogonali. Quindi è certo che, teoricamente, dopo al massimo n iterazioni il metodo arriva a convergenza.