Metodi Iterativi per la Risoluzione di Sistemi Lineari

Luca Gemignani
luca.gemignani@unipi.it

21 marzo 2018

Indice

Lezione 1: Generalità sui Metodi Iterativi. **1**

Lezione 2: I Metodi di Jacobi e Gauss-Seidel. **4**

Lezione 3: Convergenza dei Metodi di Jacobi e Gauss-Seidel. **6**

Lezione 4: Raffinamento Iterativo. **8**

Lezione 1: Generalità sui Metodi Iterativi.

Sistemi lineari $A\boldsymbol{x} = \boldsymbol{b}$ dove la matrice dei coefficienti $A \in \mathbb{R}^{n\times n}$ è sparsa o di elevate dimensioni ($n > 10^6$) sono generalmente risolti numericamente mediante metodi iterativi che a partire da un vettore iniziale $\boldsymbol{x}^{(0)}$ generano una sequenza di approssimazioni $\boldsymbol{x}^{(k)}$, $k > 0$, che converge alla soluzione del sistema lineare, i.e.,

$$\lim_{k\to+\infty} \| \boldsymbol{x}^{(k)} - \boldsymbol{x} \| = 0.$$

In pratica la costruzione della successione termina dopo un numero finito di passi determinato in base alla verifica di opportuni *criteri di arresto*. La qualità e l'efficienza di un metodo iterativo è pertanto determinata dalle proprietà di convergenza della successione generata.

Una tecnica generale per derivare un metodo iterativo si basa sulla decomposizione additiva $A = M - N$ con M matrice invertibile. Si ha allora

$$A\boldsymbol{x} = \boldsymbol{b} \iff (M - N)\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{x} = M^{-1}N\boldsymbol{x} + M^{-1}\boldsymbol{b}.$$

Posto dunque $P = M^{-1}N$ detta *matrice di iterazione* e $q = M^{-1}\boldsymbol{b}$ si ottiene che

$$A\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{x} = P\boldsymbol{x} + \boldsymbol{q},$$

ovvero $\boldsymbol{x}$ è soluzione del sistema lineare se e soltanto se

$$g(\boldsymbol{x}) = \boldsymbol{x}, \quad g\colon \mathbb{R}^n \to \mathbb{R}^n,\ g(\boldsymbol{z}) = P\boldsymbol{z} + \boldsymbol{q}. \tag{1}$$

Per la soluzione del problema *di punto fisso* (1) vale il seguente risultato.

Teorema 1.1. Dato $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$ sia $\boldsymbol{x}^{(k+1)} = g(\boldsymbol{x}^{(k)})$, $k \geq 0$. Se $\lim_{k\to+\infty} \boldsymbol{x}^{(k+1)} = \boldsymbol{x}$ allora $g(\boldsymbol{x}) = \boldsymbol{x}$ e dunque $A\boldsymbol{x} = \boldsymbol{b}$.

Dimostrazione. Segue dalla continuità di g. Vale infatti

$$0 \leq \| g(\boldsymbol{x}^{(k)}) - g(\boldsymbol{x}) \| = \| P\boldsymbol{x}^{(k)} - P\boldsymbol{x} \| \leq \| P \| \| \boldsymbol{x}^{(k)} - \boldsymbol{x} \| .$$

□

Questo risultato motiva l'introduzione del seguente metodo iterativo per la risoluzione del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ con $A \in \mathbb{R}^{n\times n}$ matrice invertibile:

$$\begin{cases} \boldsymbol{x}^{(0)} \in \mathbb{R}^n \\ \boldsymbol{x}^{(k+1)} = P\boldsymbol{x}^{(k)} + \boldsymbol{q}, \quad k \geq 0, \end{cases} \tag{2}$$

con

$$P = M^{-1}N, \quad \boldsymbol{q} = M^{-1}\boldsymbol{b}, \quad A = M - N.$$

Si osservi che (2) può essere scritto in maniera formalmente (ma non computazionalmente) equivalente come segue

$$\begin{cases} \boldsymbol{x}^{(0)} \in \mathbb{R}^n \\ M\boldsymbol{x}^{(k+1)} = N\boldsymbol{x}^{(k)} + \boldsymbol{b}, \quad k \geq 0. \end{cases} \tag{3}$$

Definizione 1.1. Il metodo (2) ((3)) si dice *convergente* se la successione generata dal metodo per ogni scelta del punto iniziale $\boldsymbol{x}^{(0)}$ converge alla soluzione $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ del sistema lineare.

Il seguente risultato fornisce una condizione sufficiente per la convergenza del metodo (2).

Teorema 1.2. Il metodo (2) è convergente se esiste una norma matriciale indotta da una norma vettoriale $\| \cdot \|$ su $\mathbb{R}^n$ tale per cui $\| P \| < 1$.

Dimostrazione. Dalle relazioni

$$\boldsymbol{x}^{(k+1)} = P\boldsymbol{x}^{(k)} + \boldsymbol{q}, \quad \boldsymbol{x} = P\boldsymbol{x} + \boldsymbol{q},$$

segue che

$$\boldsymbol{e}^{(k+1)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x} = P(\boldsymbol{x}^{(k)} - \boldsymbol{x}) = P\boldsymbol{e}^{(k)}, \quad k \geq 0,$$

e quindi

$$\boldsymbol{e}^{(k+1)} = P^{k+1}\boldsymbol{e}^{(0)}, \quad k \geq 0.$$

Passando alla norma vettoriale si ha

$$\| \boldsymbol{e}^{(k+1)} \| = \| P^{k+1} \boldsymbol{e}^{(0)} \| \leq \| P^{k+1} \| \| \boldsymbol{e}^{(0)} \|,$$

da cui

$$0 \leq \| \boldsymbol{e}^{(k+1)} \| \leq \| P \|^{k+1} \| \boldsymbol{e}^{(0)} \|,$$

da cui per il teorema del confronto segue che $\forall \boldsymbol{e}^{(0)}$ o, equivalentemente, $\forall \boldsymbol{x}^{(0)}$

$$\lim_{k \to +\infty} \| \boldsymbol{e}^{(k+1)} \| = \lim_{k \to +\infty} \| \boldsymbol{x}^{(k+1)} - \boldsymbol{x} \| = 0.$$

□

Il seguente risultato descrive una condizione necessaria per la convergenza del metodo (2). Ricordiamo che il *raggio spettrale* di una matrice $B \in \mathbb{R}^{n \times n}$ è definito come $\rho(B) = \max_i |\lambda_i|$, $\lambda_1, \ldots, \lambda_n$ autovalori di B.

Teorema 1.3. Se il metodo (2) è convergente allora $\rho(P) < 1$.

Dimostrazione. Sia λ tale che $|\lambda| = \rho(P)$ e $\boldsymbol{v}$ un corrispondente autovettore di P, i.e., $P\boldsymbol{v} = \lambda \boldsymbol{v}$, $\boldsymbol{v} \neq \boldsymbol{0}$. Sia $\boldsymbol{x}^{(0)} = \boldsymbol{x} + \boldsymbol{v}$ con $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ soluzione del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$. La successione generata dal metodo (2) con punto iniziale $\boldsymbol{x}^{(0)}$ è convergente ad $\boldsymbol{x}$. D'altra parte si ha

$$\boldsymbol{e}^{(k+1)} = P^{k+1} \boldsymbol{e}^{(0)} = P^{k+1} \boldsymbol{v} = \lambda^k \boldsymbol{v},$$

da cui

$$\| \boldsymbol{e}^{(k+1)} \| = \| \lambda^k \boldsymbol{v} \| = |\lambda|^K \| \boldsymbol{v} \|,$$

e quindi

$$\lim_{k \to +\infty} |\lambda|^K = 0$$

che implica

$$|\lambda| < 1.$$

□

Dal Teorema di Hirsch segue che per ogni norma matriciale indotta vale

$$\rho(A) \leq \| A \|, \quad \forall A \in \mathbb{R}^{n \times n}$$

per cui la condizione sufficiente implica la condizione necessaria. Inoltre vale il seguente risultato di cui omettiamo la dimostrazione.

Teorema 1.4. Sia $A \in \mathbb{R}^{n \times n}$ con $\rho(A) < 1$. Allora esiste una norma matriciale indotta tale per cui $\| A \| < 1$.

Combinando tra loro i teoremi 1.2, 1.3 e 1.4 si perviene infine al seguente risultato.

Teorema 1.5. Condizione necessaria e sufficiente per la convergenza del metodo iterativo (2) è che $\rho(P) < 1$.

Lezione 2: I Metodi di Jacobi e Gauss-Seidel.

Sia $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ invertibile con *elementi diagonali non nulli*, i.e.,

$$a_{i,i} \neq 0, \quad 1 \leq i \leq n. \tag{4}$$

Poniamo $A = D - L - U$ con $D = (d_{i,j})$, $L = (l_{i,j})$ e $U = (u_{i,j})$ definite come segue:

$$d_{i,j} = \begin{cases} a_{i,j} \text{ se } i = j; \\ 0 \textit{ altrimenti}, \end{cases}$$

$$l_{i,j} = \begin{cases} -a_{i,j} \text{ se } i > j; \\ 0 \textit{ altrimenti}, \end{cases}$$

$$u_{i,j} = \begin{cases} -a_{i,j} \text{ se } i < j; \\ 0 \textit{ altrimenti}. \end{cases}$$

Il *metodo iterativo di Jacobi* per la risoluzione del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ è definito dal partizionamento

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

Il *metodo iterativo di Gauss-Seidel* per la risoluzione del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ è definito dal partizionamento

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

Poichè per entrambi i metodi M risulta triangolare inferiore con elementi diagonali di A si ha che la condizione (4) garantisce *l'applicabilità* dei metodi. Sotto tale assunzione i metodi sono implementati nella formulazione (3). Per il metodo di Jacobi si ottiene per $i = 1, 2, \ldots, n$,

$$a_{i,i}\boldsymbol{x}_i^{(k+1)} = \boldsymbol{b}_i - \sum_{j=1, j\neq i}^{n} a_{i,j}x_j^{(k+1)} \;\rightarrow\; \boldsymbol{x}_i^{(k+1)} = \frac{\boldsymbol{b}_i - \sum_{j=1, j\neq i}^{n} a_{i,j}x_j^{(k+1)}}{a_{i,i}}.$$

Per il metodo di Gauss-Seidel si ottiene

$$\sum_{j=1}^{i} a_{i,j}\boldsymbol{x}_j^{(k+1)} = \boldsymbol{b}_i - \sum_{j=i+1}^{n} a_{i,j}x_j^{(k+1)}, \quad i = 1, 2, \ldots, n,$$

da cui

$$\boldsymbol{x}_i^{(k+1)} = \frac{\boldsymbol{b}_i - \sum_{j=1}^{i-1} a_{i,j}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j}x_j^{(k)}}{a_{i,i}}, \quad i = 1, 2, \ldots, n.$$

I seguenti programmi MatLab prendono in input la matrice A, il vettore $\boldsymbol{b}$ ed una approssimazione $\boldsymbol{x}\_old$ di $\boldsymbol{x}$ e restituiscono in output la nuova approssimazione $\boldsymbol{x}\_new$ di $\boldsymbol{x}$ generata dal metodo corrispondente. Per il metodo di Gauss-Seidel $\boldsymbol{x}\_new$ è sovrascritto direttamente in $\boldsymbol{x}\_old$.

```
function [x_new] = jacobi_mio(A,b,x_old)
n=length(b);
for k=1:n
    s=0;
    for j=1:k-1
        s=s+A(k,j)*x_old(j);
    end
    for j=k+1:n
        s=s+A(k,j)*x_old(j);
    end
    x_new(k)=(b(k)-s)/A(k,k);
end
end
```

```
function [x_old] = gauss_seidel_mio(A,b,x_old)
n=length(b);
for k=1:n
    s=0;
    for j=1:k-1
        s=s+A(k,j)*x_old(j);
    end
    for j=k+1:n
        s=s+A(k,j)*x_old(j);
    end
    x_old(k)=(b(k)-s)/A(k,k);
end
end
```

Detto nnz(A) il numero di elementi non nulli della matrice A si osserva che una iterazione del metodo di Jacobi e di Gauss-Seidel applicati per la risoluzione del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ costa nnz(A) operazioni moltiplicative. Pertanto i metodi sono particolarmente interessanti per la risoluzione numerica di sistemi lineari sparsi (nnz(A) $<< n^2$).

La risoluzione numerica del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ con un metodo iterativo richiede la determinazione di un *criterio di arresto* che consenta di terminare l'elaborazione. Criteri usualmente utilizzati sono del tipo

$$\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \| \leq tol;$$

$$\frac{\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \|}{\| \boldsymbol{x}^{(k)} \|} \leq tol;$$

$$\| A\boldsymbol{x}^{(k+1)} - \boldsymbol{b} \| \leq tol;$$

$$\frac{\| A\boldsymbol{x}^{(k+1)} - \boldsymbol{b} \|}{\| \boldsymbol{x}^{(k+1)} \|} \leq tol,$$

dove *tol* indica una tolleranza prefissata, eventualmente combinati con una condizione sul numero massimo di iterazioni *max_iter* eseguite in modo da garantire

comunque (anche in caso di non convergenza) la terminazione del programma. Per i criteri basati sulla valutazione dell'errore assoluto e relativo in norma si osserva che se $\rho(P) < 1$ allora

$$\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k+1)} - \boldsymbol{x} + \boldsymbol{x} - \boldsymbol{x}^{(k)} = (P - I_n)(\boldsymbol{x}^{(k)} - \boldsymbol{x}),$$

da cui

$$\| \boldsymbol{x}^{(k)} - \boldsymbol{x} \| \leq \| (P - I_n)^{-1} \| \| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \| \leq tol \, \| (P - I_n)^{-1} \| .$$

Per i criteri basati sulla valutazione del residuo $\boldsymbol{r}^{(k+1)} = A\boldsymbol{x}^{(k+1)} - \boldsymbol{b}$ si ha

$$\boldsymbol{r}^{(k+1)} = A\boldsymbol{x}^{(k+1)} - \boldsymbol{b} = A\boldsymbol{x}^{(k+1)} - A\boldsymbol{x} = A(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}),$$

da cui

$$\| \boldsymbol{x}^{(k)} - \boldsymbol{x} \| \leq \| A^{-1} \| \| \boldsymbol{r}^{(k+1)} \| \leq tol \, \| A^{-1} \| .$$

Il seguente programma Matlab implementa il metodo di Jacobi arrestandosi quando $\| \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \|_\infty \leq tol$ o $k > max\_iter$.

```
function [x_new] = jacobi_solve(A,b,x_old, tol, max_iter)
err=+inf;
it=0;
while(err>tol && it<=max_iter)
x_new=jacobi_mio(A,b,x_old);
err=norm(x_new'-x_old, 'inf');
x_old=x_new';
it=it+1;
end
it
end
```

Lezione 3: Convergenza dei Metodi di Jacobi e Gauss-Seidel.

Nello studio della convergenza dei metodi iterativi di Jacobi e Gauss-Seidel siamo interessati a condizioni esprimibili in termini di proprietà della matrice A dei coefficienti piuttosto che della matrice di iterazione P che non è esplicitamente disponibile. Tra queste proprietà particolarmente rilevante è la seguente.

Definizione 3.1. Una matrice $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$ si dice *predominante diagonale* (per righe) se

$$|a_{i,i}| > \sum_{j=1, j \neq i}^{n} |a_{i,j}|, \quad 1 \leq i \leq n.$$

Per sistemi predominanti diagonali vale

Teorema 3.1. Sia $A = (a_{i,j}) \in \mathbb{R}^{n\times n}$ predominante diagonale. Allora:

1. A è invertibile;
2. i metodi di Jacobi e Gauss-Seidel per la risoluzione di un sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ sono applicabili;
3. i metodi di Jacobi e Gauss-Seidel per la risoluzione di un sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ sono convergenti.

Dimostrazione. Dimostriamo le tre proprietà.

1. L'invertibilità di A segue dal teorema di Gershgorin. Infatti vale

$$|0 - a_{i,i}| = |a_{i,i}| > \sum_{j=1, j\neq i}^{n} |a_{i,j}|, \quad 1 \leq i \leq n,$$

   e dunque

$$0 \notin U_{i=1}^{n} K_i.$$

2. Per l'applicabilità si ha

$$|a_{i,i}| > \sum_{j=1, j\neq i}^{n} |a_{i,j}| \Rightarrow |a_{i,i}| \neq 0, \quad 1 \leq i \leq n.$$

3. Dimostriamo quindi la convergenza. Dalla relazione

$$\det(P - \lambda I_n) = \det(M^{-1}N - \lambda I_n) = \det(N - \lambda M) = (-1)^n \det(\lambda M - N),$$

   segue che $\lambda \in \mathbb{C}$ è autovalore di P se e soltanto se $\det(\lambda M - N) = 0$. Si assuma ora $\lambda \in \mathbb{C}$, $|\lambda| \geq 1$. Si mostra che la matrice $\lambda M - N$ è predominante diagonale. Si ha infatti che

$$|a_{i,i}| > \sum_{j=1, j\neq i}^{n} |a_{i,j}| = \sum_{j=1}^{i-1} |a_{i,j}| + \sum_{j=i+1}^{n} |a_{i,j}|, \quad 1 \leq i \leq n,$$

   implica

$$|\lambda||a_{i,i}| > |\lambda| \sum_{j=1}^{i-1} |a_{i,j}| + |\lambda| \sum_{j=i+1}^{n} |a_{i,j}| \geq |\lambda| \sum_{j=1}^{i-1} |a_{i,j}| + \sum_{j=i+1}^{n} |a_{i,j}|, \quad 1 \leq i \leq n,$$

   e quindi

$$|\lambda a_{i,i}| > \sum_{j=1}^{i-1} |\lambda a_{i,j}| + \sum_{j=i+1}^{n} |a_{i,j}|, \quad 1 \leq i \leq n,$$

   e

$$|\lambda a_{i,i}| > \sum_{j=1}^{i-1} |a_{i,j}| + \sum_{j=i+1}^{n} |a_{i,j}|, \quad 1 \leq i \leq n.$$

Queste relazioni esprimono la predominanza diagonale della matrice $\lambda M - N$ ottenuta rispettivamente nel metodo di Gauss-Seidel e di Jacobi per $\lambda \in \mathbb{C}$, $|\lambda| \geq 1$. Ma per il punto (1) sappiamo che una matrice predominante diagonale è invertibile e dunque $\lambda \in \mathbb{C}$, $|\lambda| \geq 1$ allora $\det(\lambda M - N) \neq 0$ per Jacobi e Gauss-Seidel. Segue che per gli autovalori delle matrici di iterazione di questi metodi deve valere $|\lambda| < 1$ e dunque $\rho(P) < 1$ e dunque la convergenza segue dal teorema 1.5.

□

Lezione 4: Raffinamento Iterativo.

Abbiamo visto che il processo di eliminazione gaussiana in aritmetica a precisione finita determina un fattore triangolare $\tilde{U}$ ed un fattore "psychologically lower triangular matrix" $\tilde{L}$ per cui si ha (vedi relazione (4) nella nota sui metodi diretti)

$$\tilde{L}\tilde{U} = \tilde{A} = A + E,$$

con $\| E \|$ piccola. Per raffinare la soluzione $\tilde{\boldsymbol{x}}$ del sistema lineare $A\boldsymbol{x} = \boldsymbol{b}$ ottenuta calcolando in macchina $\tilde{U}\tilde{\boldsymbol{x}} = \tilde{L}^{-1}\boldsymbol{b}$ si può procedere come segue. Dalla relazione $A\boldsymbol{x} = \boldsymbol{b}$ segue che

$$A\boldsymbol{x} = \tilde{A}\boldsymbol{x} - E\boldsymbol{x} = \boldsymbol{b},$$

e quindi

$$\boldsymbol{x} = \tilde{A}^{-1}E\boldsymbol{x} + \boldsymbol{b},$$

che motiva il metodo iterativo

$$\boldsymbol{x}^{(0)} = \tilde{\boldsymbol{x}}, \quad \boldsymbol{x}^{(k+1)} = \tilde{A}^{-1}E\boldsymbol{x}^{(k)} + \tilde{A}^{-1}\boldsymbol{b}, \ k \geq 0.$$

Si osserva che il metodo può essere riscritto come

$$\boldsymbol{x}^{(0)} = \tilde{\boldsymbol{x}}, \quad \tilde{A}\boldsymbol{x}^{(k+1)} = (\tilde{A} - A)\boldsymbol{x}^{(k)} + \boldsymbol{b}, \quad k \geq 0,$$

da cui si ottiene

$$\boldsymbol{x}^{(0)} = \tilde{\boldsymbol{x}}, \quad \tilde{A}\boldsymbol{x}^{(k+1)} = (\tilde{A} - A)\boldsymbol{x}^{(k)} + \boldsymbol{b}, \quad k \geq 0,$$

e quindi

$$\boldsymbol{x}^{(0)} = \tilde{\boldsymbol{x}}, \quad \tilde{A}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}) = \boldsymbol{b} - A\boldsymbol{x}^{(k)}, \quad k \geq 0.$$

Il lettore dovrebbe dimostrare che se $\| \tilde{A}^{-1}E \| < 1$ allora il metodo iterativo converge. Si consideri quindi la procedura seguente che utilizza il programma *gauss_pp* descritto nella nota precedente.

```
function [x] = itera_ref(n)
a=invhilb(n);
b=zeros(n,1);
b(1)=1;
x=gauss_pp(a,b)
pause
for k=1:n
    r=double(b-a*sym(x,'f'));
    y=gauss_pp(a,r);
    x=double(sym(y, 'f')+sym(x, 'f'))
    pause
end
end
```

Si descriva la procedura implementata e si commentino i risultati ottenuti sperimentalmente per $n = 8, 12, 16$.