

Metodi iterativi per la soluzione di sistemi lineari

Alvise Sommariva

Università degli Studi di Padova
Dipartimento di Matematica

3 aprile 2017

Metodi iterativi e diretti

Problema.

Supponiamo che siano $A \in \mathbb{R}^{n \times n}$ (matrice non singolare), $b \in \mathbb{R}^n$ (vettore colonna). Bisogna risolvere il problema $Ax = b$ avente sol. unica x^* .

A tal proposito si può utilizzare la fatt. LU con pivoting. Il costo computazionale è in generale di $O(n^3/3)$ operazioni moltiplicative. Questo diventa proibitivo se n è particolarmente elevato.

Nota.

Sia A una matrice invertibile. Allora $PA = LU$ dove

- P è una matrice di permutazione,
- L è una matrice triangolare inferiore a diagonale unitaria ($l_{ii} = 1, \forall i$),
- U una matrice triangolare superiore.

Se $Ax = b$, nota $PA = LU$, per risolvere $Ax = b$ da

$Ax = b \Leftrightarrow PAX = Pb \Leftrightarrow LUX = Pb$ si risolve prima $Ly = Pb$ e detta y^* la soluzione, si determina x^* tale che $Ux^* = y^*$.

Metodi iterativi e diretti

- L'idea dei **metodi iterativi** è quello di ottenere una **successione di vettori** $x^{(k)} \rightarrow x^*$ **cosicchè per $\bar{k} \ll n$ sia $x^{(\bar{k})} \approx x^*$.**
- In generale, la soluzione non è ottenuta esattamente come nei metodi diretti in un numero finito di operazioni (in aritmetica esatta), ma quale limite, accontentandosi di poche iterazioni ognuna dal costo quadratico. Il costo totale sarà $O(\bar{k} \cdot n^2)$.

Metodi iterativi stazionari

Supponiamo che la matrice non singolare $A \in \mathbb{R}^{n \times n}$ sia tale che

$$A = M - N, \text{ con } M \text{ non singolare.}$$

Allora, se $Ax = b$ necessariamente $Mx - Nx = Ax = b$.

Di conseguenza da $Mx = Nx + b$

- moltiplicando ambo i membri per M^{-1} ,
- posto $\phi(x) = M^{-1}Nx + M^{-1}b$,

$$x = M^{-1}Nx + M^{-1}b = \phi(x).$$

Viene quindi naturale utilizzare la succ. del metodo di punto fisso

$$x^{(k+1)} = \phi(x^{(k)}) = M^{-1}Nx^{(k)} + M^{-1}b.$$

La matrice $P = M^{-1}N$ si dice di *iterazione* e non dipende, come pure b dall'indice di iterazione k . Per questo motivo tali metodi si chiamano **iterativi stazionari**.

Esempio A, E, F

Risulta utile anche scrivere la matrice A come $A = D - E - F$ dove

- D è una matrice diagonale,
- E è triangolare inferiore con elementi diagonali nulli,
- F è triangolare superiore con elementi diagonali nulli.

Definizione

Sia $A = M - N$, con M non singolare. Un metodo iterativo le cui iterazioni sono

$$x^{(k+1)} = \phi(x^{(k)}) = M^{-1}Nx^{(k)} + M^{-1}b.$$

si dice **iterativo stazionario**.

La matrice $P = M^{-1}N$ si dice di **matrice iterazione**.

Nota.

Si osservi che la matrice di iterazione $M^{-1}N$ come pure $M^{-1}b$ non dipendono dall'indice di iterazione k .

Esempio A, E, F

Esempio

```
>> A=[1 2 3 4; 5 6 7 2; 8 9 1 2; 3 4 5 1]
A =
     1     2     3     4
     5     6     7     2
     8     9     1     2
     3     4     5     1
>> E=-(tril(A)-diag(diag(A)))
E =
     0     0     0     0
    -5     0     0     0
    -8    -9     0     0
    -3    -4    -5     0
>> F=-(triu(A)-diag(diag(A)))
F =
     0    -2    -3    -4
     0     0    -7    -2
     0     0     0    -2
     0     0     0     0
>> % A=diag(diag(A))-E-F.
```

Metodo di Jacobi

Definizione (Metodo di Jacobi, 1846)

Sia $A = D - E - F \in \mathbb{R}^{n \times n}$ dove

- D è una matrice diagonale, non singolare
- E è triangolare inferiore con elementi diagonali nulli,
- F è triangolare superiore con elementi diagonali nulli.

Sia fissato $x^{(0)} \in \mathbb{R}^n$. Il *metodo di Jacobi* genera $\{x^{(k)}\}_k$ con

$$x^{(k+1)} = \phi(x^{(k)}) = M^{-1}N x^{(k)} + M^{-1}b$$

dove

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

Si vede che la matrice di iterazione $P = M^{-1}N$ è in questo caso
 $P = M^{-1}N = D^{-1}(E + F) = D^{-1}(D - D + E + F) = I - D^{-1}A.$

Metodo di Jacobi

Nota.

Se D è non singolare il metodo di Jacobi

$$x^{(k+1)} = \phi(x^{(k)}) = D^{-1}(E + F)x^{(k)} + D^{-1}b$$

può essere descritto come *metodo delle sostituzioni simultanee*

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

Nota.

Si osservi che se D è non singolare, ovviamente $a_{i,i} \neq 0$ per ogni indice i e quindi la successione è ben definita.

Metodo di Jacobi

Vediamo perchè questa derivazione. Da $Ax = b$, con $A \in \mathbb{R}^{n \times n}$ abbiamo

$$\sum_{j < i} a_{i,j}x_j + a_{i,i}x_i + \sum_{j > i} a_{i,j}x_j = \sum_j a_{i,j}x_j = b_i, \quad i = 1, \dots, n$$

e quindi evidenziando x_i , supposto $a_{i,i} \neq 0$ deduciamo

$$x_i = \frac{1}{a_{i,i}} \left(b_i - \left(\sum_{j < i} a_{i,j}x_j + \sum_{j > i} a_{i,j}x_j \right) \right), \quad i = 1, \dots, n$$

Quindi note delle approssimazioni della soluzione $x^{(k)} = \{x_j^{(k)}\}$ per $j = 1, \dots, n$ è naturale introdurre il metodo iterativo

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

che si vede facilmente essere il metodo di Jacobi.

Metodo di Gauss-Seidel

Definizione (Metodo di Gauss-Seidel, 1823-1874)

Sia $A = D - E - F \in \mathbb{R}^{n \times n}$ dove

- D è una matrice diagonale, non singolare
- E è triangolare inferiore con elementi diagonali nulli,
- F è triangolare superiore con elementi diagonali nulli.

Sia fissato $x^{(0)} \in \mathbb{R}^n$. Il **metodo di Gauss-Seidel** genera $\{x^{(k)}\}_k$ con

$$x^{(k+1)} = \phi(x^{(k)}) = M^{-1}N x^{(k)} + M^{-1}b$$

dove

$$M = D - E, N = F.$$

Si vede che la matrice di iterazione $P = M^{-1}N$ è in questo caso

$$P = M^{-1}N = (D - E)^{-1}F. \quad (2)$$

Metodo di Gauss-Seidel

Nota.

Se D è non singolare il metodo di Gauss-Seidel

$$x^{(k+1)} = \phi(x^{(k)}) = (D - E)^{-1}Fx^{(k)} + (D - E)^{-1}b$$

può essere descritto come *metodo delle sostituzioni successive*

$$x_i^{(k+1)} = \left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} \right) / a_{ii}. \quad (3)$$

Nota.

Si osservi che se D è non singolare, ovviamente $a_{i,i} \neq 0$ per ogni indice i e quindi la successione è ben definita.

Metodo di Gauss-Seidel

Vediamo perchè questa derivazione. Da $Ax = b$, con $A \in \mathbb{R}^{n \times n}$ abbiamo

$$\sum_{j < i} a_{i,j}x_j + a_{i,i}x_i + \sum_{j > i} a_{i,j}x_j = \sum_j a_{i,j}x_j = b_i, \quad i = 1, \dots, n$$

e quindi evidenziando x_i , supposto $a_{i,i} \neq 0$ deduciamo

$$x_i = \frac{1}{a_{i,i}} \left(b_i - \left(\sum_{j < i} a_{i,j}x_j + \sum_{j > i} a_{i,j}x_j \right) \right), \quad i = 1, \dots, n$$

Quindi note delle approssimazioni della soluzione $x^{(k)} = \{x_j^{(k)}\}$ per $j > i$, e già calcolate $x^{(k+1)} = \{x_j^{(k+1)}\}$ per $j < i$ è naturale introdurre il metodo iterativo

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

Metodo di Gauss-Seidel

Nota.

- *Secondo Forsythe:*

The Gauss-Seidel method was not known to Gauss and not recommended by Seidel.

- *Secondo Gauss, in una lettera a Gerling (26 dicembre 1823):*

One could do this even half asleep or one could think of other things during the computations.

Definizione (Successive over relaxation, (SOR), 1884-1950)

Sia $A = D - E - F \in \mathbb{R}^{n \times n}$ dove

- D è una matrice diagonale, non singolare
- E è triangolare inferiore con elementi diagonali nulli,
- F è triangolare superiore con elementi diagonali nulli.

Sia fissato $x^{(0)} \in \mathbb{R}^n$ e $\omega \in \mathbb{R} \setminus 0$. Il **metodo SOR** genera $\{x^{(k)}\}_k$ con

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

dove

$$M = \frac{D}{\omega} - E, \quad N = \left(\frac{1}{\omega} - 1 \right) D + F$$

Nota. ((cf.[1, p.555]))

SOR deriva dallo descrivere esplicitamente una iter. di Gauss-Seidel

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

e sostituirla con la combinazione convessa

$$x_i^{(k+1)} = \omega z_i^{(k+1)} + (1 - \omega) x_i^{(k)}.$$

cioè

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

Per convincersi dell'equivalenza tra formulazione matriciale ed equazione per equazione, osserviamo che

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

diventa

$$x^{(k+1)} = \omega D^{-1} [b + E x^{(k+1)} + F x^{(k)}] + (1 - \omega) x^{(k)}$$

e dopo noiosi calcoli in cui si evidenziano $x^{(k)}$ e $x^{(k+1)}$

$$\frac{D}{\omega} (I - \omega D^{-1} E) x^{(k+1)} = \frac{D}{\omega} (\omega D^{-1} F + (1 - \omega)) x^{(k)} + b$$

da cui, dovendo essere $M x^{(k+1)} = N x^{(k)} + b$,

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

Nota.

Gauss-Seidel è SOR in cui si è scelto $\omega = 1$.

Nota.

Notiamo che le iterazioni di SOR verificano l'uguaglianza

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

con

$$M = \frac{D}{\omega} - E, \quad N = \left(\frac{1}{\omega} - 1\right)D + F$$

ed è

$$M - N = \left(\frac{D}{\omega} - E\right) - \left(\left(\frac{1}{\omega} - 1\right)D + F\right) = -E + D - F = A.$$

Quindi effettivamente $A = M - N$.

I metodi di Richardson stazionari

Definizione (Residuo)

Sia A una matrice quadrata $n \times n$, $b \in \mathbb{R}^n$, $x \in \mathbb{R}^n$. La quantità

$$r = b - Ax \quad (4)$$

è nota come **residuo** (relativamente ad A e b).

Definizione (Metodo di Richardson stazionario, 1910)

Fissato α , un metodo di *Richardson stazionario*, con *matrice di preconditionamento* P , verifica

$$P(x^{(k+1)} - x^{(k)}) = \alpha r^{(k)}. \quad (5)$$

dove il residuo alla k -sima iterazione è

$$r^{(k)} = b - Ax^{(k)}. \quad (6)$$

I metodi di Richardson non stazionari

Definizione (Metodo di Richardson non stazionario)

Fissati α_k dipendenti dalla iterazione k , un metodo di **Richardson non stazionario**, con *matrice di preconditionamento* P , verifica

$$P(x^{(k+1)} - x^{(k)}) = \alpha_k r^{(k)}. \quad (7)$$

Nota.

Si osservi che se $\alpha_k = \alpha$ per ogni k , allora il metodo di Richardson non stazionario diventa stazionario.

I metodi di Richardson

Proposizione.

I metodi di Jacobi e di Gauss-Seidel, SOR, sono metodi di Richardson stazionari.

Dimostrazione.

I metodi di Jacobi e di Gauss-Seidel, SOR, sono metodi iterativi del tipo

$$Mx^{(k+1)} = Nx^{(k)} + b, \quad (8)$$

per opportune scelte delle matrici M (che dev'essere invertibile), N tali che $A = M - N$. Essendo $r^{(k)} = b - Ax^{(k)}$,

$$M(x^{(k+1)} - x^{(k)}) = Nx^{(k)} + b - Mx^{(k)} = b - Ax^{(k)} = r^{(k)}. \quad (9)$$

I metodi di Richardson

Ne consegue che i metodi di Jacobi e di Gauss-Seidel, SOR, verificano

$$M(x^{(k+1)} - x^{(k)}) = r^{(k)} \quad (10)$$

In altri termini sono dei metodi di Richardson sono metodi di Richardson stazionari, con $\alpha = 1$ e matrice di preconditionamento $P = M$. △

Per quanto riguarda i **metodi di Richardson preconditionati e non stazionari**, un classico esempio è il **metodo del gradiente classico** che vedremo in seguito.

Norma di matrici

Definizione (Raggio spettrale)

Siano $\lambda_1, \dots, \lambda_n$ gli autovalori di matrice $A \in \mathbb{R}^{n \times n}$. La quantità

$$\rho(A) := \max_i (|\lambda_i|)$$

è detto *raggio spettrale* di A .

Definizione (Norma naturale)

Sia $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}_+$ una norma vettoriale. Definiamo *norma naturale* (o indotta) di una matrice $A \in \mathbb{R}^{n \times n}$ la quantità

$$\|A\| := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Nota.

Questa definizione coincide con quella di norma di un operatore lineare e continuo in spazi normati.

Norma di matrici

Esempio ([4], p. 24)

Sia x un arbitrario elemento di R^n , $A \in R^{n \times n}$.

| <i>Norma vettoriale</i> | <i>Norma naturale</i> |
|---|--|
| $\ x\ _1 := \sum_{k=1}^n x_k $ | $\ A\ _1 = \max_j \sum_{i=1}^n a_{i,j} $ |
| $\ x\ _2 := (\sum_{k=1}^n x_k ^2)^{1/2}$ | $\ A\ _2 = \rho^{1/2}(A^T A)$ |
| $\ x\ _\infty := \max_k x_k $ | $\ A\ _\infty = \max_i \sum_{j=1}^n a_{i,j} $ |

Nota.

- Nella norma matriciale 1, di ogni colonna si fanno i moduli di ogni componente e li si somma. Quindi si considera il massimo dei valori ottenuti, al variare della riga.
- Nella norma matriciale ∞ , di ogni riga si fanno i moduli di ogni componente e li si somma. Quindi si considera il massimo dei valori ottenuti, al variare della colonna.

Norma di matrici

Per quanto riguarda un esempio chiarificatore in Matlab/Octave

```
>> A=[1 5; 7 13]
A =
     1     5
     7    13
>> norm(A,1)
ans =
    18
>> norm(A,inf)
ans =
    20
>> norm(A,2)
ans =
    15.5563
>> eig(A*A')
ans =
     2
    242
>> sqrt(242)
ans =
    15.5563
>> raggio_spettrale_A=max(abs(eig(A)))
raggio_spettrale_A =
    15.4261
>>
```

Norma di matrici

Si dimostra che (cf. [4, p.28])

Teorema

Per ogni norma naturale $\| \cdot \|$ e ogni matrice quadrata A si ha $\rho(A) \leq \|A\|$. Inoltre per ogni matrice A di ordine n e per ogni $\epsilon > 0$ esiste una norma naturale $\| \cdot \|$ tale che

$$\rho(A) \leq \|A\| \leq \rho(A) + \epsilon.$$

e inoltre (cf. [4, p.29], [3, p.232])

Teorema (Hensel 1926, Oldenburger 1940, Householder 1958)

Fissata una norma naturale $\| \cdot \|$, i seguenti asserti sono equivalenti

- 1** $A^m \rightarrow 0$;
- 2** $\|A^m\| \rightarrow 0$;
- 3** $\rho(A) < 1$.

Sul raggio spettrale

Nota.

- Siano $\{\lambda_k\}_{k=1,\dots,n}$ autovalori di $A \in \mathbb{R}^{n \times n}$. Il **raggio spettrale**

$$\rho(A) = \max_k (|\lambda_k|)$$

non è una norma. Infatti la matrice

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

ha raggio spettrale nullo, ma non è la matrice nulla.

- Osserviamo che dagli esempi il raggio spettrale di una matrice A **non coincide in generale con la norma 1, 2, ∞** , ma che a volte $\rho(A) = \|A\|_2$ come nel caso di una matrice diagonale A (essendo gli autovalori di una matrice diagonale, proprio i suoi elementi diagonali).

Convergenza dei metodi iterativi stazionari

Definizione (Matrice diagonalizzabile)

Una matrice $A \in \mathbb{R}^{n \times n}$ è *diagonalizzabile* se e solo se \mathbb{R}^n possiede una *base composta di autovettori* di A .

Equivalentemente $A \in \mathbb{R}^{n \times n}$ è diagonalizzabile se e solo se è *simile ad una matrice diagonale* \mathcal{D} , cioè esiste S tale che

$$A = S^{-1}\mathcal{D}S.$$

Definizione (Metodo consistente)

Sia A una matrice non singolare e $Ax^* = b$. Il metodo stazionario

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

si dice *consistente* relativamente al problema $Ax = b$ se e solo se

$$x^* = Px^* + c.$$

Convergenza dei metodi iterativi stazionari

Lemma

Sia $x^{(k+1)} = Px^{(k)} + c$ un metodo iterativo stazionario.

Supponiamo sia consistente, relativamente al problema $Ax = b$, cioè

$$Ax^* = b \text{ implica } x^* = Px^* + c.$$

Allora, posto $e^{(k)} = x^{(k)} - x^{(*)}$ si ha che $e^{(k+m)} = P^m e^{(k)}$. In particolare, per $k = 0$ si ottiene $e^{(m)} = P^m e^{(0)}$.

Dimostrazione.

Basta osservare che

$$\begin{aligned} e^{(k+m)} &:= x^{(k+m)} - x^* = Px^{(k+m-1)} + c - (Px^* + c) \\ &= P(x^{(k+m-1)} - x^*) = \dots = P^m(x^{(k)} - x^*) = P^m e^{(k)}. \end{aligned} \quad (11)$$

Convergenza dei metodi iterativi stazionari

Teorema

Se P è diagonalizzabile allora un metodo iterativo stazionario consistente $x^{(k+1)} = Px^{(k)} + c$ **converge** per ogni vettore iniziale x_0 se e solo se $\rho(P) < 1$.

Dimostrazione.

Siccome P è diagonalizzabile allora \mathbb{R}^n possiede una base composta di autovettori di P .

Consideriamo un metodo iterativo stazionario $x^{(k+1)} = Px^{(k)} + c$ in cui scelto $x^{(0)}$ si abbia

$$e^{(0)} := x^{(0)} - x^* = \sum_{s=1}^n c_s u_s$$

dove $\{u_k\}_k$ è una base di autovettori di P avente autovalori $\{\lambda_k\}_k$.

Convergenza dei metodi iterativi stazionari

Supponiamo $|\lambda_s| < 1$ per $s = 1, \dots, n$.

Se il metodo è consistente, cioè

$$x^* = Px^* + c$$

da un lemma precedente abbiamo $e^{(k)} = P^k e^{(0)}$ e da $Pu_s = \lambda_s u_s$,

$$x^{(0)} - x^* = \sum_{s=1}^n c_s u_s$$

$$\begin{aligned} P^k e^{(0)} &= P^k \left(\sum_{s=1}^n c_s u_s \right) = \sum_{s=1}^n c_s P^k u_s = \sum_{s=1}^n c_s P^{k-1} P u_s \\ &= \sum_{s=1}^n c_s \lambda_s P^{k-1} u_s = \sum_{s=1}^n c_s \lambda_s^2 P^{k-2} u_s = \dots = \sum_{s=1}^n c_s \lambda_s^k u_s. \end{aligned}$$

Convergenza dei metodi iterativi stazionari

Dimostrazione.

Quindi se $|\lambda_s^k| < 1$ per ogni $s = 1, \dots, n$ e $k = 1, 2, \dots$, abbiamo

$$\|x^{(k)} - x^*\| = \left\| \sum_{s=1}^n c_s \lambda_s^k u_s \right\| \leq \sum_{s=1}^n |c_s| |\lambda_s^k| \|u_s\| \rightarrow 0.$$

Se invece per qualche k si ha $|\lambda^k| \geq 1$ e $c_k \neq 0$ allora $\|x^{(k)} - x^*\|$ non converge a 0 al crescere di k .

Infatti, se $\lambda_l \geq 1$ è l'autovalore di massimo modulo, posto

$$x^{(0)} - x^* = u_l$$

abbiamo da (11)

$$\|x^{(k)} - x^*\| = \|\lambda_l^k u_l\| = |\lambda_l^k| \|u_l\|$$

che non tende a 0. Di conseguenza non è vero che il metodo è convergente per qualsiasi scelta del vettore $x^{(0)}$. △

Convergenza dei metodi iterativi stazionari

Introduciamo il teorema di convergenza nel caso generale, basato sul teorema di Hensel.

Teorema

Un metodo iterativo stazionario consistente $x^{(k+1)} = Px^{(k)} + c$ converge per ogni vettore iniziale x_0 se e solo se $\rho(P) < 1$.

Nota.

- *Si noti che il teorema riguarda la convergenza **per ogni vettore iniziale** x_0 ed è quindi di convergenza globale.*
- *Non si richiede che la matrice P sia diagonalizzabile.*

Convergenza dei metodi iterativi stazionari

Dimostrazione. ([3], p. 236)

\Leftrightarrow Se $\rho(P) < 1$, allora $x = Px + c$ ha una e una sola sol. x^* .

Infatti,

$$x = Px + c \Leftrightarrow (I - P)x = c$$

e la matrice $I - P$ ha autovalori $1 - \lambda_k$ con $k = 1, \dots, n$ tali che

$$0 < |1 - |\lambda_k|_{\mathbb{C}}|_{\mathbb{R}} \leq |1 - \lambda_k|_{\mathbb{C}},$$

poichè $|\lambda_k|_{\mathbb{C}} \leq \rho(P) < 1$ e quindi

$$\det(I - P) = \prod_{k=1}^n (1 - \lambda_k) \neq 0,$$

per cui la matrice $I - P$ è invertibile e *il sistema $(I - P)x = c$ ha una e una sola soluzione x^* .*

Convergenza dei metodi iterativi stazionari

Dimostrazione.

Sia $e(k) = x^{(k)} - x^*$. Come stabilito dal Teorema che lega norme a raggi spettrali, sia inoltre una norma naturale $\|\cdot\|$ tale che

$$\rho(P) \leq \|P\| = \rho(P) + (1 - \rho(P))/2 < 1.$$

Da

- $x^{(k+1)} = Px^{(k)} + c,$

- $x = Px + c,$

da un lemma precedente $e^{(k)} = P^k e^{(0)}$ e quindi essendo $\|\cdot\|$ una norma naturale

$$\|e^{(k)}\| = \|P^k e^{(0)}\| \leq \|P^k\| \|e^{(0)}\|. \quad (12)$$

Essendo $\rho(P) < 1$, abbiamo che $\|P^k\| \rightarrow 0$ da cui per (12) è $\|e^{(k+1)}\| \rightarrow 0$ e quindi per le proprietà delle norme $e^{(k)} \rightarrow 0$ cioè $x^{(k)} \rightarrow x^*$.

Convergenza dei metodi iterativi stazionari

Dimostrazione.

⇒ Supponiamo che la successione $x^{(k+1)} = Px^{(k)} + c$ converga a x^* per qualsiasi $x^{(0)} \in \mathbb{R}^n$ ma che sia $\rho(P) \geq 1$. Sia λ_{\max} il massimo autovalore in modulo di P e scegliamo $x^{(0)}$ tale che $e^{(0)} = x^{(0)} - x^*$ sia autovettore di P relativamente all'autovalore λ_{\max} . Essendo $Pe^{(0)} = \lambda_{\max}e^{(0)}$ e $e^{(k+1)} = P^k e^{(0)}$ abbiamo che

$$e^{(k+1)} = \lambda_{\max}^k e^{(0)}$$

da cui, qualsiasi sia la norma $\|\cdot\|$, per ogni $k = 1, 2, \dots$ si ha

$$\|e^{(k)}\| = |\lambda_{\max}^k| \|e^{(0)}\| \geq \|e^{(0)}\|$$

il che comporta che la successione non è convergente (altrimenti per qualche k sarebbe $e^{(k)} < e^{(0)}$). △

Velocità di convergenza

Proposito.

L'analisi che faremo in questa sezione non è rigorosamente matematica. Ciò nonostante permette di capire il legame tra il raggio spettrale della matrice di iterazione P e la riduzione dell'errore.

Supponiamo di voler approssimare x^* tale che $Ax^* = b$ con il metodo stazionario

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

Supporremo tale metodo consistente e quindi se $Ax^* = b$ allora

$$x^* = Px^* + c.$$

Inoltre indicheremo con $e^{(k)}$ l'errore

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

Velocità di convergenza

- Da un lemma precedente

$$e^{(k+m)} = P^m e^{(k)}. \quad (13)$$

- Si dimostra che se $A \in \mathbb{C}^{n \times n}$ e $\|\cdot\|$ è una norma naturale allora

$$\lim_k \|A^k\|^{1/k} = \rho(A)$$

Quindi per k sufficientemente grande si ha $\|P^k\|^{1/k} \approx \rho(P)$
ovvero $\|P^k\| \approx \rho^k(P)$.

Velocità di convergenza

Sotto queste ipotesi, per k sufficientemente grande, da

- $e^{(k+m)} = P^m e^{(k)}$,
- $\|P^k\| \approx \rho^k(P)$,

abbiamo

$$\|e^{(k+m)}\| = \|P^m e^{(k)}\| \leq \|P^m\| \|e^{(k)}\| \approx \rho^m(P) \|e^{(k)}\| \quad (14)$$

e quindi

$$\|e^{(k+m)}\| / \|e^{(k)}\| \lesssim \rho^m(P). \quad (15)$$

Per cui da

$$\rho^m(P) \leq \epsilon \Rightarrow \|e^{(k+m)}\| / \|e^{(k)}\| \lesssim \rho^m(P) \leq \epsilon$$

applicando il logaritmo naturale ad ambo i membri, per avere una riduzione dell'errore di un fattore ϵ , asintoticamente servono al più m iterazioni con

$$m \log(\rho(P)) \approx \log(\epsilon)$$

e quindi

$$m \approx \log(\epsilon) / \log(\rho(P)).$$

Velocità di convergenza

Se

$$R(P) = -\log(\rho(P))$$

è la cosiddetta **velocità di convergenza asintotica** del metodo iterativo relativo a P abbiamo

$$m \approx \left\lceil \frac{\log \epsilon}{\log(\rho(P))} \right\rceil = \left\lceil \frac{-\log(\epsilon)}{R(P)} \right\rceil.$$

Se P è la matrice d'iterazione di un metodo stazionario convergente (e consistente), essendo $\rho(P) < 1$, minore è $\rho(P)$ necessariamente è maggiore $R(P)$ e si può *stimare* il numero di iterazioni per ridurre l'errore di un fattore ϵ .

Si desidera quindi cercare metodi

- metodi con m minore possibile o equivalentemente
- $\rho(P)$ più piccolo possibile o equivalentemente
- con $R(P)$ il più grande possibile.

Alcuni teoremi di convergenza

Definizione (Matrice tridiagonale)

Una matrice A si dice **tridiagonale** se $A_{i,j} = 0$ qualora $|i - j| > 1$.

Esempio:

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 1 & 5 & 7 & 0 & 0 \\ 0 & 2 & 0 & 9 & 0 \\ 0 & 0 & 4 & 4 & 2 \\ 0 & 0 & 0 & 5 & 3 \end{pmatrix}$$

Alcuni teoremi di convergenza

Teorema

Per matrici *tridiagonali* $A = (a_{i,j})$ con *componenti diagonali non nulle*, i metodi di *Jacobi e Gauss-Seidel* sono

- o entrambi convergenti
- o entrambi divergenti.

La velocità di convergenza del metodo di *Gauss-Seidel* è il *doppio* di quello del metodo di *Jacobi*.

Nota.

Questo significa vuol dire che asintoticamente sono necessarie metà iterazioni del metodo di *Gauss-Seidel* per ottenere la stessa precisione del metodo di *Jacobi*.

Alcune definizioni

Definizione (Matrice a predominanza diagonale)

La matrice A è a *predominanza diagonale* (per righe) se per ogni $i = 1, \dots, n$ risulta

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

e per almeno un indice s si abbia

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

La matrice A è a predominanza diagonale per colonne se A^T a predominanza diagonale per righe.

Alcune definizioni

Definizione (Matrice a predom. diagonale in senso stretto)

Se

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

allora la matrice A si dice a **predominanza diagonale (per righe) in senso stretto**.

La matrice A è a **predominanza diagonale (per colonne) in senso stretto** se A^T a predominanza diagonale per righe in senso stretto.

Ad esempio la matrice

$$A = \begin{pmatrix} 4 & -4 & 0 \\ -1 & 4 & -1 \\ 0 & -4 & 4 \end{pmatrix}$$

è a predominanza diagonale per righe (non stretta).

Teoremi di convergenza

Teorema

Sia A una matrice quadrata a *predominanza diagonale per righe in senso stretto*. Allora il metodo di *Jacobi converge* alla soluzione di $Ax = b$, qualsiasi sia il punto $x^{(0)}$ iniziale.

Teorema

Sia A è a *predominanza diagonale per righe in senso stretto*. Allora il metodo di *Gauss-Seidel converge*, qualsiasi sia il punto $x^{(0)}$ iniziale.

- Tali teoremi valgono anche se A è a predominanza diagonale per colonne in senso stretto.
- Si basano sul teorema di convergenza dei metodi iterativi stazionari.

Alcune definizioni

Definizione (Matrice simmetrica)

La matrice A è *simmetrica* se $A = A^T$.

Definizione (Matrice definita positiva)

Una matrice simmetrica A è *definita positiva* se ha tutti gli autovalori positivi.

Equivalentemente una matrice simmetrica A è definita positiva se

$$x^T Ax > 0, \text{ per ogni } x \neq 0.$$

Alcune definizioni

Esempio

Ad esempio la matrice

$$A = \begin{pmatrix} 4 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 4 \end{pmatrix}$$

è simmetrica e definita positiva. Per convincersene basta vedere che ha tutti gli autovalori positivi.

```
>> A=[4 -1 0; -1 4 -1; 0 -1 4];  
>> eig(A) % AUTOVALORI DI A.  
ans =  
    2.5858  
    4.0000  
    5.4142  
>>
```

Alcuni teoremi di convergenza

Teorema (Kahan)

Sia A una matrice quadrata. Condizione necessaria affinché il metodo SOR converga globalmente alla soluzione di $Ax = b$ è che sia $|1 - \omega| < 1$. Se $\omega \in \mathbb{R}$ in particolare è che sia $\omega \in (0, 2)$.

Teorema (Ostrowski)

Sia A

- *simmetrica,*
- *con elementi diagonali positivi.*

Allora il metodo SOR converge se e solo se

- $0 < \omega < 2,$
- *A è definita positiva.*

Per una dimostrazione si veda [?, p.215].

Test di arresto

Problema. (Test di arresto)

Consideriamo il sistema lineare $Ax = b$ avente un'unica soluzione x^* e supponiamo di risolverlo numericamente con un metodo iterativo stazionario del tipo

$$x^{(k+1)} = Px^{(k)} + c,$$

che sia consistente cioè

$$x^* = Px^* + c.$$

Desideriamo introdurre un **test di arresto** che interrompa le iterazioni, qualora una certa quantità relativa al sistema lineare $Ax = b$ e alle iterazioni eseguite, sia al di sotto di una tolleranza $\epsilon > 0$ fissata dall'utente.

Test di arresto: criterio dello step

Proposizione. (Stima a posteriori)

Sia $Ax = b$ con A non singolare e $x^{(k+1)} = Px^{(k)} + c$ un metodo iterativo stazionario consistente, con $\|P\| < 1$. Allora

$$\|e^{(k)}\| := \|x^* - x^{(k)}\| \leq \frac{\|x^{(k+1)} - x^{(k)}\|}{1 - \|P\|}.$$

Dimostrazione.

Posto $\Delta^{(k)} := x^{(k+1)} - x^{(k)}$ e $e^{(k)} = x^* - x^{(k)}$, essendo $e^{(k)} = Pe^{(k-1)}$ abbiamo

$$\begin{aligned}\|e^{(k)}\| &= \|x^* - x^{(k)}\| = \|(x^* - x^{(k+1)}) + (x^{(k+1)} - x^{(k)})\| \\ &= \|e^{(k+1)} + \Delta^{(k)}\| = \|Pe^{(k)} + \Delta^{(k)}\| \leq \|P\| \cdot \|e^{(k)}\| + \|\Delta^{(k)}\|\end{aligned}$$

da cui

$$(1 - \|P\|)\|e^{(k)}\| \leq \|\Delta^{(k)}\| \Leftrightarrow \|e^{(k)}\| := \|x^* - x^{(k)}\| \leq \frac{\|x^{(k+1)} - x^{(k)}\|}{1 - \|P\|}.$$



Test di arresto: criterio dello step

Fissata dall'utente una tolleranza tol , si desidera interrompere il processo iterativo quando

$$\|x^* - x^{(k)}\| \leq tol.$$

Definizione (Criterio dello step)

Il **criterio dello step**, consiste nell'interrompere il metodo iterativo che genera la successione $\{x^{(k)}\}$ alla $k + 1$ -sima iterazione qualora

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

(dove tol è una tolleranza fissata dall'utente).

Di seguito desideriamo vedere quando tale criterio risulti attendibile cioè $|x^{(k+1)} - x^{(k)}| \approx |x^* - x^{(k)}|$.

Test di arresto: criterio dello step

Teorema

Sia $Ax^* = b$ con A non singolare e $x^{(k+1)} = Px^{(k)} + c$ un metodo iterativo stazionario

- consistente,
- convergente,
- con P simmetrica.

Allora

$$\|e^{(k)}\|_2 := \|x^* - x^{(k)}\|_2 \leq \frac{\|x^{(k+1)} - x^{(k)}\|_2}{1 - \rho(P)}.$$

Test di arresto: criterio dello step

Lemma

Se P è *simmetrica*, allora esistono

- una matrice ortogonale U , cioè tale che $U^T = U^{-1}$,
- una matrice diagonale a coefficienti reali Λ

per cui

$$P = U\Lambda U^T$$

ed essendo P e Λ simili, le matrici P e Λ hanno gli stessi autovalori $\{\lambda_k\}_k$.

Test di arresto: criterio dello step

Dimostrazione.

Dal lemma, se P è simmetrica, essendo

- $\|P\|_2 := \sqrt{\rho(PP^T)},$

- $U^T U = I$

$$\begin{aligned}\|P\|_2 &= \sqrt{\rho(PP^T)} = \sqrt{\rho(U\Lambda U^T(U\Lambda U^T)^T)} \\ &= \sqrt{\rho(U\Lambda U^T U\Lambda U^T)} = \sqrt{\rho(U\Lambda^2 U^T)}\end{aligned}\quad (16)$$

Essendo $U\Lambda^2 U^T$ simile a Λ^2 , $U\Lambda^2 U^T$ e Λ^2 hanno gli stessi autovalori uguali a $\{\lambda_k^2\}_k$ e di conseguenza lo stesso raggio spettrale, da cui

$$\rho(U\Lambda^2 U^T) = \rho(\Lambda^2).$$

Test di arresto: criterio dello step

Da $\rho(U\Lambda^2 U^T) = \rho(\Lambda^2)$ ricaviamo

$$\begin{aligned}\|P\|_2 &= \sqrt{\rho(\Lambda^2)} = \sqrt{\max_k |\lambda_k^2|} = \sqrt{\max_k |\lambda_k|^2} \\ &= \sqrt{(\max_k |\lambda_k|)^2} = \max_k |\lambda_k| = \rho(P).\end{aligned}\quad (17)$$

Se il metodo stazionario $x^{(k+1)} = Px^{(k)} + c$ è convergente, per il teorema di convergenza $\|P\|_2 = \rho(P) < 1$ e per la precedente proposizione

$$\|e^{(k)}\|_2 := \|x^* - x^{(k)}\|_2 \leq \frac{\|x^{(k+1)} - x^{(k)}\|_2}{1 - \|P\|_2} = \frac{\|x^{(k+1)} - x^{(k)}\|_2}{1 - \rho(P)}.\quad \triangle$$

Nota.

Il teorema afferma che se la matrice di iterazione P , di un metodo consistente e convergente, è simmetrica, il criterio dello step è affidabile se $\rho(P) < 1$ è piccolo.

Test di arresto: criterio del residuo

Definizione (Criterio del residuo)

Il **criterio del residuo**, consiste nell'interrompere il metodo iterativo che genera la successione $\{x^{(k)}\}$ alla $k + 1$ -sima iterazione qualora

$$\|r^{(k)}\| \leq tol$$

dove

- tol è una tolleranza fissata dall'utente,
- $r^{(k)} = b - Ax^{(k)}$.

Test di arresto: criterio del residuo

Teorema

Sia $Ax^* = b$ con A non singolare e $\|\cdot\|$ una norma naturale. Sia $\{x^{(k)}\}$ la successione generata da un metodo iterativo. Allora

$$\frac{\|x^{(k)} - x^*\|}{\|x^*\|} = \frac{\|e^{(k)}\|}{\|x^*\|} \leq \kappa(A) \frac{\|r^{(k)}\|}{\|b\|}.$$

dove $\kappa(A)$ è il numero di condizionamento di A .

Nota.

Dal teorema si evince che il criterio $\frac{\|r^{(k)}\|}{\|b\|} \leq \text{tol}$ non offre indicazioni su $\frac{\|x^{(k)} - x^*\|}{\|x^*\|}$ per $\kappa(A) \gg 1$.

Test di arresto: criterio del residuo

Dimostrazione.

Se $b = Ax^*$ allora, essendo A invertibile

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

e quindi, essendo $\|\cdot\|$ una norma naturale

$$\|e^{(k)}\| \leq \|A^{-1}\| \|r^{(k)}\|;$$

Inoltre, poichè $b = Ax^*$ abbiamo $\|b\| \leq \|A\| \|x^*\|$ e quindi

$$\frac{1}{\|x^*\|} \leq \frac{\|A\|}{\|b\|}.$$

Di conseguenza, posto $\kappa(A) = \|A\| \|A^{-1}\|$, se $x^* \neq 0$ ricaviamo

$$\frac{\|e^{(k)}\|}{\|x^*\|} \leq \frac{\|A\|}{\|b\|} \cdot \|A^{-1}\| \|r^{(k)}\| = \kappa(A) \frac{\|r^{(k)}\|}{\|b\|}.$$



I metodi di discesa

Sia A **una matrice simmetrica definita positiva**.

Si osserva che se x^* è l'unica soluzione di $Ax = b$ allora è pure il minimo del **funzionale dell'energia**

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

in quanto

$$\text{grad}(\phi(x)) = Ax - b = 0 \Leftrightarrow Ax = b.$$

Quindi invece di calcolare la soluzione del sistema lineare, intendiamo calcolare il minimo del funzionale ϕ .

I metodi di discesa

Un generico **metodo di discesa** consiste nel generare una successione

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

dove $p^{(k)}$ è una direzione fissata secondo qualche criterio.

Lo scopo ovviamente è che

$$\phi(x^{(k+1)}) < \phi(x^{(k)}),$$

e che il punto x^* , in cui si ha il minimo di ϕ , venga calcolato rapidamente.

I metodi di discesa

Definizione (Metodo del gradiente classico, Cauchy 1847)

Il metodo del **gradiente classico** è un metodo di discesa

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

con α_k e $p^{(k)}$ scelti così da ottenere la massima riduzione del funzionale dell'energia a partire dal punto $x^{(k)}$.

Differenziando $\phi(x) = \frac{1}{2}x^T Ax - b^T x$, $x \in \mathbb{R}^n$ si mostra che tale scelta coincide con lo scegliere

$$p^{(k)} = r^{(k)}, \quad \alpha_k = \frac{\|r^{(k)}\|_2^2}{(r^{(k)})^T A r^{(k)}}. \quad (18)$$

Nota.

Con qualche facile conto si vede che è un **metodo di Richardson non stazionario** con $P = I$ e parametro α_k definito da (18).

Il metodo del gradiente coniugato.

Definizione (Gradiente coniugato, Hestenes-Stiefel, 1952)

Il metodo del *gradiente coniugato* è un metodo di discesa

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}$$

con

- $\alpha_k = \frac{(r^{(k)})^T r^{(k)}}{(p^{(k)})^T A p^{(k)}}$,
- $p^{(0)} = r^{(0)}$,
- $p^{(k)} = r^{(k)} + \beta_k p^{(k-1)}$, $k = 1, 2, \dots$
- $\beta_k = \frac{(r^{(k)})^T r^{(k)}}{(r^{(k-1)})^T r^{(k-1)}} = \frac{\|r^{(k)}\|_2^2}{\|r^{(k-1)}\|_2^2}$.

Con questa scelta si prova che $p^{(k)}$ e $p^{(k-1)}$ sono *A-coniugati*.

$$(p^{(k)})^T A p^{(k-1)} = 0.$$

Il metodo del gradiente coniugato

Nota. (Brezinski, Wuytack)

Hestenes was working at the I.N.A. at UCLA. In June and early July 1951, he derived the conjugate gradient algorithm.

When, in August, Stiefel arrived at I.N.A. from Switzerland for a congress, the librarian gave him the routine written by Hestenes. Shortly after, Stiefel came to Hestenes office and said about the paper "this is my talk!".

Indeed, Stiefel had invented the same algorithm from a different point of view. So they decided to write a joint paper from a different point of view.

Nota. (Saad, van der Vorst)

The anecdote told at the recent conference ... by Emer. Prof. M. Stein, the post-doc who programmed the algorithm for M. Hestenes the first time, is that Stiefel was visiting UCLA at the occasion of a conference in 1951. Hestenes then a faculty member at UCLA, offered to demonstrate this effective new method to Stiefel, in the evening after dinner. Stiefel was impressed by the algorithm. After seeing the deck of cards he discovered that this was the same method as the one he had developed independently in Zurich. Stiefel also had an assistant, by the name of Hochstrasser, who programmed the method.

Il metodo del gradiente coniugato: proprietà

Il metodo del gradiente coniugato ha molte proprietà particolari. Ne citiamo alcune.

Teorema

Se A è una matrice simmetrica e definita positiva di ordine n , allora il metodo del gradiente coniugato è convergente e fornisce in aritmetica esatta la soluzione del sistema $Ax = b$ in al massimo n iterazioni.

Questo teorema tradisce un po' le attese, in quanto

- in generale i calcoli non sono compiuti in aritmetica esatta,
- in molti casi della modellistica matematica n risulta essere molto alto.

Il metodo del gradiente coniugato: proprietà

Definizione (Spazio di Krylov)

Lo spazio

$$\mathcal{K}_k = \text{span}(r^{(0)}, Ar^{(0)}, \dots, A^{k-1}r^{(0)})$$

per $k \geq 1$ si dice *spazio di Krylov*.

Teorema

Sia

$$\mathcal{K}_k = \text{span}(r^{(0)}, Ar^{(0)}, \dots, A^{k-1}r^{(0)})$$

per $k \geq 1$. Allora la k -sima iterata dal metodo del gradiente coniugato, minimizza il funzionale ϕ nell'insieme $x^{(0)} + \mathcal{K}_k$.

Per una dimostrazione si veda [7, p.12]. Si osservi che essendo la k -sima iterazione del gradiente classico pure in $x^{(0)} + \mathcal{K}_k$, il gradiente classico non minimizza in generale il funzionale ϕ in $x^{(0)} + \mathcal{K}_k$.

Il metodo del gradiente coniugato: proprietà

Teorema

Se

- A è *simmetrica e definita positiva*,

- $\|x\|_A = \sqrt{x^T A x}$,

- $e^{(k)} = x^* - x^{(k)}$, con $x^{(k)}$ iterazione del gradiente coniugato

allora (cf. [6, p.530])

$$\|e^{(k)}\|_A \leq 2 \left(\frac{\sqrt{K_2(A)} - 1}{\sqrt{K_2(A)} + 1} \right)^k \|e^{(0)}\|_A.$$

Il metodo del gradiente coniugato: proprietà

Nota.

Questo risultato stabilisce che la convergenza del gradiente coniugato è lenta qualora si abbiano alti numeri di condizionamento

$$K_2(A) := \|A\|_2 \|A^{-1}\|_2 = \frac{\max_i |\lambda_i|}{\min_j |\lambda_j|}, \quad \{\lambda_i\} = \text{spettro}(A).$$

Posto

$$\mu(x) = \frac{\sqrt{x} - 1}{\sqrt{x} + 1}$$

abbiamo ad esempio

- $\mu(10) \approx 0.5195$,
- $\mu(1000) \approx 0.9387$,
- $\mu(100000) \approx 0.9937$.

Bibliografia



K. Atkinson, *Introduction to Numerical Analysis*, Wiley, 1989.



K. Atkinson e W. Han, *Theoretical Numerical Analysis*, Springer, 2001.



D. Bini, M. Capovani e O. Menchi, *Metodi numerici per l'algebra lineare*, Zanichelli, 1988.



V. Comincioli, *Analisi Numerica, metodi modelli applicazioni*, Mc Graw-Hill, 1990.



S.D. Conte e C. de Boor, *Elementary Numerical Analysis, 3rd Edition*, Mc Graw-Hill, 1980.



G.H. Golub, C.F. Van Loan *Matrix Computations*, third edition, The John Hopkins University Press, 1996.



C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, 1995.



A. Quarteroni e F. Saleri, *Introduzione al calcolo scientifico*, Springer Verlag, 2006.