

# Metodi iterativi per la soluzione di sistemi lineari

Alvise Sommariva

Università degli Studi di Padova  
Dipartimento di Matematica Pura e Applicata

11 gennaio 2012

Supponiamo che siano  $A \in \mathbb{R}^{n \times n}$  (matrice non singolare),  $b \in \mathbb{R}^n$  (vettore colonna) e di dover 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.

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. Quindi il costo totale sarà di ordine  $O(\bar{k} \cdot n^2)$ .

# 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  $Mx - Nx = Ax = b$  e quindi  $Mx = Nx + b$ . Moltiplicando ambo i membri per  $M^{-1}$  e posto  $\phi(x) = M^{-1}Nx + b$  abbiamo  $x = M^{-1}Nx + M^{-1}b = \phi(x)$ . 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**.

Quale utile notazione, sia inoltre  $A = D - E - F$  con  $D$  la matrice diagonale estratta da  $A$ ,  $E$ ,  $F$  rispettivamente triangolari inferiore e superiore.

Il **metodo di Jacobi** fu scoperto nel 1845, nell'ambito di alcune ricerche su problemi di piccole oscillazioni che comportavano alla risoluzione di sistemi lineari con matrici diagonalmente dominanti. Nel caso del metodo di Jacobi si ha

$$M = D, N = E + F \quad (1)$$

e quindi

$$\begin{aligned} P &= M^{-1}N = D^{-1}(E + F) = D^{-1}(D - D + E + F) \\ &= D^{-1}(D - A) = I - D^{-1}A \end{aligned} \quad (2)$$

Si osservi che se  $D$  è non singolare allora il metodo di Jacobi, almeno in questa versione di base, non può essere utilizzato visto che in (5) non ha senso la scrittura  $D^{-1}$ .

# Metodo di Jacobi

Qualora sia  $a_{ii} \neq 0$  per ogni  $i = 1, \dots, n$ , il metodo di Jacobi può essere descritto come

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

L'idea è che se  $Ax = B$  allora  $\sum_{j=1}^n a_{s,j}x_j = b$  allora

$$a_{i,i}x_i = b - \sum_{j=1}^{i-1} a_{s,j}x_j - \sum_{j=i+1}^n a_{s,j}x_j$$

da cui (3) ricordando il metodo delle appross. successive, qualora  $a_{i,i} \neq 0$ . Si osservi che ogni componente di  $x^{(k+1)}$  dipende dai soli valori all'iterazione precedente  $x^{(k)}$ , e questo rende il metodo facilmente implementabile in computers paralleli. Da (3) si capisce perchè tale metodo è noto anche come **metodo delle sostituzioni simultanee**.

Il **metodo di Gauss-Seidel** fu scoperto nel 1874, da studi preliminari di Gauss (1823) completati dal suo allievo Seidel per lo studio di problemi ai minimi quadrati del tipo  $Sx = f$  con  $S$  non quadrata, che venivano risolti quali soluzione del sistema di equazioni normali  $S^T Sx = S^T f$ . Mentre Gauss oltre a problemi di Astronomia era interessato a problemi di Geodesia (triangolazione di Hannover usando una *catena* di 26 triangoli), Seidel si interessava alla risoluzione di un sistema di equazioni con 72 incognite per uno studio di luminosità stellare. Il metodo di Gauss-Seidel è definito quale metodo stazionario in cui

$$M = D - E, N = F \quad (4)$$

e quindi

$$P = M^{-1}N = (D - E)^{-1}F \quad (5)$$

Similmente al metodo di Jacobi, possiamo riscrivere più semplicemente anche Gauss-Seidel come

$$x_i^{(k+1)} = \left( b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^n a_{ij}x_j^{(k)} \right) / a_{ii}. \quad (6)$$

Da (6) si capisce perchè tale metodo è noto anche come **metodo delle sostituzioni successive**.

- ▶ Siano  $\{\lambda_k\}_{k=1,\dots,n}$  gli autovalori di una matrice  $A$ . Si definisce **raggio spettrale**  $\rho(A)$  la quantità

$$\max_{k=1,\dots,n} |\lambda_k|.$$

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

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

Per ogni norma naturale  $\|\cdot\|$  si ha  $\rho(A) \leq \|A\|$ .

- ▶ Un metodo iterativo stazionario  $x^{(k+1)} = Px^{(k)} + c$  è **consistente** per la soluzione di  $Ax = b$  se per  $x^*$  tale che  $Ax^* = b$  si ha  $x^* = Px^* + c$ .

Una matrice  $A$  si dice **diagonalizzabile** se vale una qualsiasi di queste (equivalenti) affermazioni:

- ▶ è simile a una matrice diagonale, cioè  $A = S^{-1}DS$  con  $D$  matrice diagonale.
- ▶ esiste una base di autovettori  $\{u_k\}$  (con autovalori  $\{\lambda_k\}$ ).
- ▶ tutti gli autovalori di  $A$  sono distinti;

**NB:** Una matrice  $A$  simmetrica (cioè  $A = A^T$ ) è diagonalizzabile.

# Teorema di Hensel

**Teorema.** Se  $P$  è diagonalizz. allora un metodo iter. staz. cons.  $x^{(k+1)} = Px^{(k)} + c$  converge per ogni vett. iniziale  $x_0 \leftrightarrow \rho(P) < 1$ .

( $\Leftarrow$ ) Se la matrice  $P$  è diagonalizz. allora esiste una base di autovett.  $\{u_k\}$  con autovalori  $\{\lambda_k\}$ . Poichè il metodo è consistente

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

Scelto arbitrariamente  $x^0$ , abbiamo per certi  $c_1, \dots, c_n$  che

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

e quindi da  $x^{(k)} = Px^{(k-1)} + c$  e  $x^* = Px^* + c$ , per linearità

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

Ma allora, ripetendo più volte tale ragionamento

$$\begin{aligned}x^{(k)} - x^* &= P(x^{(k-1)} - x^*) = P(P(x^{(k-2)} - x^*)) \\ &= P^2(x^{(k-2)} - x^*) = \dots = P^k(x^0 - x^*).\end{aligned}$$

Quindi

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

Così da  $|\lambda_s^k| \leq (\rho(P))^k < 1$  per ogni  $s = 1, \dots, n$  e  $k = 1, 2, \dots$ ,

$$\begin{aligned}\|x^{(k)} - x^*\| &= \left\| \sum_{s=1}^n c_s \lambda_s^k u_s \right\| \leq \sum_{s=1}^n |c_s| |\lambda_s^k| \|u_s\| \\ &\leq (\rho(P))^k \sum_{s=1}^n |c_s| \|u_s\| \rightarrow 0\end{aligned}$$

( $\Rightarrow$ ) Supponiamo per assurdo che per qualche  $k$  sia  $|\lambda_s| \geq 1$ .  
Posto  $c \neq 0$  sia  $x^{(0)} = x^* + cu_s$  con  $Pu_s = \lambda_s u_s$ . Similmente a quanto visto in precedenza:

$$\|x^{(k)} - x^*\| = \|c\lambda_s^k u_s\| \leq |\lambda_s^k| \|cu_s\| \rightarrow \infty$$

Di conseguenza abbiamo trovato un punto iniziale  $x^{(0)}$  per cui il metodo non converge, il che è assurdo.

Esiste una versione più generale del precedente teorema (per  $P$  non diagonalizzabili).

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

## Alcuni teoremi di convergenza

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

Esempio:

$$\begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 1 & 5 & 7 & 0 & 0 \\ 0 & 2 & 6 & 9 & 0 \\ 0 & 0 & 4 & 4 & 2 \\ 0 & 0 & 0 & 5 & 3 \end{pmatrix}$$

**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 divergenti e il tasso di convergenza del metodo di Gauss-Seidel è il doppio di quello del metodo di Jacobi (il che vuol dire che *asintoticamente* sono necessarie metà iterazioni del metodo di Gauss-Seidel per ottenere la stessa precisione del metodo di Jacobi).

## Alcune definizioni

Ricordiamo che  $A$  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

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

Se per tutti gli indici  $s$  si ha  $|a_{s,s}| > \sum_{j=1, j \neq s}^n |a_{s,j}|$  allora la predominanza diagonale è 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

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

Tali teoremi valgono se  $A^T$  è a predominanza diagonale per righe in senso stretto (cioè  $A$  è a predominanza diagonale per colonne in senso stretto).

## Alcune definizioni

La matrice  $A$  è **simmetrica** se  $A = A^T$ . Una matrice  $A$  è **definita positiva** se ha tutti gli autovalori positivi. Ad esempio la matrice

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

è simmetrica e definita positiva:

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

Ricordiamo che equivalentemente una matrice  $A$  è definita positiva se

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

**Teorema.** Sia  $A$  una matrice simmetrica, non singolare con elementi principali  $a_{i,i} \neq 0$ . Allora il metodo di Gauss-Seidel è convergente per qualsiasi scelta del punto iniziale  $x^{(0)}$  se e solo se  $A$  è definita positiva.

Una classica famiglia di metodi iterativi è quella dei 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

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

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. Il gradiente coniugato è un classico esempio di metodo di discesa.

## Il metodo del gradiente coniugato.

Il metodo del grad. coniugato fu descritto nel 1952 da Hestenes e Stiefel ma non venne molto utilizzato fino al 1971, quando Reid suggerì il suo utilizzo per la risoluzione di sistemi sparsi (cioè con molte componenti nulle) di grandi dimensioni. Il metodo *non* è iterativo stazionario.

La succ. delle iterazioni del gradiente coniugato è quella propria dei metodi di discesa,

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}, \quad \alpha_k = \frac{(r^{(k)})^T r^{(k)}}{(p^{(k)})^T A p^{(k)}}$$

dove  $p^{(0)} = r^{(0)}$  e

$$p^{(k)} = r^{(k)} + \beta_k p^{(k-1)}, \quad \beta_k = \frac{(r^{(k)})^T r^{(k)}}{(r^{(k-1)})^T r^{(k-1)}}.$$

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: proprietà

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

- ▶ Se  $A$  è una matrice simmetrica e definita positiva di ordine  $n$ , si può dimostrare che il metodo 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, sia perchè in generale i calcoli non sono compiuti in aritmetica esatta, sia perchè in molti casi della modellistica matematica  $n$  risulta essere molto alto.

# Il metodo del gradiente coniugato: proprietà

Si può dimostrare che se  $A$  è simmetrica e definita positiva,

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

e

$$e_k = x^* - x^{(k)}$$

allora

$$\|e_k\|_A \leq \left( \frac{\sqrt{\kappa_2(A)} - 1}{\sqrt{\kappa_2(A)} + 1} \right)^{2k} \|e_0\|_A.$$

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

$$\kappa_2(A) := \|A\|_2 \|A^{-1}\|_2 = \frac{\max_j |\lambda_j|}{\min_j |\lambda_j|}$$

(ove al solito  $\{\lambda_j\}$  sono gli autovalori di  $A$ ).

# Gradiente coniugato in Matlab

Per quanto riguarda il codice del Gradiente Coniugato, un esempio è il file `cg.m` tratto da Netlib:

```
function [x,error,iter,flag]=cg(A,x,b,M,max_it,tol)
flag = 0; iter = 0; bnorm2 = norm( b );
if ( bnorm2 == 0.0 ), bnorm2 = 1.0; end
r = b - A*x; error = norm( r ) / bnorm2;
if ( error < tol ) return, end
for iter = 1:max_it
    z = M \ r; rho = (r'*z);
    if ( iter > 1 )
        beta = rho / rho_1; p = z + beta*p;
    else
        p = z;
    end
    q = A*p; alpha = rho / (p'*q); x = x + alpha * p;
    r = r - alpha*q; error = norm( r ) / bnorm2;
    if ( error <= tol ), break, end
    rho_1 = rho;
end
if ( error > tol ) flag = 1; end
```

- ▶ Ha Matlab una sua versione del gradiente coniugato? Se la risposta è negativa cercare in internet una versione del gradiente coniugato per Matlab.
- ▶ Matlab dispone di una gallery di matrici. Aiutandosi con  
`|>> help gallery`

si calcoli la matrice di Poisson  $P_{20}$  di ordine 20.

- ▶ Sia  $b$  il vettore composto di componenti uguali a 1, avente lo stesso numero di righe di  $P_{20}$ . Si risolva col gradiente coniugato il problema  $P_{20}x = b$ . Quante iterazioni servono?