Consiglio Nazionale delle Ricerche

ISTITUTO DI ELABORAZIONE DELLA INFORMAZIONE

PISA

METODI DIRETTI PER LA RISOLUZIONE DI SISTEMI LINEARI

M. Arioli, A. Laratta

Nota interna B82-09

Luglio 1982

METODI DIRETTI PER LA RISOLUZIONE DI SISTEMI LINEARI(1)

M. Arioli, A. Laratta(2)

Abstract

Numerical methods are studied, based on ortogonal decomposition of matrices, for the computation of the solution of an underdeterminate system of linear equations of least distance from an assigned point.

These methods are especially useful for the computation of the general solution.

(1) Lavoro sviluppato nell'ambito del Progetto Finalizzato Informatica del C.N.R. (Sottoprogetto P1, Sofmat)

(2) Istituto di Elaborazione della Informazione, C.N.R., Via S. Maria 46 - 56100 Pisa

l'inversa generalizzata di Penrose di A^T; con

$$N_{A^T} = A^{T+} A^T$$

la matrice di proiezione sulla varietà ortogonale a $A^T x=0$ e con

$$P_{A^T} = I - N_{A^T}$$

la matrice di proiezione su $A^T x = 0$, la (3) assume la forma

$$\hat{x} = P_{A^T} x^\circ + A^{T+} b \quad .$$

Quindi $\hat{x}$ è la somma della proiezione di x° sulla varietà lineare $A^T x = 0$ e della soluzione di minima norma del sistema $A^T x = b$. Inoltre la (3) fornisce la soluzione generale di (1) se x° è un generico n-vettore.

Supponiamo ora $m \lesseqgtr n$ e A^T di rango k con $k \leq m$ $(k \leq n)$. Se si riesce ad estrarre da (1) il sistema

$$A'^T x = b'$$

di k equazioni linearmente indipendenti, la soluzione $\hat{x}$ si ottiene da (3) sostituendo A' ad A.

I metodi numerici che studieremo nei prossimi paragrafi controllano se il sistema (1) è inconsistente e in caso contrario calcolano la soluzione $\hat{x}$ e il rango della matrice A^T.

2. Perturbazione dei dati

Vogliamo ora studiare come si propagano le perturbazioni dei dati A, b, x sulla soluzione di (2).

Esaminiamo il caso in cui $m \leq n$ e A^T di massimo rango; se $m \lesseqgtr n$ e A^T di rango k con $k \leq m$ $(k \leq n)$ i risultati che si ottengono si applicano al sistema:

$$A'^T x = b'$$

di k equazioni linearmente indipendenti estratto da (1).

Richiamiamo innanzitutto alcune proprietà della inversa generalizzata di una m x n matrice B^T di rango m e alcuni Lemmi allo scopo di dimostrare il teorema che fornisce una maggiorazione dell'errore della soluzione del problema (2). Valgono le seguenti proprietà [1]

$$(4) \qquad \begin{cases} B^{T+} = B^{+T} \\ B^T B^{T+} = I \\ B^T N_{B^T} = B^T \\ B^T P_{B^T} = 0 \\ N_{B^T} B^{T+} = B^{T+} \end{cases}$$

Indichiamo con

$$\tilde{A} = A + \Delta A$$

$$\tilde{b} = b + \Delta b$$

$$\tilde{x}° = x° + \Delta x°$$

$$G = \tilde{A}^{T+} - A^{T+}$$

$$K = \| A^+ \| \; \| A \| \quad \text{(numero di condizionamento di } A^T\text{)}$$

e con $\hat{x} + \Delta \hat{x}$ la soluzione del problema perturbato

$$\min \| x - \tilde{x}° \|^2$$

$$\tilde{A}^T x = \tilde{b}$$

Lemma 1

Se rango $(\tilde{A})$ = rango $(A) = m$

si ha

$$\| GA^T \| \leq (\| \tilde{A}^+ \| + \| A^+ \|) \| \Delta A \|$$

Dimostrazione. In [1] si dimostra che

$$G = \tilde{A}^{T+} \Delta A^T A^{T+} + P_{\tilde{A}^T} \Delta A A^+ A^{T+} \quad ;$$

la tesi si ottiene utilizzando le proprietà (4) e ricordando che

$$\| N_{\tilde{A}^T} \| = \| N_{A^T} \| = \| P_{\tilde{A}^T} \| = \| P_{A^T} \| = 1.$$

Il seguente lemma è dimostrato in [1].

Lemma 2

Se rango $(\tilde{A})$ = rango $(A) = m$ e $\| A^+ \| \| \Delta A \| < 1$, allora

$$\| \tilde{A}^+ \| \leq \frac{\| A^+ \|}{1 - \| A^+ \| \| \Delta A \|}$$

Teorema

Nelle ipotesi del lemma 2, si ha

$$\frac{\| \Delta \hat{x} \|}{\| \hat{x} \|} \leq \hat{K} \left(2\alpha + \gamma \beta + 3\alpha \frac{\| x^\circ \|}{\| \hat{x} \|} \right) + \delta \frac{\| x^\circ \|}{\| x \|}$$

dove:

$$\hat{K} = \frac{K}{1 - K \dfrac{\| \Delta A \|}{\| A \|}}$$

$$\alpha = \frac{\| \Delta A \|}{\| A \|}, \quad \beta = \frac{\| \Delta b \|}{\| b \|}, \quad \gamma = \frac{\| b \|}{\| A \| \| \hat{x} \|} \leq 1, \quad \delta = \frac{\| \Delta x^\circ \|}{\| x^\circ \|}$$

Dimostrazione

Da

$$\hat{x} = A^{T+} b + P_{A^T} x^o$$

$$\hat{x} + \Delta \hat{x} = \tilde{A}^{T+}(b + \Delta b) + P_{\tilde{A}^T}(x^o + \Delta x^o)$$

si ottiene

$$\Delta \hat{x} = GA^T(\hat{x} - x^o) - \tilde{A}^{T+} \Delta A^T x^o + P_{\tilde{A}^T} \Delta x^o + \tilde{A}^{T+} \Delta b$$

La tesi segue applicando il lemma 1 e il lemma 2.

Osserviamo che la maggiorazione della perturbazione della soluzione dipende linearmente dal numero di condizionamento di A^T, mentre, come è noto, per sistemi sopradeterminati tale dipendenza è quadratica [1].

Se $x^o = 0$, si ottiene il risultato dato in [1] relativo alla soluzione di minima norma.

3. Metodi per il calcolo della soluzione.

Scriviamo ora la soluzione $\hat{x}$ attraverso una decomposizione ortogonale della matrice A allo scopo di introdurre metodi di soluzione stabili. Supponiamo $m \leq n$. E' noto che esiste una $m \times m$ matrice triangolare superiore R e una $n \times m$ matrice a colonne mutuamente ortogonali Q, tali che

(5) $\quad A = QR$

Se con a_j, q_j ($q_i^T q_j = 0$, $i \neq j$), r_{ij}($r_{ij} = 0$ $i > j$) si indicano rispettivamente le colonne di A, le colonne di Q e gli elementi di R da (5) si ricava

(6) $$a_i = \sum_{k=1}^{i} r_{ki} \quad q_k \qquad , \qquad i=1,\dots,m$$

e da questa

(7) $$q_j^T q_j \; r_{ji} = q_j^T \; a_i \; , \qquad i=1,\dots,m \quad ; \; j \leqq i$$

Inoltre è noto che, la decomposizione (5) è unica se si fissano gli elementi $r_{ii} \neq 0$ oppure i moduli dei q_i ($q_i^T q_i \neq 0$). Se a_i è linearmente dipendente dagli a_j, j=1, ..., i-1, da (7) segue

$$r_{ii} q_i^T q_i = 0$$

Quindi se si fissano i valori di $r_{ii} \neq 0$ oppure di $q_i^T q_i \neq 0$, si hanno rispettivamente $q_i = 0$ oppure $r_{ii} = 0$.

Una caratterizzazione della matrice Q si può ottenere osservando che, se si indica con $P_i (P_o = I)$ il proiettore costruito con la matrice $(A^{(i)})^T$ formata dalle prime i righe di A^T, $(P_i = P_{(A^{(i)})^T})$ si ha

$$P_i = I - \sum_{j=1}^{i} (q_j q_j^T)/(q_j^T q_j),$$

$$P_{i-1} q_k = 0 \quad k = 1,\dots, i-1,$$

$$P_{i-1} \; q_i = q_i$$

da cui segue per la (6)

(8) $$P_{i-1} \; a_i = r_{ii} \; q_i$$

Se a_i è linearmente indipendente dagli a_j, j=1,..,i-1, $P_{i-1} \; a_i \neq 0$ e quindi anche $r_{ii} \; q_i \neq 0$. Si può allora dire che nella decomposizione (5) le colonne di Q sono date da

(9) $$q_i = (1/r_{ii}) \; P_{i-1} \; a_i \quad i = 1,\dots, m, \quad P_o = I$$

e se si fissa $r_{ii} = 1$, i=1,..., m

(10) $\quad q_i = P_{i-1}\, a_i, \; i = 1,\ldots, m, \quad P_o = I.$

Supponiamo A^T di rango massimo;con la decomposizione (5) la varietà $A^T x=b$ è descritta anche da

(11) $\quad Q^T x = d, \quad d = (R^T)^{-1} b$

e $\hat{x}$ è la soluzione del problema

$$\min_{Q^T x = d} \| x - x^\circ \|^2$$

La (3) si scrive in questo caso

$$(12)\quad \begin{cases} \hat{x} = (I-Q(Q^TQ)^{-1}Q^T)x^\circ + Q(Q^TQ)^{-1}d \\ d = (R^T)^{-1}b \end{cases}$$

da cui si ottiene anche

$$(13)\quad \begin{cases} \hat{x} = x^\circ - \sum_{j=1}^{m} (q_j^T x^\circ)/(q_j^T q_j)q_j + \sum_{j=1}^{m} d_j/(q_j^T q_j)q_j \\ \\ d_1 = b_1/r_{11}, \quad d_j = (1/r_{jj})(b_j - \sum_{j=1}^{j-1} r_{kj} d_k) \qquad j = 2,\ldots,m. \end{cases}$$

Le d_i si possono calcolare con le formule ricorrenti:

$$(14)\quad \begin{cases} d_1 = b_1/r_{11} \qquad r_{11} = (q_i^T a_i)/(q_i^T q_i) \\ \left.\begin{array}{l} d_i^{(1)} = b_i \\ d_i^{(j+1)} = d_i^{(j)} - r_{ji} d_j, \quad r_{ji} = (q_j^T a_i)/(q_j^T q_j) \quad j=1,\ldots,i-1 \\ d_i = d_i^{(i)}/r_{ii} \quad r_{ii} = (q_i^T a_i)/(q_i^T q_i) \end{array}\right\} i=2,\ldots,m \end{cases}$$

oppure

$$
(15)\quad \begin{cases}
d_i^{(1)} = b_i & i = 1,\dots,m \\
d_j = d_j^{(j)}/r_{jj} \qquad r_{jj} = (q_j^T a_j)/(q_j^T q_j) & \\
d_i^{(j+1)} = d_i^{(j)} - r_{ji} d_j \qquad r_{ji} = (q_j^T a_i)/(q_j^T q_j) \quad i=j+1,..,m & \Big\} \; j=1,..,m-1 \\
d_m = d_m^{(m)}/r_{mm} &
\end{cases}
$$

Sia ora

$$
A^{(i)^T} x = b^{(i)} \qquad\qquad i=1,2,\dots,m
$$

Il sistema formato dalle prime i equazioni di (1). Il problema

$$
\begin{gathered}
\min \| x-x^\circ \|^2 \\
A^{(i)^T} x = b^{(i)}
\end{gathered}
$$

ha soluzione unica x^i $(\hat{x}=x^m)$

$$
(16)\quad \begin{cases}
x^i = x^\circ - \sum\limits_{j=0}^{i} ((q_j^T x^\circ)/(q_j^T q_j)) q_j + \sum\limits_{j=0}^{i} (d_j/q_j^T q_j) q_j \qquad i=1,..,m \\
d_1 = b_1/r_{11}, \; d_j = b_j - \sum\limits_{k=1}^{j-1} r_{kj} d_k \qquad j=2,..,i
\end{cases}
$$

La prima delle (16) può anche essere scritta

$$
(17)\qquad x^i = x^{i-1} - (q_i^T x^\circ)/(q_i^T q_i)\; q_i + (d_i/q_i^T q_i) q_i \qquad i=1,\dots,m
$$

da cui si ricava, poichè $q_j^T x^i = d_j \qquad j=1,\dots,i,$

$$
(18)\qquad q_i^T x^\circ = q_i^T x^{i-1} \quad .
$$

Dalle (17), (18) e dalla seconda delle (16) si ottiene il procedimento per il calcolo delle x^i, $i=1,\dots, m$, $x^m=\hat{x}$

$$(19)\quad \left.\begin{cases} x^i=x^{i-1}-(r^i(x^{i-1})/(q_i^Tq_i))q_i \\ r^i(x^{i-1})=q_i^Tx^{i-1}-d_i \\ d_1=b_1/r_{11},\ d_i=(1/r_{ii})(b_i-\sum_{j=1}^{i-1} r_{ji}d_j) \end{cases}\right\}\quad i=1,\dots, m$$

L'interpretazione geometrica delle (19) è la seguente: x^{i-1} e q_{i-1} appartengono rispettivamente alle varietà parallele $Q^{(i-1)T}x=d^{(i-1)}$ (sistema formato dalle prime i-1 equazioni di $Q^Tx=d$) e $Q^{(i-1)T}x=0$; x^i, che appartiene a $Q^{(i)T}x=d^{(i)}$, è l'intersezione della retta passante per x^{i-1} e di direzione q_{i-1} con l'iperpiano $q_i^T x=d_i$.

Vogliamo ora ricavare da (19) un algoritmo per la risoluzione del problema (2) in cui figurano i dati A e b invece di Q e d.

Indichiamo con $R(x)$ e $R^i(x)$ i residui

$$R(x)=A^Tx-b=R^T(Q^Tx-d)$$

$$R^i(x)=a_i^Tx-b_i=\sum_{j=1}^{i} r_{ji}(q_j^Tx-d_j)$$

Poichè

$$q_j^Tx^{i-1}-d_j=0 \qquad j=1,2,\dots,i-1$$

da (6) e (11) si ottiene

$$(20)\qquad R^i(x^{i-1})=r_{ii}r^i(x^{i-1})$$

Huang, per il calcolo della soluzione di minima norma, ha ricavato le (26) dalle (22).

Il confronto fra i vari metodi verrà fatto nei paragrafi seguenti; per ora, in relazione ai metodi (25) e (27) osserviamo che Rice [4] ha provato sperimentalmente che il procedimento di ortogonalizzazione di Gram-Schmidt è meno stabile di quello di Gram-Schmidt modificato. Una ragione di questo si può avere, [1], osservando che i due procedimenti eseguono le stesse operazioni e differiscono nel calcolo degli elementi di R; in Gram-Schmidt

$$r_{ij}=(q_i^T a_j)/(q_i^T q_i);$$

in Gram-Schmidt modificato

$$r_{ij}=(q_i^T a_j^{(i)})/(q_i^T q_i)$$

e si ha evidentemente

$$\| a_j^{(i)} \| \leq \| a_j \| \qquad\qquad j=i,\dots,m$$

Sarebbe interessante riuscire a fare una analisi dell'errore degli algoritmi introdotti, ad esempio vedere se la soluzione calcolata è soluzione esatta di un problema perturbato.

Qui osserviamo che il passaggio dalle formule (12) alle (13) dalle quali sostanzialmente si costruiscono gli algoritmi di questo paragrafo, presuppone che le colonne di Q siano mutuamente ortogonali. Sembra quindi importante che i valori numerici $\overline{q}_i$ di q_i siano tali che se $\overline{q}_i^T \overline{q}_j = \varepsilon_{ij}$ e $\varepsilon = \max |\varepsilon_{ij}|$ sia $\varepsilon \simeq 0$.

D'altra parte è noto che nel procedimento di ortogonalizzazione di Gram-Schmidt modificato ci si può aspettare che la deviazione della ortogonalità della matrice calcolata Q dipenda essenzialmente dal numero di condizionamento della matrice A: al crescere del numero di condizionamento di A, ε cresce [5].

Nel procedimento di ortogonalizzazione di Householder [1], invece, la deviazione dell'ortogonalità non dipende dal numero di condizionamento di A.

Queste osservazioni suggeriscono di prendere in considerazione nelle (19) anche il procedimento di ortogonalizzazione di Householder.

5. Un metodo costruito con la decomposizione di Householder.

Col procedimento di ortogonalizzazione di Householder si costruisce una nxn matrice ortogonale H, una mxm matrice triangolare superiore R e una (n-m)xm matrice nulla 0 tali che

$$(28) \qquad HA=\begin{bmatrix} R \\ 0 \end{bmatrix}$$

La decomposizione

$$A=QR$$

si ottiene da (28) indicando

$$(29) \qquad H=\begin{bmatrix} Q^T \\ S^T \end{bmatrix}$$

Le $q_i (q_i^T q_i=1, q_i^T q_j=0, i\neq j)$ sono quindi le prime m righe di H.

Un metodo per ottenere la decomposizione (28) è descritta in [6]. Esso consiste nella applicazione di una successione di matrici di rotazione per annullare gli elementi al di sotto della diagonale della matrice A. Se si pone $A^{(1)}=A$ e si calcola

$$A^{(k+1)}=P^k A^{(k)} \quad (k=1,\dots, m)$$

Si ottiene

$$A^{(m+1)} = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

$$H=P^m P^{m-1} \dots P^1$$

I P^k sono definiti da

$$P^k=I-c_k u^{(k)} u^{(k)^T} \tag{30}$$

e i c_k e $u^{(k)}$ si determinano col seguente algoritmo:

$$\left\{\begin{array}{ll} v_k=(\sum_{i=k}^{m} (a_{ik}^{(k)})^2)^{1/2} & \\ c_k=1/(v_k(v_k+|a_{kk}^{(k)}|)) & \\ u_i^{(k)}=0 & i<k \\ u_k^{(k)}=\mathrm{sgn}(a_{kk}^{(k)})(v_k+|a_{kk}^{(k)}|) & \\ u_i^{(k)}=a_{ik}^{(k)} & i>k \end{array}\right. \tag{31}$$

in cui con $a_{ij}^{(k)}$ si sono indicati gli elementi di $A^{(k)}$.

Scriviamo ora l'algoritmo (19) in cui le q_i, $i=1,\dots, m$, sono le prime m righe di H. Se le $u^{(h)}$ sono calcolate con l'algoritmo (31), e con h_k si indica la k-sima riga di H si ottiene

$$
(32)\quad \left\{ \begin{array}{lll}
A^{(1)}=A & & \\
H^{(1)}=I & & \\
A^{(k+1)}=A^{(k)}-u^{(k)}(c_k u^{(k)^T} A^{(k)}) & & \\
H^{(k+1)}=H^{(k)}-u^{(k)}(c_k u^{(k)^T} H^{(k)}) & & \\
q_k=h_k & & \\
r_{jk}=a^{(k)}_{jk} & j=1,\ldots,k & \\
d_1=b_1/r_{11} & & \text{se } k=1 \\
d_k=(b_k-\sum_{j=1}^{k-1} r_{jk}d_j)/r_{kk} & & \text{se } k>1 \\
r^k(x^{k-1}) = q_k^T x^{k-1}-d_k & & \\
x^k=x^{k-1}-r^k(x^{k-1})q_k & &
\end{array} \right\} \quad k=1,\ldots,m
$$

6. <u>Considerazioni numeriche</u>.

Esaminiamo alcuni aspetti numerici degli algoritmi esposti nei paragrafi precedenti di cui si deve tenere conto in fase di implementazione.

a) Iniziamo col riportare nella tabella seguente i valori asintotici (per n grande) del numero di operazioni (una operazione = una moltiplicazione + una addizione) dei metodi (23), (25), (27) e (32) nei casi $n \gg m$ e $n \gtrsim m$

	(23)	(25)	(27)	(32)
$n \gg m$	$2\,n^2 m$	$n\,m(m+2)$	$nm(m+2)$	$4n^2 m$
$n \gtrsim m$	$2\,n^2 m$	nm^2	nm^2	$(4/3)m^3$

b) L'algoritmo (23) fornisce alla fine del calcolo anche il proiettore $P_{A^T} = P_m$. Se si applicano gli algoritmi (25), (27), (32) P_{A^T} può essere calcolato con la formula

$$P_{A^T}=I-\sum_{j=1}^{m} (q_j q_j^T)/(q_j^T q_j)$$

Il numero di operazioni che si debbono eseguire è, per n grande, $n^2 m$

c) Vogliamo vedere come gli errori nel calcolo di q_i e d_i si propaghino in un passo dell'algoritmo (19). Sia

$$q_i^T q_i=1$$

e supponiamo di conoscere il valore esatto di x^{i-1}

$$q_j^T x^{i-1}=d_j \qquad j=1,\dots, i$$

Indichiamo con

$$\tilde{q}_i=q_i+\varepsilon_i \qquad \tilde{d}_i=d_i+\delta_i \qquad ;$$

da

$$x^i=x^{i-1}-(q_i^T x^{i-1}-d_i)q_i$$
$$\tilde{x}^i=x^{i-1}-(\tilde{q}_i x^{i-1}-\tilde{d}_i)\tilde{q}_i$$

si ottiene

$$\tilde{x}^i-x^i \doteq -r^i(x^{i-1})\,\varepsilon_i-\gamma_i(x^{i-1})q_i \tag{33}$$

in cui

$$\gamma_i(x^{i-1})=(\tilde{q}_i^T x^{i-1}-\tilde{d}_i)-(q_i^T x^{i-1}-d_i)=\varepsilon_i^T x^{i-1}-\delta_i$$

e $\doteq$ sta ad indicare che il segno di uguaglianza vale solo in prima approssimazione.

Da (33) discendono le due disuguaglianze

$$\|\tilde{x}^i - x^i\| \overset{\bullet}{\leq} |r^i(x^{i-1})| \|\varepsilon_i\| + |\gamma_i(x^{i-1})|$$

$$\|\tilde{x}^i - x^i\| \overset{\bullet}{\leq} (\|x^i - x^{i-1}\| + \|x^{i-1}\|)\|\varepsilon_i\| + |\delta_i|.$$

dove $\overset{\bullet}{\leq}$ sta ad indicare che $\leq$ vale solo in prima approssimazione

Se in (19) è

$$q_i^T q_i \neq 1$$

si ottengono le relazioni

$$\tilde{x}^i - x^i \doteq -(r^i(x^{i-1})/(q_i^T q_i))\varepsilon_i - (\gamma_i(x^{i-1})/(q_i^T q_i))q_i + \\ +2((\varepsilon_i^T q_i) r^i(x^{i-1})/(q_i^T q_i)^2) q_i$$

$$\|\tilde{x}^i - x^i\| \overset{\bullet}{\leq} 3\ (|r^i(x^{i-1})|/\|q_i\|^2)\|\varepsilon_i\| + |\gamma_i(x^{i-1})|/\|q_i\|$$

$$\|\tilde{x}^i - x^i\| \overset{\bullet}{\leq} (1/\|q_i\|)\left\{(3\|x^i - x^{i-1}\| + \|x^{i-1}\|)\|\varepsilon_i\| + |\delta_i|\right\}.$$

Naturalmente, ricordando l'interpretazione geometrica di (19), non sorprende che nel fattore di amplificazione di ε_i figuri la distanza di x^i da x^{i-1}.

d) Se il sistema (1) è consistente e A^T ha caratteristica k=n, per quanto detto in c, può essere conveniente utilizzare la soluzione calcolata come nuovo punto iniziale e riapplicare il metodo. In questo caso i valori asintotici delle operazioni riportate nella tabella del punto c non cambiano, perchè il procedimento d'ortogonalizzazione è eseguito una sola volta. Se la caratteristica di A^T è minore di n e si riapplica il metodo, si ottiene una soluzione che il generale non è la soluzione di minima distanza da x°.

e) Per la (20), nei metodi (25) e (27) può essere sostituito $r^i(x^{i-1})$ con $R^i(x^{i-1})$ e in (32) $r^i(x^{i-1})$ con $R^i(x^{i-1})/r_{ii}$, evitando così di risolvere il sistema triangolare per il calcolo di d. I valori asintotici delle operazioni riportate nella tabella del punto c non cambiano, però si opera su dati affetti solo da errori di arrotondamento.

Ciò nonostante i risultati numerici sconsigliano tali modifiche; questo può dipendere dal fatto che il calcolo numerico del prodotto scalare $a_i^T x^{i-1}$ è affetto da un errore generalmente più grande del calcolo di $q_i^T x^{i-1}$ in quanto

$$\| q_i \| \leq \| a_i \|$$

7. <u>Esperimenti numerici</u>.

Abbiamo provato gli algoritmi (23), (25), (27) e (32) su sistemi generati a caso di m equazioni in n incognite (con $m \leq n$ e per vari valori di m e n) e su problemi test noti in letteratura.

Per il confronto dei vari metodi abbiamo considerato i due errori

$$E_1 = \| \hat{x} - \bar{x} \| \;/\; \| \hat{x} \|$$

$$E_2 = \max_i \left\{ |\hat{x}_i - \bar{x}_i| \;/\; |\hat{x}_i| \right\}$$

in cui

$$\hat{x} = (\hat{x}_1, \ldots, \hat{x}_n)$$

è la soluzione esatta e

$$\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$$

è la soluzione calcolata.

Abbiamo trovato, come d'altra parte era prevedibile, che gli algoritmi più affidabili sono il (27) e il (32).

I risultati che saranno riportati si riferiscono a due problemi test a matrice quadrata dipendente da un parametro, al variare del quale varia il numero di condizionamento. Ricordando il punto d del paragrafo precedente, gli algoritmi sono stati anche rieseguiti utilizzando la soluzione calcolata come nuovo punto iniziale. Gli esperimenti sono stati eseguiti su calcolatore IBM 3033.

I problemi test sono stati costruiti in doppia precisione (precisione di macchina 16^{-13}) e risolti in precisione semplice (precisione di macchina 16^{-5}).

Esempio 1. [7]

Consideriamo le matrici 2 x 2

$$R=\begin{bmatrix} -\cot\vartheta & \csc\vartheta \\ -\csc\vartheta & \cot\vartheta \end{bmatrix} \qquad S=\begin{bmatrix} 1-\cot\vartheta & \csc\vartheta \\ -\mathrm{cso}\,\vartheta & 1+\cot\vartheta \end{bmatrix} \qquad T=\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$0<\vartheta<\pi/2$$

e con esse costruiamo la matrice A quadrata di ordine pari

$$A=\begin{bmatrix} R & S & T & . & . & . & T \\ S & R & S & T & . & . & T \\ T & . & . & . & . & . & T \\ . & . & . & . & . & . & . \\ T & . & . & . & . & . & T \\ . & . & . & T & S & R & S \\ T & . & . & T & T & S & R \end{bmatrix}$$

Fissiamo il vettore $\hat{x}=(\hat{x}_1,\ldots\hat{x}_n)$, con

$\hat{x}_i=(-1)^i i \qquad i=1,\ldots,n$

e calcoliamo

$$b = A^T \hat{x}$$

Otteniamo così il sistema

$$A^T x = b \tag{34}$$

con soluzione nota $\hat{x}$.

I risultati del calcolo numerico della soluzione di (34) sono stati riassunti nelle tabelle 1,2,3,4.

Si è fissato $n = 20$ e uguali a zero le prime 19 componenti del vettore iniziale.

$$x^\circ=(x^\circ_1,x^\circ_2,..,x^\circ_{20}).$$

Le tabelle 1,2,3,4 si riferiscono alla risoluzione del sistema (34) avendo scelto rispettivamente

$$x^\circ_{20}=0, \quad x^\circ_{20}=10^3, \quad x^\circ_{20},=10^4, \quad x^\circ_2=10^5$$

Nella prima colonna sono indicati i valori di ϑ, mentre nelle colonne 2,3,4,5 sono riportati i risultati relativi ai metodi (27)(GSM), (27) con doppia iterazione (GSM1), (32)(HS), (32) con doppia iterazione (HS1): il primo numero è il valore di E_1 e il secondo il valore di E_2.

E' stato inoltre valutato il reciproco del numero di condizionamento K di A, per vari valori di ϑ , con un codice della libreria LINPACK[8] che implementa un metodo proposto in [9]. Tale metodo da una stima di 1/K senza il calcolo dei valori singolari di A. I valori di 1/K relativi agli ϑ che figurano nelle tabelle 1,2,3,4 sono riportati nella tabella seguente:

ϑ	$1/K$
$\pi/2$	$0{,}105 \quad 10^{-1}$
$\pi/4$	$0{,}675 \quad 10^{-2}$
$\pi/8$	$0{,}840 \quad 10^{-2}$
$\pi/16$	$0{,}820 \quad 10^{-3}$
$\pi/32$	$0{,}248 \quad 10^{-3}$
$\pi/64$	$0{,}694 \quad 10^{-4}$
$\pi/128$	$0{,}186 \quad 10^{-4}$
$\pi/256$	$0{,}483 \quad 10^{-5}$
$\pi/512$	$0{,}122 \quad 10^{-5}$
$\pi/1024$	$0{,}309 \quad 10^{-6}$

Tab. 5

Esempio 2

Si considera la matrice triangolare superiore R, i cui elementi, se ϑ è un numero reale, sono definiti da

$$r_{11}=\vartheta \quad , \quad r_{nn}=1/\vartheta \quad , \quad r_{ii}=1 \qquad i=2,\dots,n-1$$
$$r_{ij}=0 \qquad i>j \quad , \quad r_{ij}=-1 \qquad i<j.$$

Si generano in modo casuale i vettori $v_i (\|v_i\|=1) i=1,\dots,n$ e si costruiscono le matrici ortogonali

$$E_i=I-2\, v_i v_i^T \qquad i=1,\dots,n$$

Si calcola la matrice

$$A=E_n\, E_{n-1}\dots E_1 R \quad .$$

Si fissa la soluzione $\hat{x}$ come nell'esempio 1

$$\hat{x}_i=(-1)^i i \qquad i=1,\dots n$$

e il termine noto

$$b=A^T\hat{x}.$$

Si ottiene così il sistema con soluzione nota

$$A^T x=b \qquad .$$

I risultati riportati si riferiscono alla scelta n=10; la tabella 6 è relativa al caso $x^o=0$, la tabella 7 al caso in cui x^o ha le prime nove componenti nulle e $x^o_{10}=1000$.

Le stime dei reciproci dei numeri di condizionamento, calcolati come nell'esempio 1, sono riportati nella tabella seguente:

ϑ	1/K
0.2	$0.229\ 10^{-3}$
0.4	$0.310\ 10^{-3}$
0.6	$0.347\ 10^{-3}$
0.8	$0.331\ 10^{-3}$
1.0	$0.363\ 10^{-3}$
20	$0.826\ 10^{-5}$
40	$0.138\ 10^{-5}$
60	$0.147\ 10^{5}$
80	$0.116\ 10^{-5}$
100	$0.537\ 10^{-6}$

Tab. 8

8. Risultati numerici.

Dagli esperimenti numerici vengono confermate alcune previsioni esposte nei paragrafi precedenti.

E' stato osservato che la deviazione dall'ortogonalità della matrice calcolata Q non dipende dal numero di condizionamento di A nel procedimento di ortogonalizzazione di Householder, mentre ne dipende nel procedimento di Gram-Schmidt modificato. Da ciò segue, come si vede dalle tabelle dei risultati numerici, che:

1) HS da risultati migliori di GSM al crescere del numero di condizionamento di A specialmente se $\| x-x^{\circ} \|$ è grande.

2) L'applicazione della doppia iterazione migliora più sensibilmente i risultati ottenuti con GSM che con HS.

Un altro risultato che si legge sulle tabelle è che E_1 è generalmente più piccolo in HS che in GSM (anche nei casi in cui E_2 è più grande in HS che in GSM); questo significa che la distribuzione dell'errore sulle singole componenti del vettore soluzione è più uniforme in HS che in GSM.

ALFA	GSM	GSM1	HS	HS1
0.1571E+01	0.1706E-04 0.1068E-03	0.1630E-04 0.1554E-03	0.1975E-04 0.2375E-03	0.1802E-04 0.1755E-03
0.7854E+00	0.4605E-04 0.1325E-02	0.1715E-04 0.2007E-03	0.1776E-04 0.2165E-03	0.2062E-04 0.1733E-03
0.3927E+00	0.2976E-04 0.1969E-03	0.1062E-04 0.5722E-04	0.3092E-04 0.7095E-03	0.2867E-04 0.7372E-03
0.1963E+00	0.2709E-03 0.3054E-02	0.1787E-04 0.2689E-03	0.1251E-03 0.5140E-03	0.1226E-03 0.4965E-03
0.9817E-01	0.3909E-03 0.5575E-02	0.2130E-04 0.1078E-03	0.1713E-03 0.1131E-02	0.1731E-03 0.1101E-02
0.4909E-01	0.1533E-02 0.8225E-02	0.2932E-03 0.5861E-02	0.5067E-03 0.4339E-02	0.5042E-03 0.4335E-02
0.2454E-01	0.5643E-02 0.5142E-01	0.3236E-02 0.4322E-01	0.3307E-02 0.5815E-01	0.3304E-02 0.5818E-01
0.1227E-01	0.4267E-01 0.2570E+00	0.2257E-02 0.1939E-01	0.2238E-01 0.1940E+00	0.2238E-01 0.1940E+00
0.6136E-02	0.8146E-01 0.1730E+01	0.1033E-01 0.1429E+00	0.5801E-01 0.2137E+00	0.5801E-01 0.2137E+00
0.3068E-02	0.3658E+00 0.2874E+01	0.2463E+00 0.1448E+01	0.5381E+01 0.3655E+02	0.5381E+01 0.3655E+02

Tab. 1

ALFA	GSM	GSM1	HS	HS1
0.1571E+01	0.2425E-04 0.5341E-03	0.1628E-04 0.1516E-03	0.8516E-04 0.2258E-03	0.1784E-04 0.1793E-03
0.7854E+00	0.9193E-04 0.1419E-02	0.1708E-04 0.1996E-03	0.1290E-03 0.3227E-03	0.2054E-04 0.1792E-03
0.3927E+00	0.7082E-03 0.1109E-01	0.1067E-04 0.6390E-04	0.1126E-03 0.5646E-03	0.2868E-04 0.7391E-03
0.1963E+00	0.2494E-02 0.2393E-01	0.1794E-04 0.2728E-03	0.1394E-03 0.4985E-03	0.1225E-03 0.4972E-03
0.9817E-01	0.5280E-02 0.2390E-01	0.2121E-04 0.1087E-03	0.1998E-03 0.1465E-02	0.1730E-03 0.1098E-02
0.4909E-01	0.3265E-01 0.8627E-01	0.2932E-03 0.5848E-02	0.4977E-03 0.3952E-02	0.5041E-03 0.4333E-02
0.2454E-01	0.7092E-01 0.6019E+00	0.3236E-02 0.4317E-01	0.3261E-02 0.5750E-01	0.3304E-02 0.5817E-01
0.1227E-01	0.2146E+00 0.1597E+01	0.2454E-02 0.1448E-01	0.2237E-01 0.1933E+00	0.2238E-01 0.1940E+00
0.6136E-02	0.9447E+00 0.9293E+01	0.3595E-01 0.4404E+00	0.5797E-01 0.2135E+00	0.5801E-01 0.2137E+00
0.3068E-02	0.3193E+01 0.1449E+02	0.1783E+01 0.8627E+01	0.5381E+01 0.3655E+02	0.5381E+01 0.3655E+02

Tab. 2

ALFA	GSM	GSM1	HS	HS1
0.1571E+01	0.2135E-03 0.5859E-02	0.1634E-04 0.1574E-03	0.8699E-03 0.3174E-02	0.1792E-04 0.1774E-03
0.7854E+00	0.1061E-02 0.1989E-01	0.1709E-04 0.1953E-03	0.1257E-02 0.4150E-02	0.2062E-04 0.1776E-03
0.3927E+00	0.7202E-02 0.1094E+00	0.1051E-04 0.6104E-04	0.1173E-02 0.3296E-02	0.2861E-04 0.7296E-03
0.1963E+00	0.2718E-01 0.2666E+00	0.1776E-04 0.2728E-03	0.9747E-03 0.2441E-02	0.1226E-03 0.4975E-03
0.9817E-01	0.5567E-01 0.2898E+00	0.2126E-04 0.1087E-03	0.9341E-03 0.5615E-02	0.1731E-03 0.1101E-02
0.4909E-01	0.3256E+00 0.8879E+00	0.2947E-03 0.5728E-02	0.8800E-03 0.2563E-02	0.5041E-03 0.4337E-02
0.2454E-01	0.7352E+00 0.6131E+01	0.3241E-02 0.4265E-01	0.2963E-02 0.5225E-01	0.3304E-02 0.5818E-01
0.1227E-01	0.2313E+01 0.1761E+02	0.1786E-01 0.1017E+00	0.2228E-01 0.1865E+00	0.2238E-01 0.1940E+00
0.6136E-02	0.1011E+02 0.1085E+03	0.3591E+00 0.5691E+01	0.5761E-01 0.2119E+00	0.5801E-01 0.2137E+00
0.3068E-02	0.3335E+02 0.1325E+03	0.1858E+02 0.8413E+02	0.5381E+01 0.3655E+02	0.5381E+01 0.3655E+02

Tab. 3

ALFA	GSM	GSM1	HS	HS1
0.1571E+01	0.2029E-02 0.4541E-01	0.1624E-04 0.1526E-03	0.9141E-02 0.2686E-01	0.1787E-04 0.1707E-03
0.7854E+00	0.1066E-01 0.2075E+00	0.1698E-04 0.1986E-03	0.1154E-01 0.4297E-01	0.2072E-04 0.1773E-03
0.3927E+00	0.7214E-01 0.1086E+01	0.1066E-04 0.5913E-04	0.1071E-01 0.3271E-01	0.2854E-04 0.7334E-03
0.1963E+00	0.2741E+00 0.2679E+01	0.1830E-04 0.2356E-03	0.9278E-02 0.3174E-01	0.1225E-03 0.4972E-03
0.9817E-01	0.5583E+00 0.2945E+01	0.2206E-04 0.9251E-04	0.8711E-02 0.3735E-01	0.1732E-03 0.1101E-02
0.4909E-01	0.3256E+01 0.8905E+01	0.4057E-03 0.4564E-02	0.8158E-02 0.2222E-01	0.5040E-03 0.4340E-02
0.2454E-01	0.7380E+01 0.6142E+02	0.3727E-02 0.3747E-01	0.8756E-02 0.2454E-01	0.3304E-02 0.5818E-01
0.1227E-01	0.2332E+02 0.1778E+03	0.1833E+00 0.1065E+01	0.2354E-01 0.1287E+00	0.2238E-01 0.1940E+00
0.6136E-02	0.1017E+03 0.1101E+04	0.3603E+01 0.5819E+02	0.5477E-01 0.2036E+00	0.5801E-01 0.2137E+00
0.3068E-02	0.3350E+03 0.1313E+04	0.1867E+03 0.8391E+03	0.5381E+01 0.3656E+02	0.5381E+01 0.3655E+02

Tab. 4

ALFA	GSM	GSM1	HS	HS1
0.2000E+00	0.2657E-04	0.2124E-04	0.2650E-03	0.2662E-03
	0.2203E-03	0.1535E-03	0.1966E-02	0.1984E-02
0.4000E+00	0.7463E-05	0.9770E-06	0.2921E-04	0.2702E-04
	0.9155E-04	0.3497E-05	0.5007E-04	0.5436E-04
0.6000E+00	0.1489E-04	0.1309E-04	0.3382E-03	0.3385E-03
	0.1210E-03	0.1907E-04	0.4802E-03	0.4785E-03
0.8000E+00	0.2575E-04	0.6703E-05	0.2413E-04	0.2366E-04
	0.3004E-03	0.1059E-04	0.3691E-04	0.3748E-04
0.1000E+01	0.5181E-04	0.4464E-05	0.3425E-03	0.3429E-03
	0.8411E-03	0.7534E-05	0.5655E-03	0.5659E-03
0.2000E+02	0.1038E-02	0.8849E-04	0.4680E-02	0.4679E-02
	0.1805E-01	0.1786E-03	0.9449E-02	0.9476E-02
0.4000E+02	0.1300E-02	0.5103E-04	0.6266E-02	0.6265E-02
	0.1320E-01	0.9928E-04	0.1219E-01	0.1219E-01
0.6000E+02	0.1476E-02	0.6570E-03	0.5713E-02	0.5714E-02
	0.1215E-01	0.1155E-02	0.1004E-01	0.1005E-01
0.8000E+02	0.4177E-02	0.1712E-02	0.7328E-02	0.7329E-02
	0.5672E-01	0.2822E-02	0.1208E-01	0.1208E-01
0.1000E+03	0.3791E-02	0.2093E-02	0.1665E-01	0.1666E-01
	0.6317E-01	0.7861E-02	0.6254E-01	0.6254E-01

Tab. 6

ALFA	GSM	GSM1	HS	HS1
0.2000E+00	0.1115E-02 0.4501E-02	0.2111E-04 0.1526E-03	0.5601E-03 0.1404E-02	0.2662E-03 0.1987E-02
0.4000E+00	0.1253E-02 0.1547E-01	0.9022E-06 0.2623E-05	0.5324E-03 0.1175E-02	0.2704E-04 0.5531E-04
0.6000E+00	0.2945E-02 0.4709E-01	0.1331E-04 0.1971E-04	0.4471E-03 0.7080E-03	0.3385E-03 0.4784E-03
0.8000E+00	0.3970E-02 0.4720E-01	0.6693E-05 0.1068E-04	0.1476E-03 0.6866E-03	0.2375E-04 0.3748E-04
0.1000E+01	0.4181E-02 0.6848E-01	0.4491E-05 0.7439E-05	0.3970E-03 0.6836E-03	0.3430E-03 0.5661E-03
0.2000E+02	0.1470E+00 0.2540E+01	0.8832E-04 0.1607E-03	0.4548E-02 0.9689E-02	0.4679E-02 0.9477E-02
0.4000E+02	0.1486E+00 0.1514E+01	0.5082E-04 0.9899E-04	0.6125E-02 0.1191E-01	0.6265E-02 0.1219E-01
0.6000E+02	0.1746E+00 0.1660E+01	0.6570E-03 0.1155E-02	0.5852E-02 0.1035E-01	0.5714E-02 0.1005E-01
0.8000E+02	0.2546E+00 0.3758E+01	0.1712E-02 0.2824E-02	0.7402E-02 0.1226E-01	0.7329E-02 0.1208E-01
0.1000E+03	0.2751E+00 0.4774E+01	0.2093E-02 0.7352E-02	0.1673E-01 0.6252E-01	0.1666E-01 0.6254E-01

Tab. 7

Bibliografia.

1 C. L. Lawson, R. J. Hanson, Solving Least Squares Problems (1974), Prentice-Hall, Englewood Cliffs, N. J.

2 J. B. Rosen, The Gradient Projection Methods for Non-Linear Programming. Part 1, Linear Constraints, SIAM J. Appl. Math., Vol. 8, n° 1, (1960)

3 H. Y. Huang, A Direct Method for the General Solution of a System of Linear Equations, JOTA, Vol. 16, n° 5, (1975)

4 J. R. Rice, Experiments on Gram-Schmidt Ortogonalization, Math. Comp., vol. 20, (1966), 325-328

5 A. Bjorck, Solving Linear Least Squares Problems by Gram-Schmidt Ortogonalization, BIT, vol. 7, (1967), 1-21

6 G. H. Golub, Numerical Methods for Solving Linear Least Squares Problems, Numer. Math., vol. 7 (1965)

7 J. R. Rice, Matrix Computations and Mathematical Software, (1981), McGraw Hill, New York, New York

8 J. J. Dongarra, J. R. Brunch, G. B. Moler, G. N. Stewart, LINPACK user's guide, (1979), SIAM

9 A. K. Cline, C. B. Moler, G. N. Stewart, J. H. Wilkinson, An estimate for the Condition Number of a Matrix, SIAM J. Numer Analysis, vol. 16, (1979), 368-375