

Appunti di Calcolo Numerico

Capitolo 2

Sistemi di equazioni lineari

Maurizio Ciampa

Dipartimento di Matematica - Università di Pisa

Gli esercizi contrassegnati dal simbolo ★ sono leggermente più astratti rispetto agli altri. Quelli contrassegnati dal simbolo ♠ richiedono direttamente, o comunque riguardano, l'uso del calcolatore. A chi legge si raccomanda di riprodurre al calcolatore i “dialoghi” con *Scilab* proposti e di prendere spunto da essi per crearne di nuovi (per ottenere *Scilab* visitare la pagina <https://www.scilab.org/>).

2 Sistemi di equazioni lineari

Siano $A \in \mathbb{R}^{n \times n}$ una matrice *invertibile*, b una colonna di \mathbb{R}^n e x^* l'unica colonna di \mathbb{R}^n soluzione del sistema di equazioni lineari $Ax = b$.

I metodi per determinare la soluzione di un sistema di equazioni lineari si suddividono in *diretti* e *iterativi*. Un metodo diretto determina la soluzione del sistema con un numero finito di operazioni elementari sui reali (operazioni aritmetiche e calcolo di radici quadrate). Un metodo iterativo determina con un numero finito di operazioni elementari sui reali un elemento di una successione che converge alla soluzione del sistema.

In questo capitolo affrontiamo il problema di *determinare un'approssimazione accurata di x^** utilizzando alcuni metodi diretti.

Si ricordi che l'asserto "la colonna y di \mathbb{R}^n è soluzione del sistema $Ax = b$ " significa che $Ay = b$ e che l'asserto " $A \in \mathbb{R}^{n \times n}$ è invertibile" significa che *esiste* $B \in \mathbb{R}^{n \times n}$ tale che: $AB = BA = I$, con $I \in \mathbb{R}^{n \times n}$ matrice identità di colonne e_1, \dots, e_n . Proprietà *equivalenti* all'invertibilità di A sono:

- $\det A \neq 0$;
- $Ax = 0 \Rightarrow x = 0$, ovvero $\ker A = \{0\}$;
- Le colonne (righe) di A sono elementi linearmente indipendenti, dunque una *base*, di \mathbb{R}^n ;
- Per ogni $b \in \mathbb{R}^n$ il sistema di equazioni $Ax = b$ ha una sola soluzione.

Se M è una matrice $n \times n$ e v una colonna di n numeri, indichiamo come usuale con m_{ij} e v_i ($i, j = 1, \dots, n$) gli elementi di M e quelli di v . Invece, se k è un numero intero e M_k una matrice $n \times n$, indichiamo i suoi elementi con la notazione $M_k(i, j)$.¹

2.1 Casi semplici

Sia $A \in \mathbb{R}^{n \times n}$. Elenchiamo un insieme di casi particolari in cui la verifica dell'invertibilità di A ed il calcolo di x^* sono particolarmente *semplici*.

(D) *A diagonale*, ovvero: per ogni i, j si ha $i \neq j \Rightarrow a_{ij} = 0$.

La matrice è *invertibile se e solo se* $a_{kk} \neq 0, k = 1, \dots, n$; una volta verificata l'invertibilità, per le componenti della soluzione si ha:

$$x_k^* = b_k / a_{kk} \quad , \quad k = 1, \dots, n$$

Il numero di operazioni aritmetiche richiesto dal calcolo di x^* è: n (precisamente: n divisioni).

(T) *A triangolare*, ovvero: per ogni i, j si ha $i > j \Rightarrow a_{ij} = 0$ (triangolare *superiore*) oppure per ogni i, j si ha $i < j \Rightarrow a_{ij} = 0$ (triangolare *inferiore*).

Anche in questo caso la matrice è *invertibile se e solo se* $a_{kk} \neq 0, k = 1, \dots, n$; una volta verificata l'invertibilità, se la matrice è triangolare superiore le componenti della soluzione si determinano con la procedura di *Sostituzione all'Indietro*:

- $x = \text{SI}(T, c)$
 - // T matrice $n \times n$ triangolare superiore invertibile, c colonna di n numeri reali;
 - // x verifica la relazione: $Tx = c$.
 - $x_n = c_n / t_{nn}$;
 - per** $k = n - 1, \dots, 1$ **ripeti**:
 - $s_k = c_k - (t_{k,k+1}x_{k+1} + \dots + t_{kn}x_n)$;
 - $x_k = s_k / t_{kk}$

Se la matrice è triangolare inferiore la soluzione si calcola con l'analoga procedura di *sostituzione in avanti* (Esercizio E1).

Il numero di operazioni aritmetiche richiesto dal calcolo di x^* è: n^2 (precisamente: n divisioni, $\frac{1}{2}n(n-1)$ moltiplicazioni ed altrettante somme).

¹Questa notazione, poco usuale in matematica, è invece usuale in *Scilab*.

(O) *A* matrice *ortogonale*, ovvero che verifica una delle tre proprietà equivalenti:

- Le colonne (righe) di *A* sono una *base ortonormale* di \mathbb{R}^n rispetto al prodotto scalare canonico ($a \cdot b = a_1b_1 + \dots + a_nb_n = b^T a$);
- $A^T A = I$;
- *A* è invertibile e $A^{-1} = A^T$.

La matrice è *certamente invertibile*; per la soluzione si ha: il sistema $Ax = b$ è equivalente al sistema $A^T Ax = A^T b$, a sua volta equivalente a: $x = A^T b$, dunque: $x^* = A^T b$.

Il numero di operazioni aritmetiche richiesto dal calcolo di x^* è quello richiesto dal prodotto di una matrice $n \times n$ per una colonna di n componenti: $2n^2 - n$ (precisamente: n^2 moltiplicazioni e $n(n - 1)$ somme).

(P) *A* matrice *di permutazione*, ovvero le cui colonne (righe) sono una *permutazione* di quelle della matrice identità. Si osservi che, se *A* è una matrice di permutazione allora:

- Le colonne (righe) di *A* sono una *base ortonormale* di \mathbb{R}^n rispetto al prodotto scalare canonico, dunque: *le matrici di permutazione sono particolari matrici ortogonali*;
- Se $v \in \mathbb{R}^n$ allora le componenti di Av si ottengono *permutando* quelle di v come indicato da *A*, in particolare il numero di operazioni aritmetiche richiesto dal calcolo di Av è *zero*;
- Anche A^T è di permutazione.

La matrice è *certamente invertibile*; per la soluzione si ha: il sistema $Ax = b$ è equivalente ai sistema $A^T Ax = A^T b$, a sua volta equivalente a: $x = A^T b = x^*$, dunque: $x^* = A^T b$.

Il numero di operazioni aritmetiche richiesto dal calcolo di x^* è quello richiesto dal prodotto di una matrice $n \times n$ di *permutazione* per una colonna di n componenti: *zero*.

Esercizi

E1 Descrivere la procedura di *Sostituzione in Avanti* di intestazione:

$$x = SA(T, c)$$

che determina, dati una matrice $n \times n$ triangolare *inferiore* invertibile e una colonna c di n numeri reali, la colonna x che verifica: $Tx = c$. Verificare anche che il numero di operazioni aritmetiche richiesto dal calcolo di $x = SA(T, c)$ è lo stesso di quello riportato per il calcolo della soluzione di un sistema nel caso di matrice triangolare superiore con la procedura *SI*.

E2 Sia $A \in \mathbb{R}^{n \times n}$. Verificare che: Le colonne di *A* sono una base ortonormale di \mathbb{R}^n rispetto al prodotto scalare canonico *se e solo se* $A^T A = I$.

E3 Sia:

$$v = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

Determinare la matrice di permutazione $P \in \mathbb{R}^{3 \times 3}$ tale che:

$$Pv = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$$

2.2 Caso generale

Sia $A \in \mathbb{R}^{n \times n}$ una matrice *non* diagonale, triangolare, ortogonale o di permutazione. Per verificare se A è invertibile ed eventualmente calcolare la soluzione del sistema $Ax = b$ si procede come segue:

– *Passo 1:*

Si fattorizza A in (si scrive A come) prodotto di fattori F_1, \dots, F_j *semplici*, ovvero ciascuno appartenente ad una delle categorie D, T, O, P, e si verifica l'invertibilità di A controllando (facilmente) l'invertibilità di *ciascuno* dei fattori.

– *Passo 2:*

Se qualcuno dei fattori F_1, \dots, F_j risulta non invertibile (e quindi A risulta non invertibile) si rinuncia a calcolare la soluzione del sistema, altrimenti si calcolano (facilmente):

- (1) la soluzione c_1 del sistema $F_1x = b$;
- (2) la soluzione c_2 del sistema $F_2x = c_1$;
- \vdots
- (j) la soluzione c_j del sistema $F_jx = c_{j-1}$.

Poiché:

$$Ac_j = F_1 \cdots F_{j-1}(F_j c_j) \stackrel{(j)}{=} F_1 \cdots F_{j-2}(F_{j-1} c_{j-1}) \stackrel{(j-1)}{=} \cdots \stackrel{(2)}{=} F_1 c_1 \stackrel{(1)}{=} b$$

dall'unicità della soluzione del sistema $Ax = b$ si ottiene $c_j = x^*$. Dunque, per determinare la soluzione del sistema $Ax = b$ si risolvono tanti sistemi *semplici* quanti sono i fattori di A ottenuti nel Passo 1.

Resta da descrivere come determinare una fattorizzazione di A in prodotto di fattori semplici. Ci limiteremo a discutere le due fattorizzazioni più comunemente usate nel contesto della soluzione dei sistemi di equazioni lineari: la *fattorizzazione LR con pivoting* e la *fattorizzazione QR*, definite nell'asserto seguente.

2.1 Definizione (fattorizzazioni LR, LR con pivoting e QR di una matrice quadrata)

Una *fattorizzazione LR* di $A \in \mathbb{R}^{n \times n}$ è una coppia di matrici $S, D \in \mathbb{R}^{n \times n}$ tali che:

- $A = SD$;
- Il fattore sinistro S è *triangolare inferiore* con $s_{kk} = 1, k = 1, \dots, n$;
- Il fattore destro D è *triangolare superiore*.

Una *fattorizzazione LR con pivoting* di $A \in \mathbb{R}^{n \times n}$ è una terna di matrici $S, D, P \in \mathbb{R}^{n \times n}$ tali che:

- La matrice P è *di permutazione*;
- La coppia S, D è una fattorizzazione LR di PA .

In particolare sussiste la fattorizzazione:

$$A = P^T S D$$

Una *fattorizzazione QR* di $A \in \mathbb{R}^{n \times n}$ è una coppia di matrici $U, T \in \mathbb{R}^{n \times n}$ tali che:

- $A = UT$;
- Il fattore sinistro U è *ortogonale*;
- Il fattore destro T è *triangolare superiore*.

Si osservi che le tre fattorizzazioni riducono l'invertibilità di A a quella del solo *fattore destro* (D per le fattorizzazioni LR e LR con pivoting, T per la fattorizzazione QR).

Nella prossima Sezione si mostra come una fattorizzazione LR con pivoting possa essere determinata rileggendo opportunamente la procedura di *eliminazione di Gauss*. Analogamente, mostriamo più avanti come una fattorizzazione QR possa essere determinata rileggendo opportunamente la procedura di *ortonormalizzazione di Gram-Schmidt*.

2.3 Fattorizzazione LR con pivoting: la procedura EGP

Assegnata una matrice $A \in \mathbb{R}^{n \times n}$, la procedura EGP (*Eliminazione di Gauss con Pivoting*), di intestazione:

$$(S, D, P) = \text{EGP}(A)$$

determina una fattorizzazione LR con pivoting di A .

La fattorizzazione determinata consente (a) di verificare l'invertibilità di A constatando se $d_{11} \neq 0, \dots, d_{nn} \neq 0$ ed eventualmente (b) di determinare la soluzione del sistema $Ax = b$ calcolando $c = SA(S, Pb)$ e poi $x^* = \text{SI}(D, c)$.

Prima di dare una descrizione della procedura, introduciamo la nozione di *matrice elementare di Gauss*.

2.2 Definizione (matrice elementare di Gauss)

Una matrice $n \times n$ ad elementi reali si chiama *matrice elementare di Gauss* se è ottenuta dalla matrice identità scegliendo un indice j in $1, \dots, n-1$, numeri reali $\lambda_{j+1,j}, \dots, \lambda_{nj}$ ed operando la sostituzione:

$$e_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \lambda_{j+1,j} \\ \vdots \\ \lambda_{nj} \end{bmatrix}$$

Si osservi che una matrice elementare di Gauss è dunque una particolare matrice *triangolare inferiore con uno sulla diagonale*, dunque *invertibile*.

2.3 Esempio

Siano λ_{21} e λ_{31} numeri reali. La matrice:

$$H = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{21} & 1 & 0 \\ \lambda_{31} & 0 & 1 \end{bmatrix}$$

è elementare di Gauss (ottenuta dalla matrice identità sostituendo la prima colonna con...). Sia poi:

$$A = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \in \mathbb{R}^{3 \times 3}$$

Costruendo il prodotto *per righe* si constata che:

$$HA = \begin{bmatrix} r_1 \\ \lambda_{21}r_1 + r_2 \\ \lambda_{31}r_1 + r_3 \end{bmatrix}$$

Inoltre si verifica che:

$$H^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -\lambda_{21} & 1 & 0 \\ -\lambda_{31} & 0 & 1 \end{bmatrix}$$

In generale: Se H è una matrice elementare di Gauss, l'inversa H^{-1} si ottiene da H *cambiando segno agli elementi al di sotto della diagonale*. Anche la matrice H^{-1} è elementare di Gauss, in particolare è triangolare inferiore con uno sulla diagonale.

La procedura EGP opera come segue:

- pone $A_1 = A$;
- per $k = 1, \dots, n-1$ determina *opportunamente* $P_k \in \mathbb{R}^{n \times n}$ di permutazione e $H_k \in \mathbb{R}^{n \times n}$ elementare di Gauss e pone:

$$A_{k+1} = H_k P_k A_k$$

– pone $D = A_n$, $P = P_{n-1} \cdots P_1$ e $S = P(P_1^{-1}H_1^{-1} \cdots P_{n-1}^{-1}H_{n-1}^{-1})$.

Le matrici di permutazione P_k ed elementari di Gauss H_k sono determinate in modo che la matrice D risulti *triangolare superiore* e la matrice S risulti *triangolare inferiore con uno sulla diagonale*.

Si osservi che al termine della procedura si ha:

$$D = H_{n-1}P_{n-1} \cdots H_1P_1 A$$

da cui, essendo ciascuno dei fattori P_k e H_k *invertibile*:

$$A = P_1^{-1}H_1^{-1} \cdots P_{n-1}^{-1}H_{n-1}^{-1} D$$

La matrice:

$$\Sigma = P_1^{-1}H_1^{-1} \cdots P_{n-1}^{-1}H_{n-1}^{-1}$$

non è, in generale, triangolare inferiore (la coppia Σ, D è una fattorizzazione di A ma *non* di tipo LR) *ma* la matrice:

$$S = P\Sigma$$

è triangolare inferiore con uno sulla diagonale e $SD = P(\Sigma D) = PA$. Dunque la *terna* di matrici S, D, P è una fattorizzazione LR con pivoting della matrice PA .

Restano da discutere due punti: (i) come la procedura determina le matrici di permutazione P_k ed elementari di Gauss H_k e (ii) come mai Σ non è triangolare inferiore e $P\Sigma$ è triangolare inferiore con uno sulla diagonale. Illustreremo questi punti descrivendo dettagliatamente il comportamento della procedura in due esempi.

2.4 Esempio

Sia:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ -2 & 0 & 0 & -1 \\ -1 & 1 & 2 & -1 \end{bmatrix}$$

La procedura opera così:

- Pone $A_1 = A$;
- Pone $k = 1$.
- Costata che $A_1(1, 1) \neq 0$ e pone di conseguenza:

$$P_1 = I \quad \text{e} \quad T_1 = P_1 A_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ -2 & 0 & 0 & -1 \\ -1 & 1 & 2 & -1 \end{bmatrix}$$

Così facendo si ha $T_1(1, 1) \neq 0$.

- Considera la matrice elementare di Gauss:

$$H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & 1 & 0 & 0 \\ \lambda_{31} & 0 & 1 & 0 \\ \lambda_{41} & 0 & 0 & 1 \end{bmatrix}$$

e cerca valori di λ_{21} , λ_{31} e λ_{41} tali che gli elementi di posto (2, 1), (3, 1) e (4, 1) della matrice $H_1 T_1$ siano *zero*. Le tre condizioni equivalgono alle equazioni:

$$\lambda_{j1} T_1(1, 1) + T_1(j, 1) = 0 \quad \text{per } j = 2, 3, 4$$

Poiché $T_1(1, 1) \neq 0$ le equazioni determinano, ciascuna, *un solo valore* di λ_{j1} :

$$\lambda_{21} = -\frac{T_1(2, 1)}{T_1(1, 1)} = -2 \quad , \quad \lambda_{31} = -\frac{T_1(3, 1)}{T_1(1, 1)} = 2 \quad , \quad \lambda_{41} = -\frac{T_1(4, 1)}{T_1(1, 1)} = 1$$

- Con i valori trovati costruisce:

$$A_2 = H_1 T_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 2 & 2 & -1 \end{bmatrix}$$

Si osservi che la prima riga di A_2 è copia della prima riga di T_1 .

– Pone $k = 2$.

- Costata che $A_2(2, 2) = 0$ e cerca $j > 2$ tale che $A_2(j, 2) \neq 0$. Costatato che $A_2(3, 2) \neq 0$, indicata con P_{23} la matrice di permutazione che *scambia le righe 2 e 3*, pone di conseguenza:

$$P_2 = P_{23} \quad \text{e} \quad T_2 = P_2 A_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 2 & 2 & -1 \end{bmatrix}$$

Così facendo si mantengono gli zeri ottenuti al passo precedente (in magenta) e si ha $T_2(2, 2) \neq 0$.

- Considera la matrice elementare di Gauss:

$$H_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \lambda_{32} & 1 & 0 \\ 0 & \lambda_{42} & 0 & 1 \end{bmatrix}$$

e cerca valori di λ_{32} e λ_{42} tali che gli elementi di posto $(3, 2)$ e $(4, 2)$ della matrice $H_2 T_2$ siano *zero*. Le due condizioni equivalgono alle equazioni:

$$\lambda_{j2} T_2(2, 2) + T_2(j, 2) = 0 \quad \text{per } j = 3, 4$$

Poiché $T_2(2, 2) \neq 0$ le equazioni determinano, ciascuna, *un solo valore* di λ_{j2} :

$$\lambda_{32} = -\frac{T_2(3, 2)}{T_2(2, 2)} = 0 \quad , \quad \lambda_{42} = -\frac{T_2(4, 2)}{T_2(2, 2)} = -1$$

- Con i valori trovati costruisce:

$$A_3 = H_2 T_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

Si noti che la scelta di H_2 *mantiene* i tre zeri ottenuti al passo precedente (in blu) e le prime *due* righe di A_3 sono copia delle prime due righe di T_2 .

– Pone $k = 3$.

- Costata che $A_3(3, 3) \neq 0$ e pone di conseguenza:

$$P_3 = I \quad \text{e} \quad T_3 = P_3 A_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

Così facendo si mantengono gli zeri ottenuti al passo precedente (in magenta) e si ha $T_3(3, 3) \neq 0$.

- Considera la matrice elementare di Gauss:

$$H_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_{43} & 1 \end{bmatrix}$$

e cerca un valore di λ_{43} tale che l'elemento di posto $(4, 3)$ della matrice $H_3 T_3$ sia *zero*. La condizione equivale all'equazione:

$$\lambda_{43} T_3(3, 3) + T_3(4, 3) = -2$$

Poiché $T_3(3, 3) \neq 0$ l'equazione determina *un solo valore* di λ_{43} :

$$\lambda_{43} = -\frac{T_3(4, 3)}{T_3(3, 3)} = 0$$

- Con il valore trovato costruisce:

$$A_4 = H_3 T_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = D$$

Si noti che la scelta di H_3 *mantiene* i gli zeri ottenuti ai passi precedenti (in blu) e le prime *tre* righe di A_4 sono copia delle prime tre righe di T_3 .

I valori $T_1(1, 1)$, $T_2(2, 2)$ e $T_3(3, 3)$ che la procedura utilizza come *divisori* per determinare i vari elementi λ_{ij} , e che *ritroviamo sulla diagonale della matrice finale D*, si chiamano *pivot*. La tecnica utilizzata per determinare le matrici P_k (e quindi i pivot) si chiama *pivoting*.

Come preannunciato, la matrice $\Sigma = H_1^{-1} P_2^{-1} H_2^{-1} H_3^{-1}$ *non è* triangolare inferiore:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 \\ -1 & 1 & 2 & 1 \end{bmatrix}$$

ma, posto $P = P_3 P_2 P_1 = P_2$ si ha invece:

$$S = P\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -1 & 1 & 2 & 1 \end{bmatrix}$$

che è triangolare inferiore con uno sulla diagonale. Per capire come ciò accada si osservi che:

$$P\Sigma = P_2 H_1^{-1} P_2^{-1} H_2^{-1} H_3^{-1}$$

e che:

$$P_2 H_1^{-1} P_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \equiv H_1^{-1}(2)$$

è triangolare inferiore con uno sulla diagonale. La matrice $H_1^{-1}(2)$ è *il risultato dell'azione della permutazione P_2 sulle righe e colonne di H_1^{-1}* .

Più in generale, se:

$$P = P_3 P_2 P_1 \quad \text{e} \quad \Sigma = P_1^{-1} H_1^{-1} P_2^{-1} H_2^{-1} P_3^{-1} H_3^{-1}$$

allora:

$$P\Sigma = P_3 P_2 P_1 P_1^{-1} H_1^{-1} P_2^{-1} H_2^{-1} P_3^{-1} H_3^{-1} = P_3 (P_2 H_1^{-1} P_2^{-1}) H_2^{-1} P_3^{-1} H_3^{-1}$$

e, con la notazione introdotta sopra:

$$P\Sigma = P_3 H_1^{-1}(2) H_2^{-1} P_3^{-1} H_3^{-1}$$

Adesso, ricordando che $P_3^{-1} P_3 = I$, si riscrive:

$$P\Sigma = (P_3 H_1^{-1}(2) P_3^{-1}) (P_3 H_2^{-1} P_3^{-1}) H_3^{-1} = H_1^{-1}(2, 3) H_2^{-1}(3) H_3^{-1}$$

Le matrici $H_1^{-1}(2, 3)$, $H_2^{-1}(3)$ e H_3^{-1} sono triangolari inferiori con uno sulla diagonale e tale è il loro prodotto.

2.5 Esempio

Sia:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ -2 & -2 & 0 & -1 \\ -1 & -1 & 2 & -1 \end{bmatrix}$$

La procedura opera così:

- Pone $A_1 = A$;
- Pone $k = 1$.
- Costata che $A_1(1, 1) \neq 0$ e pone di conseguenza:

$$P_1 = I \quad \text{e} \quad T_1 = P_1 A_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ -2 & -2 & 0 & -1 \\ -1 & -1 & 2 & -1 \end{bmatrix}$$

Così facendo si ha $T_1(1, 1) \neq 0$.

- Considera la matrice elementare di Gauss:

$$H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & 1 & 0 & 0 \\ \lambda_{31} & 0 & 1 & 0 \\ \lambda_{41} & 0 & 0 & 1 \end{bmatrix}$$

e cerca valori di λ_{21} , λ_{31} e λ_{41} tali che gli elementi di posto (2, 1), (3, 1) e (4, 1) della matrice $H_1 T_1$ siano *zero*. Le tre condizioni equivalgono alle equazioni:

$$\lambda_{j1} T_1(1, 1) + T_1(j, 1) = 0 \quad \text{per } j = 2, 3, 4$$

Poiché $T_1(1, 1) \neq 0$ le equazioni determinano, ciascuna, *un solo valore* di λ_{j1} :

$$\lambda_{21} = -\frac{T_1(2, 1)}{T_1(1, 1)} = -2 \quad , \quad \lambda_{31} = -\frac{T_1(3, 1)}{T_1(1, 1)} = 2 \quad , \quad \lambda_{41} = -\frac{T_1(4, 1)}{T_1(1, 1)} = 1$$

- Con i valori trovati costruisce:

$$A_2 = H_1 T_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 2 & -1 \end{bmatrix}$$

Si osservi che la prima riga di A_2 è copia della prima riga di T_1 .

- Pone $k = 2$.
- Costata che $A_2(2, 2) = A_2(3, 2) = A_2(4, 2) = 0$ (non esiste $j > 2$ tale che $A_2(j, 2) \neq 0$). Pone di conseguenza: $P_2 = I$ e $H_2 = I$ da cui:

$$A_3 = P_2 H_2 A_2 = A_2 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 2 & -1 \end{bmatrix}$$

Si noti che la scelta di H_2 *mantiene* i tre zeri ottenuti al passo precedente (in blu) e le prime *due* righe di A_3 sono copia delle prime due righe di T_2 .

– Pone $k = 3$.

- Costata che $A_3(3, 3) = 0$ e cerca $j > 3$ tale che $A_3(j, 3) \neq 0$. Costata che $A_3(4, 3) \neq 0$ e pone di conseguenza:

$$P_3 = P_{34} \quad e \quad T_3 = P_3 A_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Così facendo si mantengono gli zeri ottenuti al passo precedente (in magenta) e si ha $T_3(3, 3) \neq 0$.

- Considera la matrice elementare di Gauss:

$$H_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & \lambda_{43} & 1 \end{bmatrix}$$

e cerca un valore di λ_{43} tale che l'elemento di posto $(4, 3)$ della matrice $H_3 T_3$ sia zero. La condizione equivale all'equazione:

$$\lambda_{43} T_3(3, 3) + T_3(4, 3) = 0$$

Poiché $T_3(3, 3) \neq 0$ l'equazione determina *un solo valore* di λ_{43} :

$$\lambda_{43} = -\frac{T_3(4, 3)}{T_3(3, 3)} = 0$$

- Con il valore trovato costruisce:

$$A_4 = H_3 T_3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 \end{bmatrix} = D$$

Si noti che la scelta di H_3 *mantiene* i gli zeri ottenuti ai passi precedenti (in blu) e le prime *tre* righe di A_4 sono copia delle prime tre righe di T_3 .

I pivot, in questo caso, sono i valori $T_1(1, 1)$ e $T_3(3, 3)$ che *ritroviamo sulla diagonale della matrice finale D*.

Anche in questo caso la matrice $\Sigma = H_1^{-1} P_3^{-1}$ *non* è triangolare inferiore:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -2 & 0 & 0 & 1 \\ -1 & 0 & 1 & 0 \end{bmatrix}$$

ma, posto $P = P_3 P_2 P_1 = P_3$ si ha invece:

$$P\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$

che è triangolare inferiore con uno sulla diagonale. Procedendo come nell'esempio precedente si osserva che:

$$S = P\Sigma = P_3 H_1^{-1} P_3^{-1} \equiv H_1^{-1}(3)$$

è triangolare inferiore con uno sulla diagonale ed è *il risultato dell'azione della permutazione P_3 sulle righe e colonne di H_1^{-1}* .

2.6 Esempio (uso della procedura EGP)

Siano: $\text{EGP}(A) = (S, D, P)$ con:

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & -1 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

e:

$$b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

In questo esempio si mostra come utilizzare la fattorizzazione determinata da EGP per calcolare $\det A$, risolvere il sistema $Ax = b$ e calcolare A^{-1} .

Si ha:

$$\det A = \det(P^{-1}SD) = \det P^T \det S \det D = (-1) \cdot 1 \cdot (-2) = 2$$

La matrice A è quindi invertibile e la soluzione del sistema si determina come segue: (i) Si calcola la soluzione del sistema $Sx = Pb$ con la procedura SA:

$$c = \text{SA}(S, Pb) = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

e poi (ii) Si calcola la soluzione x^* del sistema $Ax = b$ risolvendo con la procedura SI il sistema $Dx = c$:

$$x^* = \text{SI}(D, c) = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Infine, siano e_1, \dots, e_n le colonne della matrice identità. Per definizione, la k -esima colonna di $A^{-1} = (y_1, \dots, y_n) \in \mathbb{R}^{n \times n}$ verifica la relazione:

$$Ay_k = e_k$$

ovvero è la soluzione del sistema $Ax = e_k$ e si calcola come mostrato nel punto precedente:

$$c_k = \text{SA}(S, Pe_k), \quad y_k = \text{SI}(D, c_k)$$

Risolvendo $2n$ sistemi con matrice triangolare si ottiene:

$$A^{-1} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -\frac{1}{2} & \frac{1}{2} \\ 1 & 1 & -1 \end{bmatrix}$$

Esercizi

E4 Siano:

$$H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & 1 & 0 & 0 \\ \lambda_{31} & 0 & 1 & 0 \\ \lambda_{41} & 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad H_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \lambda_{32} & 1 & 0 \\ 0 & \lambda_{42} & 0 & 1 \end{bmatrix}$$

Calcolare $H_1 H_2$ per colonne e verificare che il numero di operazioni aritmetiche richiesto per costruire $H_1 H_2$ è zero.

E5 Siano $A, B \in \mathbb{R}^{4 \times 4}$ matrici triangolari inferiori con uno sulla diagonale. Costruire il prodotto AB per righe e verificare che a sua volta è una matrice triangolare inferiore con uno sulla diagonale.

E6 Siano $A, B \in \mathbb{R}^{4 \times 4}$ matrici triangolari superiori. Costruire il prodotto AB per colonne e verificare che a sua volta è una matrice triangolare superiore.

E7 Siano:

$$H_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \lambda_{21} & 1 & 0 & 0 \\ \lambda_{31} & 0 & 1 & 0 \\ \lambda_{41} & 0 & 0 & 1 \end{bmatrix}$$

e $P_2 = P_{24}, P_3 = P_{34}$. Calcolare $H_1^{-1}(2, 3)$ e constatare che è triangolare inferiore con uno sulla diagonale.

E8 Siano:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 0 \\ -2 & 0 & 0 & -1 \\ -1 & 1 & 2 & 0 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Applicare la procedura EGP ed utilizzare la fattorizzazione ottenuta per lo studio del sistema di equazioni lineari $Ax = b$.

E9 Siano S, D, P come nell'Esempio 2.6 ed $A = P^TSD$. Al primo passo della procedura EGP applicata ad A è necessario scegliere se scambiare la prima riga con la seconda o con la terza. Constatare che le due scelte portano a valori *diversi* delle matrici prodotte da EGP.

2.4 Norme di vettori e matrici

La procedura EGP consente di ricercare la soluzione x^* del sistema $Ax = b$ con il seguente procedimento descritto in un linguaggio che consente l'uso del tipo *numero reale*:

```
(S, D, P) = EGP(A);
se esiste k tale che dkk = 0 allora arresta il procedimento e dichiara A non invertibile;
altrimenti
  c = SA(S, Pb);
  x* = SI(D, c)
```

La discussione dell'uso del calcolatore per eseguire il procedimento richiede di sostituire al tipo *numero reale* il tipo *numero in virgola mobile e precisione finita*. Questa sostituzione consiste di *due* passaggi nel primo dei quali si sostituiscono le costanti a valore in \mathbb{R} con gli arrotondati in M . Tra le costanti presenti nel procedimento vi sono *i dati* che individuano il sistema da studiare: la matrice A e la colonna b . La sostituzione *cambia* la matrice A di elementi a_{ij} nella matrice A' di elementi $\text{rd}(a_{ij})$ e la colonna b di elementi b_i nella colonna b' di elementi $\text{rd}(b_i)$. Dunque la sostituzione di tipo *cambia il sistema in esame*: il calcolatore decide dell'invertibilità di A' ed eventualmente determina un'approssimazione ξ della soluzione \hat{x} del sistema $A'x = b'$. Supponendo che ξ sia un'approssimazione accurata di \hat{x} , occorre chiedersi se essa risulti *anche* un'approssimazione accurata di x^* . Quest'ultima condizione è l'oggetto dello studio del *condizionamento del calcolo di x^** che consiste appunto nel determinare quanto *lontano* può essere \hat{x} da x^* rispetto alla *distanza* di A' da A e di b' da b .

Allo studio del condizionamento premettiamo alcune nozioni riguardanti la *norma* di vettori e matrici.

2.7 Definizione (norma, spazio normato)

Sia V uno spazio vettoriale su \mathbb{R} . Una funzione $N : V \rightarrow \mathbb{R}$ si dice *norma* in V se ha le tre proprietà seguenti:

- (1) Per ogni $v \in V$: $N(v) \geq 0$ e $N(v) = 0 \Rightarrow v = 0$
- (2) Per ogni $v \in V$ e $\alpha \in \mathbb{R}$: $N(\alpha v) = |\alpha| N(v)$
- (3) Per ogni $v, w \in V$: $N(v + w) \leq N(v) + N(w)$ (disuguaglianza triangolare)

Il numero reale $N(v)$ si chiama *norma di v* e la coppia (V, N) si chiama *spazio normato*.

2.8 Esempio

(1) Sia V lo spazio vettoriale su \mathbb{R} dei vettori del piano. La funzione che a v associa *la lunghezza del segmento orientato che rappresenta v* verifica le proprietà richieste dalla definizione di norma.

(2) Sia $V = \mathbb{R}^n$. Le funzioni $N_1, N_2, N_\infty : V \rightarrow \mathbb{R}$ definite da:

$$N_1(v) = |v_1| + \dots + |v_n| \equiv \|v\|_1$$

$$N_2(v) = \sqrt{v_1^2 + \dots + v_n^2} \equiv \|v\|_2$$

$$N_\infty(v) = \max\{|v_1|, \dots, |v_n|\} \equiv \|v\|_\infty$$

verificano la definizione di norma (N_2 è la usuale *norma euclidea*).

2.9 Definizione (distanza tra vettori)

Con la nozione di norma è possibile introdurre quella di *distanza*: se (V, N) è uno spazio normato, per ogni $a, b \in V$ il numero reale $N(v - w)$ si chiama *distanza* tra a e b .

2.10 Esercizio

Si considerino i seguenti elementi di \mathbb{R}^2 :

$$a = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad c = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Decidere quale tra b e c è più lontano da a utilizzando, per misurare la distanza tra elementi di \mathbb{R}^2 , prima la norma N_1 poi la norma N_∞ .

2.11 Definizione (intorno sferico)

Siano $v \in \mathbb{R}^n$ e r un numero reale non negativo. Si chiama *intorno sferico* di *centro v* e *raggio r* l'insieme:

$$I(v, r) = \{x \in \mathbb{R}^n : N(x - v) \leq r\}$$

ovvero l'insieme degli elementi di \mathbb{R}^n che distano da v non più di r .

2.12 Esempio

Nella Figura 1 sono rappresentati gli intorni sferici di centro $v = 0 \in \mathbb{R}^2$ e raggio $r = 1$ nei casi $N = N_2, N_1$ e N_∞ .

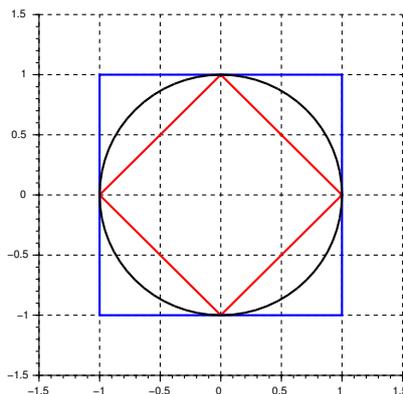


Figura 1: $I(0, 1)$ con N_2 (in nero), N_1 (in rosso) e N_∞ (in blu).

2.13 Definizione (norma di matrice)

Siano N una norma in \mathbb{R}^n e $A \in \mathbb{R}^{n \times n}$. La *norma di A indotta da N* è:

$$\|A\|_N = \sup \left\{ \frac{N(Av)}{N(v)}, v \neq 0 \right\}$$

2.14 Esempio

In \mathbb{R}^n con norma N si ha: $\|I\|_N = 1, \|0_{n \times n}\| = 0$.

2.15 Osservazione (sulla definizione di norma di matrice)

(1) Per ogni $v \neq 0$ si ha:

$$\frac{N(Av)}{N(v)} = N\left(\frac{Av}{N(v)}\right) = N\left(A \frac{v}{N(v)}\right)$$

e quindi:

$$\left\{ \frac{N(Av)}{N(v)}, v \neq 0 \right\} = \{N(Av), N(v) = 1\}$$

L'insieme B dei vettori v tali che $N(v) = 1$ è *chiuso e limitato* e la funzione $F : B \rightarrow \mathbb{R}$ definita da $F(v) = N(Av)$ è *continua*. Allora, per il Teorema di Weierstrass, F ha *massimo* e *minimo*, ovvero: esistono v^* e v_* tali che:

$$N(Av^*) = \max\{N(Av), N(v) = 1\} \quad \text{e} \quad N(Av_*) = \min\{N(Av), N(v) = 1\}$$

Allora:

$$\|A\|_N = \max\{N(Av), N(v) = 1\}$$

In particolare, posto:

$$C = \{Av, N(v) = 1\}$$

si ha: *la norma di A è il minimo valore di r tale che $C \subset I(0, r)$.*

(2) Se A è invertibile si ha:

$$\|A^{-1}\|_N = (\min\{N(Av), N(v) = 1\})^{-1}$$

Infatti:

$$\|A^{-1}\| = \sup \left\{ \frac{N(A^{-1}v)}{N(v)}, v \neq 0 \right\}$$

ovvero, posto $w = A^{-1}v$ e osservato che essendo A^{-1} invertibile si ha $v \neq 0 \Leftrightarrow w \neq 0$:

$$\|A^{-1}\| = \sup \left\{ \frac{N(w)}{N(Aw)}, w \neq 0 \right\}$$

Adesso si osservi che se $\Omega \subset \mathbb{R}$ si ha:

$$\sup \Omega = (\inf\{1/x, x \in \Omega\})^{-1}$$

dunque:

$$\|A^{-1}\| = \left(\inf \left\{ \frac{N(Aw)}{N(w)}, w \neq 0 \right\} \right)^{-1}$$

2.16 Osservazione (formule di calcolo)

Il *calcolo* di $\|A\|$ per N generica è proibitivo. Nei *casi particolari* $N = N_1, N_2$ e N_∞ si ha, dette a_1, \dots, a_n le colonne di A :

$$\|A\|_1 = \max\{N_1(a_1), \dots, N_1(a_n)\}$$

$$\|A\|_2 = \sqrt{\max\{\text{autovalori di } A^T A\}}$$

$$\|A\|_\infty = \|A^T\|_1$$

2.17 Osservazione (Proprietà della norma indotta)

Siano N una norma in \mathbb{R}^n ed $A \in \mathbb{R}^{n \times n}$. Allora:

(i) Per ogni elemento v di \mathbb{R}^n si ha: $N(Av) \leq \|A\|_N N(v)$;

- (ii) Esiste un elemento non nullo w di \mathbb{R}^n tale che: $N(Aw) = \|A\|_N N(w)$. In particolare esiste $w \in \mathbb{R}^n$ con $N(w) = 1$ tale che $N(Aw) = \|A\|_N$.

(Il primo asserto segue dalla definizione di norma di matrice, il secondo da quanto osservato nel punto (1) dell'Osservazione 2.15.)

Sia poi $B \in \mathbb{R}^{n \times n}$. Allora $AB \in \mathbb{R}^{n \times n}$ e:

- (iii) $\|AB\|_N \leq \|A\|_N \|B\|_N$.

(Infatti: Sia $v^* \in \mathbb{R}^n$ con $N(v^*) = 1$ tale che $N(ABv^*) = \|AB\|_N$. Utilizzando due volte la proprietà precedente si ottiene: $\|AB\|_N = N(ABv^*) \leq \|A\|_N N(Bv^*) \leq \|A\|_N \|B\|_N$.)

2.18 Osservazione

L'insieme $\mathbb{R}^{n \times n}$ con le usuali definizioni di somma e multiplo è uno spazio vettoriale su \mathbb{R} . Si ha:

- (1) Se N è una norma in \mathbb{R}^n allora la funzione $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ definita da $f(A) = \|A\|_N$ è una norma in $\mathbb{R}^{n \times n}$ e per ogni $A, B \in \mathbb{R}^{n \times n}$ il numero reale $\|A - B\|_N$ si chiama *distanza* tra A e B .
- (2) Lo spazio vettoriale $\mathbb{R}^{n \times n}$ è *isomorfo* allo spazio vettoriale \mathbb{R}^{n^2} . La corrispondenza che realizza l'isomorfismo è quella che alla matrice A di colonne a_1, \dots, a_n associa, prendendo a prestito la notazione da *Scilab*, il vettore $a = [a_1; \dots; a_n]$. Ciascuna delle funzioni $f_1, f_2, f_\infty : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ definite da:

- $f_1(A) = N_1(a) = \sum_{i,j=1}^n |a_{ij}|$
- $f_2(A) = N_2(a) = \sqrt{\sum_{i,j=1}^n |a_{ij}|^2}$ (detta anche *norma di Frobenius* di A)
- $f_\infty(A) = N_\infty(a) = \max\{|a_{ij}| \text{ con: } i, j = 1, \dots, n\}$

è una *norma* in $\mathbb{R}^{n \times n}$.

- (3) Siano $A \in \mathbb{R}^{n \times n}$ e $v \in \mathbb{R}^n$. Si ha:

$$\|Av\|_2 \leq f_2(A) \|v\|_2$$

(Infatti, dette r_1, \dots, r_n le righe di A ed omettendo il pedice alla norma due:

$$\|Av\| = \sqrt{|r_1v|^2 + \dots + |r_nv|^2}$$

Utilizzando la *disuguaglianza di Schwarz* si ha:

$$\sqrt{|r_1v|^2 + \dots + |r_nv|^2} \leq \sqrt{\|r_1\|^2 \|v\|^2 + \dots + \|r_n\|^2 \|v\|^2}$$

ed infine:

$$\sqrt{\|r_1\|^2 \|v\|^2 + \dots + \|r_n\|^2 \|v\|^2} = \sqrt{(\|r_1\|^2 + \dots + \|r_n\|^2) \|v\|^2}$$

da cui l'asserto.)

Esercizi

E10 In Figura 2, ottenuta utilizzando *Google Maps*, sono riportate due porzioni della cartina stradale di Manhattan. Quale funzione tra N_1, N_2 e N_∞ è utilizzata per misurare le distanze nei due casi?

E11 Verificare che le funzioni N_1 ed N_∞ sono norme in \mathbb{R}^n secondo la Definizione 2.7.

E12 Siano N una norma in \mathbb{R}^n e $\alpha \in \mathbb{R}$. Utilizzare la Definizione 2.13 per calcolare $\|\alpha I\|_N$.

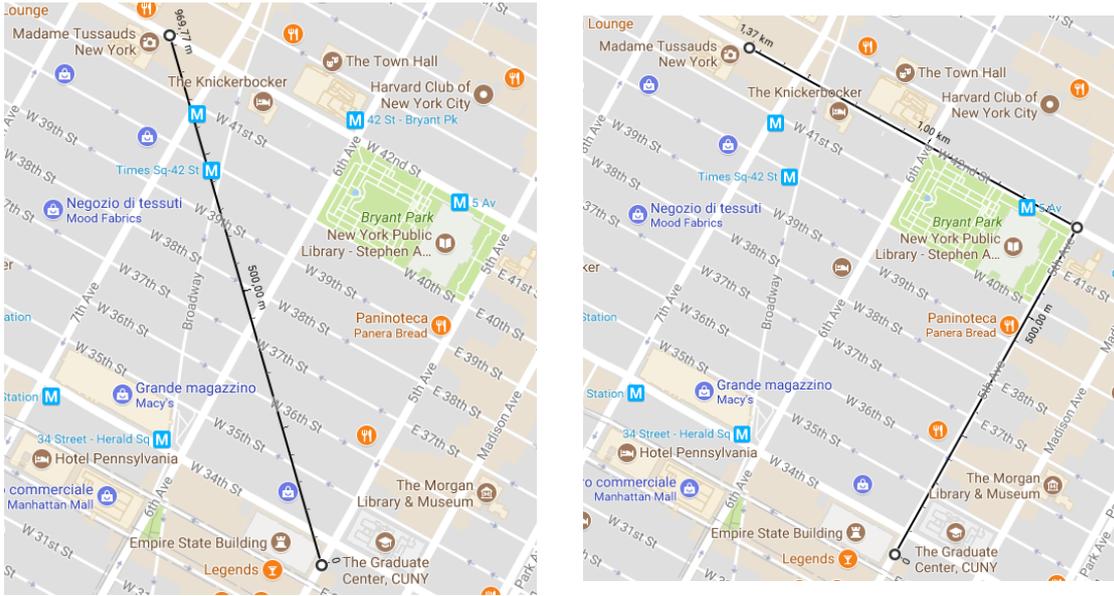


Figura 2: Distanze tra l'Empire State Building ed il museo delle cere Madame Tussauds.

E13 ★ Sia N una norma in \mathbb{R}^n . Dimostrare che: (i) se $A \in \mathbb{R}^{n \times n}$ e $\|A\|_N = 0$ allora $A = 0_{n \times n}$.
 (ii) per ogni $A, B \in \mathbb{R}^{n \times n}$ si ha: $\|A + B\|_N \leq \|A\|_N + \|B\|_N$.
 (Suggerimento per il punto (ii): si consideri $w \in \mathbb{R}^n$ tale che $N(w) = 1$ e $\|A + B\|_N = N((A + B)w)$.)

E14 Si considerino le funzioni f_1 e f_2 definite nell'Osservazione 2.18. Calcolare $f_1(I)$ e $f_2(I)$ e dedurre dal risultato che f_1 e f_2 sono norme *non indotte*.

E15 Si consideri la funzione f_∞ definita nell'Osservazione 2.18 e siano:

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Calcolare $f_\infty(A)$, $f_\infty(B)$ e $f_\infty(AB)$. Dedurre dal risultato che f_∞ è una norma *non indotta*.

E16 ★ Sia:

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

Determinare $\|A\|_\infty$ e $v^* \in \mathbb{R}^3$ tale che $N_\infty(v^*) = 1$ e $N_\infty(Av^*) = \|A\|_\infty$.

2.5 Condizionamento del calcolo della soluzione di un sistema

Siano $A \in \mathbb{R}^{n \times n}$ una matrice *invertibile*, $b \in \mathbb{R}^n$ una colonna *non nulla* e si consideri il sistema di equazioni lineari $Ax = b$. Il sistema ha una sola soluzione: l'unico vettore $x^* \in \mathbb{R}^n$ (non nullo) tale che $Ax^* = b$. Siano poi $A' \in \mathbb{R}^{n \times n}$ una matrice *invertibile*, $b' \in \mathbb{R}^n$ e si consideri il sistema di equazioni lineari $A'x = b'$. Anche questo sistema ha una sola soluzione: l'unico vettore $\hat{x} \in \mathbb{R}^n$ tale che $A'\hat{x} = b'$.

2.19 Definizione (perturbazioni, scostamento e loro misure relative)

La matrice $\delta A = A' - A$ si chiama *perturbazione* del dato A , la colonna $\delta b = b' - b$ si chiama *perturbazione* del dato b e la colonna $\delta x = \hat{x} - x^*$ si chiama *scostamento* della soluzione x^* .

La matrice A' e la colonna b' si chiamano *dati perturbati* ed il sistema $A'x = b'$ si chiama *sistema perturbato*.

Scelta una norma in \mathbb{R}^n , le misure *relative* di δA , δb e δx sono, rispettivamente:

$$\epsilon_A = \frac{\|\delta A\|}{\|A\|}, \quad \epsilon_b = \frac{\|\delta b\|}{\|b\|}, \quad \epsilon_x = \frac{\|\delta x\|}{\|x^*\|}$$

Si osservi che tutte le quantità sono ben definite perché in ciascuna il denominatore è certamente diverso da zero.

Lo studio del condizionamento del calcolo della soluzione del sistema $Ax = b$ consiste nel considerare delle *piccole perturbazioni* dei dati A, b e discutere quanto grande *può essere* ϵ_x rispetto a quanto grandi *sono* ϵ_A e ϵ_b .

2.20 Osservazione (sull'ipotesi di invertibilità di A')

Perché il sistema perturbato abbia una sola soluzione si è posta l'ipotesi che la matrice perturbata A' sia invertibile. A tal proposito si ha: Se ϵ_A è *sufficientemente piccolo* allora la matrice perturbata $A + \delta A$ è certamente invertibile.

L'ipotesi fatta è quindi *ragionevole* purché si considerino perturbazioni *piccole*, come usuale nel contesto dello studio del condizionamento.

Si analizzano prima due casi particolari poi il caso generale.

- $\delta A = 0$, $\delta b \neq 0$

Per lo scostamento δx si ha:

$$\delta x = \hat{x} - x^* = A^{-1}(b + \delta b) - A^{-1}b = A^{-1}\delta b$$

da cui, per il punto (i) dell'Osservazione 2.17:

$$\|\delta x\| = \|A^{-1}\delta b\| \leq \|A^{-1}\| \|\delta b\|$$

In termini di misura relativa si ottiene:

$$\epsilon_x = \frac{\|\delta x\|}{\|x^*\|} \leq \frac{\|A^{-1}\| \|\delta b\|}{\|x^*\|}$$

Ricordando che x^* è la soluzione del sistema $Ax = b$ si ha:

$$\|b\| = \|Ax^*\| \leq \|A\| \|x^*\| \quad \text{ovvero} \quad \frac{1}{\|x^*\|} \leq \frac{\|A\|}{\|b\|}$$

Infine, sostituendo:

$$\epsilon_x \leq \|A^{-1}\| \|A\| \frac{\|\delta b\|}{\|b\|} = \|A^{-1}\| \|A\| \epsilon_b$$

Si osservi che per il punto (ii) dell'Osservazione 2.17 esistono una colonna $w \neq 0$ tale che:

$$\|A^{-1}w\| = \|A^{-1}\| \|w\|$$

ed una colonna $y \neq 0$ tale che:

$$\|Ay\| = \|A\| \|y\|$$

Allora, per $\delta b = w$ e $x^* = y$ (ovvero $b = Ay \neq 0$) si ha:

$$\epsilon_x = \frac{\|A^{-1}\delta b\|}{\|x^*\|} = \frac{\|A^{-1}w\|}{\|y\|} = \frac{\|A^{-1}\| \|w\|}{\|y\|} = \|A^{-1}\| \|A\| \frac{\|w\|}{\|b\|} = \|A^{-1}\| \|A\| \epsilon_b$$

Introdotta il *numero di condizionamento* di A :

$$c(A) = \|A\| \|A^{-1}\|$$

il risultato ottenuto si riscrive nella forma:

2.21 Teorema (di condizionamento, I)

Sia $A \in \mathbb{R}^{n \times n}$ invertibile. Allora:

(i) Per ogni vettore non nullo $b \in \mathbb{R}^n$ ed ogni $\delta b \in \mathbb{R}^n$ si ha:

$$\epsilon_x \leq c(A) \epsilon_b$$

(ii) Esistono un vettore δb ed un vettore non nullo b tali che:

$$\epsilon_x = c(A) \epsilon_b$$

- $\delta A \neq 0, \delta b = 0$

Si ricordi che si considerano solo perturbazioni δA tali che $A + \delta A$ invertibile. Si ha:

$$(A + \delta A)\hat{x} = b = Ax^*$$

da cui:

$$A \delta x = A(\hat{x} - x^*) = -\delta A \hat{x}$$

Per lo scostamento δx si ottiene allora:

$$\delta x = -A^{-1} \delta A \hat{x}$$

e quindi per il punto (i) dell'Osservazione 2.17:

$$\|\delta x\| = \|A^{-1} \delta A \hat{x}\| \leq \|A^{-1} \delta A\| \|\hat{x}\|$$

Per punto (iii) dell'Osservazione 2.17 si ha poi:

$$\|A^{-1} \delta A\| \leq \|A^{-1}\| \|\delta A\|$$

Introducendo come misura relativa dello scostamento la quantità (si osservi che $\|\hat{x}\| \neq 0$ perché $b \neq 0$):

$$\hat{\epsilon}_x = \frac{\|\delta x\|}{\|\hat{x}\|}$$

si ottiene:

$$\hat{\epsilon}_x \leq \|A^{-1}\| \|\delta A\| = \|A^{-1}\| \|A\| \frac{\|\delta A\|}{\|A\|} = \|A^{-1}\| \|A\| \epsilon_A$$

Si osservi che per il punto (ii) dell'Osservazione 2.17 per ogni δA esiste una colonna $w \neq 0$ tale che:

$$\|A^{-1} \delta A w\| = \|A^{-1} \delta A\| \|w\|$$

Inoltre, per qualche matrice Z tale che $A + Z$ invertibile (ad esempio per $Z = \alpha I$ con $\alpha \in \mathbb{R}$ sufficientemente piccolo) si ha:

$$\|A^{-1} Z\| = \|A^{-1}\| \|Z\|$$

Allora, per $\delta A = Z$ e $\hat{x} = w$ (e quindi per $b = (A + Z)w$) si ha:

$$\|\delta x\| = \|A^{-1} \delta A \hat{x}\| = \|A^{-1} Z w\| = \|A^{-1} Z\| \|w\| = \|A^{-1}\| \|Z\| \|w\| = \|A^{-1}\| \|\delta A\| \|\hat{x}\|$$

e quindi:

$$\|\hat{\epsilon}_x\| = \|A^{-1}\| \|\delta A\| = \|A^{-1}\| \|A\| \epsilon_A$$

Con le notazioni introdotte si riscrive il risultato ottenuto nella forma:

2.22 Teorema (di condizionamento, II)

Sia $A \in \mathbb{R}^{n \times n}$ invertibile. Allora:

(i) Per ogni vettore non nullo $b \in \mathbb{R}^n$ ed ogni $\delta A \in \mathbb{R}^{n \times n}$ tale che $A + \delta A$ invertibile si ha:

$$\hat{\epsilon}_x \leq c(A) \epsilon_A$$

(ii) Esistono una matrice δA tale che $A + \delta A$ invertibile ed un vettore non nullo b tali che:

$$\hat{\epsilon}_x = c(A) \epsilon_A$$

- $\delta A \neq 0$ e $\delta b \neq 0$

Dati $\delta b \in \mathbb{R}^n$ tale che $b + \delta b$ è non nullo e $\delta A \in \mathbb{R}^{n \times n}$ tale che $A + \delta A$ invertibile, siano: \hat{x} la soluzione del sistema perturbato $(A + \delta A)x = b + \delta b$, \hat{x}_b la soluzione del sistema perturbato $Ax = b + \delta b$ e x^* la soluzione del sistema $Ax = b$. Allora, posto:

$$\epsilon_A = \frac{\|\delta A\|}{\|A\|} \quad , \quad \epsilon_b = \frac{\|\delta b\|}{\|b\|} \quad \text{e} \quad \epsilon_x = \frac{\|\hat{x} - x^*\|}{\|x^*\|}$$

si ha:

- (1) Per quanto mostrato nei casi particolari:

$$\frac{\|\hat{x} - \hat{x}_b\|}{\|\hat{x}\|} \leq c(A) \epsilon_A \quad \text{e} \quad \frac{\|\hat{x}_b - x^*\|}{\|x^*\|} \leq c(A) \epsilon_b$$

$$(2) \quad \epsilon_x = \frac{\|\hat{x} - x^*\|}{\|x^*\|} \leq \frac{\|\hat{x} - \hat{x}_b\|}{\|x^*\|} + \frac{\|\hat{x}_b - x^*\|}{\|x^*\|} = \frac{\|\hat{x} - \hat{x}_b\|}{\|\hat{x}\|} \frac{\|\hat{x}\|}{\|x^*\|} + \frac{\|\hat{x}_b - x^*\|}{\|x^*\|}$$

$$(3) \quad \frac{\|\hat{x}\|}{\|x^*\|} \leq \frac{\|\hat{x} - x^*\|}{\|x^*\|} + 1 = \epsilon_x + 1$$

Quindi:

$$\epsilon_x \leq c(A) \epsilon_A (\epsilon_x + 1) + c(A) \epsilon_b$$

ovvero:

$$(1 - c(A) \epsilon_A) \epsilon_x \leq c(A) (\epsilon_A + \epsilon_b)$$

Si ottiene infine:

2.23 Teorema (di condizionamento, III)

Sia $A \in \mathbb{R}^{n \times n}$ invertibile. Allora: per ogni vettore non nullo $b \in \mathbb{R}^n$, ogni $\delta b \in \mathbb{R}^n$ tale che $b + \delta b$ è non nullo e ogni $\delta A \in \mathbb{R}^{n \times n}$ tale che $A + \delta A$ invertibile e $c(A) \epsilon_A < 1$ si ha:

$$\epsilon_x \leq \frac{c(A)}{1 - c(A) \epsilon_A} (\epsilon_A + \epsilon_b)$$

2.24 Osservazione

- (1) Sia N una norma in \mathbb{R}^n ed $A \in \mathbb{R}^{n \times n}$ invertibile. Allora: $c(A) \geq 1$.
(Infatti: $I = A^{-1}A$ e quindi $1 = \|I\|_N = \|A^{-1}A\|_N \leq \|A^{-1}\|_N \|A\|_N = c(A)$.)
- (2) Siano $A \in \mathbb{R}^{n \times n}$ invertibile e $\delta A \in \mathbb{R}^{n \times n}$. Se $c(A) \epsilon_A < 1$ allora $A + \delta A$ invertibile.
(Infatti: $c(A) \epsilon_A = \|A^{-1}\| \|\delta A\|$ e quindi per l'ipotesi:

$$\|A^{-1}\delta A\| \leq \|A^{-1}\| \|\delta A\| < 1$$

Inoltre: $A + \delta A = A(I + A^{-1}\delta A)$, dunque:

$$A + \delta A \text{ invertibile} \quad \Leftrightarrow \quad I + A^{-1}\delta A \text{ invertibile}$$

Infine: Se per qualche $v \neq 0$ si ha $(I + A^{-1}\delta A)v = 0$ allora si ha anche: $v = -A^{-1}\delta A v$ e quindi $N(v) = N(A^{-1}\delta A v) \leq \|A^{-1}\delta A\| N(v)$, ovvero $\|A^{-1}\delta A\| \geq 1$.)

Dunque:

$$\epsilon_A < \frac{1}{c(A)} \quad \Rightarrow \quad A + \delta A \text{ invertibile}$$

- (3) Sia $x^* \in \mathbb{R}^n$ un vettore non nullo e $\delta x \in \mathbb{R}^n$. Si consideri, per ogni k tale che $x_k^* \neq 0$, la misura relativa della *componente k-esima* dello scostamento:

$$\frac{|\delta x_k|}{|x_k^*|}$$

Poiché per ogni vettore $y \in \mathbb{R}^n$ ed ogni k si ha $|y_k| \leq \|y\|$, allora:

$$\frac{|\delta x_k|}{|x_k^*|} \leq \frac{\|\delta x\|}{\|x^*\|} = \frac{\|\delta x\|}{\|x^*\|} \frac{\|x^*\|}{|x_k^*|} = \epsilon_x \frac{\|x^*\|}{|x_k^*|}$$

con:

$$\frac{\|x^*\|}{|x_k^*|} \geq 1$$

Dunque: se la componente k -esima del vettore x^* è molto vicina a zero, la misura relativa della componente k -esima dello scostamento può essere molto maggiore della misura relativa del vettore scostamento.

- (4) Scelti $M = F(\beta, m)$ e la funzione arrotondamento rd , siano: $A \in \mathbb{R}^{n \times n}$ di elementi a_{ij} una matrice invertibile, $b \in \mathbb{R}^n$ di elementi b_i una colonna non nulla, A' la matrice di elementi $\text{rd}(a_{ij})$, b' la colonna di elementi $\text{rd}(b_i)$, $\delta A = A' - A$ e $\delta b = b' - b$. Scelta una norma in \mathbb{R}^n tra N_1, N_2 e N_∞ e detta u la precisione di macchina in M si ha:

$$\epsilon_b \leq u \quad \text{e} \quad \epsilon_A \leq u$$

Se:

$$c(A) u < 1$$

allora:

- (4.a) Per quanto mostrato nel punto (2), la matrice A' è invertibile;
 (4.b) Dette x^* e \hat{x} rispettivamente la soluzione del sistema $Ax = b$ e la soluzione del sistema $A'x = b'$ per il Teorema di condizionamento, III si ha:

$$\epsilon_x \leq \frac{2c(A) u}{1 - c(A) u}$$

2.25 Osservazione (applicazione del Teorema di condizionamento)

Dati $A \in \mathbb{R}^{n \times n}$ invertibile, b elemento non nullo di \mathbb{R}^n e $\hat{x} \in \mathbb{R}^n$, si utilizza \hat{x} per approssimare la soluzione x^* del sistema $Ax = b$. Per ottenere informazioni sull'accuratezza dell'approssimazione, si introduce il vettore:

$$r = A\hat{x} - b$$

detto *residuo* di $Ax = b$ associato ad \hat{x} .

- (1) Si consideri la seguente *interpretazione* di \hat{x} :

$$\hat{x} \text{ è la soluzione del sistema perturbato } Ax = b + r$$

Per il Teorema di condizionamento, I: posto $\delta b = r$ si ha:

$$\frac{\|\hat{x} - x^*\|}{\|x^*\|} \leq c(A) \frac{\|r\|}{\|b\|}$$

Si ottiene una limitazione dell'errore relativo commesso approssimando x^* con \hat{x} .

- (2) Siano $\hat{x} \neq 0$ e $M \in \mathbb{R}^{n \times n}$ tale che $M\hat{x} = -r$.

(La condizione $\hat{x} \neq 0$ è sufficiente a garantire l'esistenza di matrici M tali che $M\hat{x} = -r$. Ad esempio:

$$M = -\frac{r\hat{x}^T}{\hat{x}^T\hat{x}}$$

Se $\hat{x} = 0$, invece, esistono matrici con la proprietà richiesta se e solo se anche $r = 0$.)

Se $A + M$ invertibile, si consideri la seguente *interpretazione* di \hat{x} :

$$\hat{x} \text{ è la soluzione del sistema perturbato } (A + M)x = b$$

Per il Teorema di condizionamento, II: posto $\delta A = M$ si ha:

$$\frac{\|\hat{x} - x^*\|}{\|\hat{x}\|} \leq c(A) \frac{\|M\|}{\|A\|}$$

Si ottiene una limitazione dell'errore relativo commesso approssimando \hat{x} con x^* .

Per ottenere una limitazione dell'errore relativo commesso approssimando x^* con $\hat{x} = x^* + \delta x$ si osserva che:

$$\epsilon_x = \frac{\|\delta x\|}{\|x^*\|} = \frac{\|\delta x\|}{\|\hat{x}\|} \frac{\|\hat{x}\|}{\|x^*\|} = \hat{\epsilon}_x \frac{\|\hat{x}\|}{\|x^*\|}$$

e:

$$\frac{\|\hat{x}\|}{\|x^*\|} = \frac{\|\hat{x} - x^* + x^*\|}{\|x^*\|} \leq \frac{\|\hat{x} - x^*\|}{\|x^*\|} + 1 = \epsilon_x + 1$$

Allora:

$$\epsilon_x \leq \hat{\epsilon}_x (\epsilon_x + 1)$$

ovvero:

$$\epsilon_x (1 - \hat{\epsilon}_x) \leq \hat{\epsilon}_x$$

Dunque:

$$\hat{\epsilon}_x < 1 \quad \Rightarrow \quad \epsilon_x \leq \frac{\hat{\epsilon}_x}{1 - \hat{\epsilon}_x}$$

Se:

$$\alpha \equiv c(A) \frac{\|M\|}{\|A\|} < 1$$

si ottiene infine una limitazione dell'errore relativo commesso approssimando x^* con \hat{x} :

$$\frac{\|\hat{x} - x^*\|}{\|x^*\|} \leq \frac{\alpha}{1 - \alpha}$$

2.26 Esempio

Si consideri \mathbb{R}^2 con norma N_1 e siano:

$$A = \begin{bmatrix} 20 & 1 \\ 0 & 20 \end{bmatrix}, \quad b = \begin{bmatrix} 10 \\ 10 \end{bmatrix}, \quad \hat{x} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Si approssima la soluzione x^* del sistema $Ax = b$ con \hat{x} . Per l'accuratezza dell'approssimazione si ha:

(1) Dopo aver calcolato A^{-1} si ottiene:

$$c(A) = \|A^{-1}\| \|A\| = \frac{21}{400} \cdot 21 = \frac{441}{400} \approx 1$$

Inoltre il residuo di $Ax = b$ associato ad \hat{x} vale:

$$r = A\hat{x} - b = \frac{1}{2} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

(2) Si interpreta \hat{x} come soluzione del sistema $Ax = b + r$. La misura relativa della perturbazione è:

$$\epsilon_b = \frac{\|r\|}{\|b\|} = \frac{1}{40}$$

e, utilizzando il Teorema di condizionamento, I:

$$\epsilon_x \leq c(A) \epsilon_b = \frac{441}{400} \frac{1}{40} \equiv \alpha_1 \approx 2.76 \cdot 10^{-2}$$

(3) Posto:

$$\delta A = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}$$

si osserva che $A + \delta A$ è invertibile e si interpreta \hat{x} come soluzione del sistema $(A + \delta A)x = b$. La misura relativa della perturbazione è:

$$\epsilon_A = \frac{\|\delta A\|}{\|A\|} = \frac{1}{21}$$

e, utilizzando il Teorema di condizionamento, II:

$$\hat{\epsilon}_x \leq c(A) \epsilon_A = \frac{441}{400} \frac{1}{21} = \frac{21}{400} \equiv \alpha_2 = 5.25 \cdot 10^{-2}$$

Essendo $\alpha_2 < 1$ si ha allora:

$$\epsilon_x \leq \frac{\alpha_2}{1 - \alpha_2} \approx 5.54 \cdot 10^{-2}$$

La limitazione ottenuta nel secondo caso è peggiore di quella ottenuta nel primo (infatti: $5.54 \cdot 10^{-2} > 2.76 \cdot 10^{-2}$). Se ne conclude che, utilizzando la norma N_1 , l'errore relativo commesso approssimando x^* con \hat{x} non supera $\alpha_1 \approx 2.76 \cdot 10^{-2}$.

Esercizi

E17 Si consideri \mathbb{R}^2 con norma N_1 e sia:

$$b = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Disegnare su un piano cartesiano l'insieme di tutti i vettori b' ottenuti sommando a b le perturbazioni δb tali che $\epsilon_b \leq \frac{1}{10}$.

E18 Si consideri \mathbb{R}^3 con norma N_∞ e siano:

$$x^* = \begin{bmatrix} 1 \\ 10^{-2} \\ 0 \end{bmatrix}, \quad \delta x = 10^{-4} \begin{bmatrix} 1 \\ 10 \\ -1 \end{bmatrix}$$

Determinare la misura relativa del vettore scostamento ϵ_x e, per ogni k tale che $x_k^* \neq 0$, la misura relativa della componente k -esima dello scostamento.

E19 Siano $M = F(2, 53)$ e A, b, A', b' come nel punto (4) dell'Osservazione 2.24. Utilizzare la limitazione mostrata nel punto (4.b) per ottenere una condizione sufficiente su $c(A)$ in modo che sia $\epsilon_x < 10^{-6}$.

E20 ★ Siano $M = F(2, 53)$ e $b, b', \delta b$ come nel punto (4) dell'Osservazione 2.24. Utilizzare il Teorema 0.27 del Capitolo 0 per dimostrare che, indicando con u la precisione di macchina, per $N = N_1, N_2$ e N_∞ si ha: $N(\delta b) \leq u N(b)$, ovvero $\epsilon_b \leq u$.

E21 ★ Dimostrare che:

$$\epsilon_x < 1 \quad \Rightarrow \quad \hat{\epsilon}_x \leq \frac{\epsilon_x}{1 - \epsilon_x}$$

E22 Nell'esempio finale si è ottenuto:

$$\epsilon_x \leq \alpha_1 \approx 2.76 \cdot 10^{-2}, \quad \hat{\epsilon}_x \leq \alpha_2 = 5.25 \cdot 10^{-2}$$

(1) Dedurre dalla limitazione su $\hat{\epsilon}_x$ che:

$$\|\delta x\|_1 \leq \alpha_2$$

(2) Utilizzare il risultato dell'Esercizio precedente per ottenere, dalla limitazione su ϵ_x , una limitazione su $\hat{\epsilon}_x$ e dedurre una nuova limitazione su $\|\delta x\|_1$.

- (3) Rappresentare su un piano cartesiano i vettori δx che verificano le limitazioni trovate in (1) e (2) e dedurne un insieme che certamente contiene l'effettivo vettore δx .

E23 Determinare la soluzione del sistema dell'Esempio 2.26 e controllare che le limitazioni trovate sono soddisfatte.

E24 Siano \hat{x} e r come nell'Esempio 2.26. Determinare:

$$M = -\frac{r\hat{x}^T}{\hat{x}^T\hat{x}}$$

e verificare che $M\hat{x} = -r$. Posto poi:

$$\epsilon_A = \frac{\|M\|}{\|A\|}$$

verificare che:

$$\alpha_3 = c(A)\epsilon_A < 1$$

Dunque $A + M$ è invertibile e \hat{x} è l'unica soluzione del sistema perturbato $(A + M)x = b$. Utilizzare il Teorema di condizionamento, II per ottenere una limitazione dell'errore relativo commesso approssimando x^* con \hat{x} e confrontare la limitazione ottenuta con quelle già ricavate.

2.6 Uso del tipo *numero in virgola mobile*: la procedura EGPP

Si ricordi che, assegnata una matrice $A \in \mathbb{R}^{n \times n}$ ed una colonna $b \in \mathbb{R}^n$, la procedura EGP permette la ricerca della soluzione della soluzione del sistema $Ax = b$ con il seguente procedimento descritto con un linguaggio che consente l'uso del tipo *numero reale*:

$(S, D, P) = \text{EGP}(A)$;
se esiste k tale che $d_{kk} = 0$ **allora** arresta il procedimento e dichiara A non invertibile;
altrimenti
 $c = \text{SA}(S, Pb)$;
 $x^* = \text{SI}(D, c)$

Sostituendo al tipo *numero reale* il tipo *numero in virgola mobile e precisione finita* il procedimento si modifica come segue:

$(\hat{S}, \hat{D}, \hat{P}) = \text{EGP}(A')$;
se esiste k tale che $\hat{d}_{kk} = 0$ **allora** arresta il procedimento e dichiara A' non invertibile;
altrimenti
 $\hat{c} = \text{SA}(\hat{S}, \hat{P}b')$;
 $\xi = \text{SI}(\hat{D}, \hat{c})$

dove: A' e b' indicano, rispettivamente, la matrice e la colonna ottenute arrotondando in M ciascuna componente di A e b ed **EGP**, **SA** ed **SI** indicano le procedure ottenute sostituendo il tipo *numero reale* rispettivamente nelle procedure EGP, SA ed SI.

Il vettore finale ξ è utilizzato per approssimare x^* . Le osservazioni seguenti mostrano che l'approssimazione è *potenzialmente non accurata* ma una piccola modifica della procedura EGP la rende invece *quasi sempre* accurata quanto consentito dal condizionamento di A .

2.27 Osservazione (interpretazione di SI)

Siano $M = F(\beta, m)$, $T \in \mathbb{R}^{2 \times 2}$ una matrice triangolare superiore invertibile e $c \in \mathbb{R}^2$. La procedura SI determina la soluzione del sistema:

$$\begin{bmatrix} t_{11} & t_{12} \\ 0 & t_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

calcolando:

- (1) $x_2 = c_2/t_{22}$;
 (2) $s_1 = c_1 - t_{12}x_2$ e poi $x_1 = s_1/t_{11}$ ovvero: $x_1 = (c_1 - t_{12}x_2)/t_{11}$.

La procedura SI, *supponendo che gli elementi di T e c siano elementi di M* , calcola:

- (1) $\xi_2 = c_2 \oslash t_{22}$;
 (2) $\sigma_1 = c_1 \ominus (t_{12} \otimes \xi_2)$ e poi $\xi_1 = \sigma_1 \oslash t_{11}$.

Ricordando la definizione di pseudo-operazioni aritmetiche, per il Teorema 0.27 del Capitolo 0 si ha: esistono numeri reali e_1, \dots, e_4 ciascuno di valore assoluto non superiore alla precisione di macchina u tali che:

$$\xi_2 = (1 + e_1) c_2 / t_{22} \quad , \quad \sigma_1 = (1 + e_3) (c_1 - (1 + e_2) t_{12} \xi_2) \quad \text{e} \quad \xi_1 = (1 + e_4) \sigma_1 / t_{11}$$

Posto poi:

$$t'_{22} = t_{22} / (1 + e_1) \neq 0 \quad , \quad t'_{12} = (1 + e_2) t_{12} \quad \text{e} \quad t'_{11} = t_{11} / ((1 + e_3)(1 + e_4)) \neq 0$$

si ottiene:

$$\xi_2 = c_2 / t'_{22} \quad \text{e} \quad \xi_1 = (c_1 - t'_{12} \xi_2) / t'_{11}$$

ovvero $\mathbf{SI}(T, c)$ è la soluzione del sistema:

$$\begin{bmatrix} t'_{11} & t'_{12} \\ 0 & t'_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

con:

$$t'_{22} \approx t_{22} \quad , \quad t'_{12} \approx t_{12} \quad \text{e} \quad t'_{11} \approx t_{11}$$

In generale si dimostra la seguente *interpretazione*:

$\mathbf{SI}(T, c) = \mathbf{SI}(T', c)$ con T' triangolare superiore invertibile tale che $t'_{ij} \approx t_{ij}$ per ogni i, j

L'interpretazione mostra anche che l'algoritmo SI è *stabile* quando utilizzato per approssimare la funzione SI. Più precisamente, l'algoritmo è *stabile all'indietro*: per ogni dato X , SI fornisce il *valore* della funzione SI in un punto vicino a X . Si ricordi che la stabilità richiede, per ogni dato X , che l'algoritmo di fornisca *un'approssimazione accurata* del valore della funzione in un punto vicino ad X .

2.28 Osservazione (inadeguatezza della procedura EGP)

L'Osservazione precedente mostra che $\xi = \mathbf{SI}(\hat{D}, \hat{c}) = \mathbf{SI}(\hat{D}', \hat{c})$. Per giudicare l'accuratezza di ξ come approssimazione di $x^* = \mathbf{SI}(D, c)$ occorre dunque studiare il *condizionamento* del calcolo della soluzione del sistema $Dx = c$, ovvero indagare il *numero di condizionamento* della matrice D .

– *Esempio*

Sia $\alpha \in (0, \frac{1}{2})$ e:

$$A = \begin{bmatrix} \alpha & 1 \\ 1 & 0 \end{bmatrix}$$

La procedura EGP applicata ad A produce:

$$S = \begin{bmatrix} 1 & 0 \\ 1/\alpha & 1 \end{bmatrix} \quad , \quad D = \begin{bmatrix} \alpha & 1 \\ 0 & -1/\alpha \end{bmatrix} \quad , \quad P = I$$

Allora:

$$D^{-1} = \begin{bmatrix} 1/\alpha & 1 \\ 0 & -\alpha \end{bmatrix}$$

e il numero di condizionamento di D , utilizzando ad esempio la norma infinito in \mathbb{R}^2 è:

$$c(D) = \|D^{-1}\| \|D\| = \frac{\alpha + 1}{\alpha^2}$$

Si osservi che:

$$\lim_{\alpha \rightarrow 0} c(D) = +\infty$$

e che, essendo:

$$A^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -\alpha \end{bmatrix}$$

risulta:

$$c(A) = (1 + \alpha)^2 < 3$$

Dunque: per α sufficientemente piccolo, *il fattore destro D prodotto dalla procedura EGP ha numero di condizionamento arbitrariamente più alto di A* . Ovvero il procedimento basato su EGP ha trasformato il sistema $Ax = b$, con *buone* proprietà di condizionamento, nel sistema *equivalente* $Dx = c$ che ha però proprietà di condizionamento, per α piccolo, *pessime*.

L'esempio mostra quindi che il procedimento che usa la procedura EGP *può* generare un'approssimazione ξ non accurata.

2.29 Osservazione (procedura EGPP)

Per ovviare al problema evidenziato nell'esempio dell'Osservazione precedente, si modifica la ricerca delle matrici di permutazione nella procedura EGP. Precisamente, per $k = 1, \dots, n-1$: se $A_k(k, k) = \dots = A_k(n, k) = 0$ si pone $P_k = I$ *altrimenti* si determina l'indice $j \geq k$ tale che:

$$\max\{|A_k(k, k)|, \dots, |A_k(n, k)|\} = |A_k(j, k)|$$

e si pone:

$$P_k = \begin{cases} I & \text{se } j = k \\ P_{kj} & \text{se } j > k \end{cases}$$

La procedura così ottenuta si chiama EGPP (*Eliminazione di Gauss con Pivotong Parziale*). Applicata alla matrice A dell'esempio fornisce:

$$\text{EGPP}(A) = \left(\begin{bmatrix} 1 & 0 \\ \alpha & 1 \end{bmatrix}, I, P_{12} \right)$$

e quindi $c(D) = 1$.

In generale si ha: *Per ogni numero intero positivo n esiste un numero reale k_n (che dipende dalla norma scelta in \mathbb{R}^n) tale che: la procedura EGPP applicata ad $A \in \mathbb{R}^{n \times n}$ invertibile produce un fattore destro D con $c(D) \leq k_n c(A)$.*

– *Dimostrazione.*

Sia $\text{EGPP}(A) = (S, D, P)$. Allora:

$$D = S^{-1}PA \quad \text{e} \quad D^{-1} = A^{-1}P^{-1}S$$

Scelta in \mathbb{R}^n una tra le norme N_1, N_2 e N_∞ , per ogni matrice di permutazione M si ha: $\|M\| = 1$ e quindi:

$$c(D) \leq c(S) c(A)$$

La tecnica del *pivoting parziale* garantisce che per ogni $k = 1, \dots, n-1$ *il valore assoluto degli elementi non nulli della matrice H_k non supera uno*. Allora lo stesso vale per gli elementi della matrice S . Quindi:

$$\|S\|_1 \leq n \quad \text{e} \quad \|S\|_\infty \leq n$$

Per ogni matrice A si ha: $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$. Allora si ha anche:

$$\|S\|_2 \leq n$$

Inoltre si ha:

$$S = PP_1^{-1}H_1^{-1} \dots P_{n-1}^{-1}H_{n-1}^{-1} \quad \Rightarrow \quad S^{-1} = H_{n-1}P_{n-1} \dots H_1P_1P^{-1}$$

da cui:

$$\|S^{-1}\|_\infty \leq \|H_{n-1}\|_\infty \dots \|H_1\|_\infty \leq 2 \dots 2 = 2^{n-1}$$

Per ogni $A \in \mathbb{R}^{n \times n}$ si ha: $\|A\|_1 \leq \sqrt{n} \|A\|_2$ e $\|A\|_2 \leq \sqrt{n} \|A\|_\infty$. Dunque:

$$\|S^{-1}\|_1 \leq n 2^{n-1} \quad \text{e} \quad \|S^{-1}\|_2 \leq \sqrt{n} 2^{n-1}$$

Dalle disuguaglianze ottenute si ricava:

$$c_1(S) \leq n^2 2^{n-1} \quad , \quad c_2(S) \leq n^{3/2} 2^{n-1} \quad , \quad c_\infty(S) = n 2^{n-1}$$

da cui l'asserto.

Esercizi

E25 Siano $T \in \mathbb{R}^{2 \times 2}$ una matrice triangolare inferiore invertibile e $c \in \mathbb{R}^2$. Mostrare che, supponendo che gli elementi di T e c siano elementi di M , sussiste la seguente interpretazione, simile a quella data per SI:

$$\text{SA}(T, c) = \text{SA}(T', c) \quad \text{con} \quad T' \text{ triangolare inferiore tale che } t'_{ij} \approx t_{ij} \text{ per ogni } i, j$$

E26 Sia:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 2 & 1 & 4 \end{bmatrix}$$

Determinare EGPP(A).

Soluzione:

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix} \quad , \quad D = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad , \quad P = P_{23}P_{13} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

2.7 Fattorizzazione QR: la procedura GS

Assegnata una matrice $A \in \mathbb{R}^{n \times n}$, la procedura GS (*procedimento di Gram-Schmidt*) di intestazione:

$$(U, T) = \text{GS}(A)$$

cerca una *fattorizzazione QR* di A .

Se la procedura trova una fattorizzazione, la soluzione del sistema $Ax = b$ si determina calcolando $x^* = \text{SI}(T, U^T b)$.

L'Esempio seguente mostra come utilizzare il procedimento di ortonormalizzazione di Gram-Schmidt per cercare una fattorizzazione QR di una matrice A .

2.30 Esempio

Si consideri \mathbb{R}^3 con prodotto scalare canonico (per ogni $a, b \in \mathbb{R}^3$: $a \cdot b = b^T a$). Sia $A \in \mathbb{R}^{3 \times 3}$ di colonne a_1, a_2, a_3 .

– *Primo passo*

Si cercano $\Omega = (\omega_1, \omega_2, \omega_3) \in \mathbb{R}^{3 \times 3}$ a colonne ortogonali non nulle e $\Theta \in \mathbb{R}^{3 \times 3}$ triangolare superiore con uno sulla diagonale tali che $\Omega\Theta = A$, ovvero tali che:

$$\omega_1 = a_1 \quad , \quad \omega_1 \theta_{12} + \omega_2 = a_2 \quad , \quad \omega_1 \theta_{13} + \omega_2 \theta_{13} + \omega_3 = a_3$$

Se esistono matrici siffatte, allora *necessariamente*:

$$\omega_1 = a_1 \quad , \quad \omega_2 = a_2 - \omega_1 \theta_{12} \quad , \quad \omega_3 = a_3 - \omega_1 \theta_{13} - \omega_2 \theta_{23}$$

e, dalla seconda uguaglianza:

$$\omega_2 \cdot \omega_1 = 0 \quad \text{se e solo se} \quad (\omega_1 \cdot \omega_1) \theta_{12} = a_2 \cdot \omega_1$$

dalla terza:

$$\omega_3 \cdot \omega_1 = 0 \quad \text{se e solo se} \quad (\omega_1 \cdot \omega_1) \theta_{13} + (\omega_2 \cdot \omega_1) \theta_{23} = a_3 \cdot \omega_1$$

$$\omega_3 \cdot \omega_2 = 0 \quad \text{se e solo se} \quad (\omega_1 \cdot \omega_2) \theta_{13} + (\omega_2 \cdot \omega_2) \theta_{23} = a_3 \cdot \omega_2$$

La procedura seguente determina Ω e Θ con le proprietà richieste se e solo se *le colonne di A sono linearmente indipendenti*:

$$\omega_1 = a_1;$$

se $\omega_1 = 0$ allora: interrompi la costruzione;

$$\text{altrimenti: } \theta_{12} = \frac{a_2 \cdot \omega_1}{\omega_1 \cdot \omega_1}; \theta_{13} = \frac{a_3 \cdot \omega_1}{\omega_1 \cdot \omega_1};$$

$$\omega_2 = a_2 - \omega_1 \theta_{12};$$

se $\omega_2 = 0$ allora: interrompi la costruzione;

$$\text{altrimenti: } \theta_{23} = \frac{a_3 \cdot \omega_2}{\omega_2 \cdot \omega_2};$$

$$\omega_3 = a_3 - \omega_1 \theta_{13} - \omega_2 \theta_{23};$$

se $\omega_3 = 0$ allora: interrompi la costruzione;

$$\text{altrimenti: } \Omega = (\omega_1, \omega_2, \omega_3) \quad , \quad \Theta = \begin{bmatrix} 1 & \theta_{12} & \theta_{13} \\ 0 & 1 & \theta_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

– *Secondo passo*

Siano Ω e Θ le matrici determinate dal *Primo passo* e:

$$\Delta = \text{diag}(\|\omega_1\|, \|\omega_2\|, \|\omega_3\|) \in \mathbb{R}^{3 \times 3}$$

Si ricordi che $\omega_1 \neq 0, \omega_2 \neq 0, \omega_3 \neq 0$ e quindi Δ è invertibile. La coppia:

$$U = \Omega \Delta^{-1} \quad , \quad T = \Delta \Theta$$

è una fattorizzazione QR di A .

La procedura seguente, descritta in un linguaggio che consente l'uso del tipo *numero reale*, formalizza il procedimento descritto nell'Esempio:

• $[U, T] = \text{GS}(A)$

// A matrice $n \times n$ ad elementi reali.

//

// ** Primo passo: cerca Ω matrice $n \times n$ a colonne ortogonali non nulle e Θ

// matrice $n \times n$ triangolare superiore con uno sulla diagonale tali che $\Omega\Theta = A$

//

$$\omega_1 = a_1;$$

per $k = 1, \dots, n-1$ ripeti:

se $\omega_k = 0$ allora: interrompi la costruzione e segnala A non invertibile;

altrimenti:

$$d_k = \omega_k \cdot \omega_k$$

$$\text{per } j = k+1, \dots, n \text{ ripeti: } \theta_{kj} = \frac{a_j \cdot \omega_k}{d_k};$$

$$\omega_{k+1} = a_{k+1} - (\omega_1 \theta_{1,k+1} + \dots + \omega_k \theta_{k,k+1});$$

se $\omega_n = 0$ allora: interrompi la costruzione e segnala A non invertibile;

//

// ** Secondo passo: costruisce la fattorizzazione QR normalizzando le colonne di Ω

//

altrimenti:

$$d_n = \omega_n \cdot \omega_n;$$

$$\Delta = \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_n});$$

$$U = (\omega_1, \dots, \omega_n) \Delta^{-1}; \quad T = \Delta \begin{bmatrix} 1 & & & \\ & \ddots & \theta_{ij} & \\ & 0 & \ddots & \\ & & & 1 \end{bmatrix}$$

La procedura GS termina correttamente (ovvero: determina una fattorizzazione QR di A) se e solo se la matrice A è invertibile.

2.31 Osservazione (non unicità della fattorizzazione QR, la funzione predefinita `qr`)

Se U, T è una fattorizzazione QR di A ed $E \in \mathbb{R}^{n \times n}$ è una matrice diagonale tale che $|e_{11}| = 1, \dots, |e_{nn}| = 1$ allora la coppia:

$$U' = UE \quad , \quad T' = ET$$

è a sua volta una fattorizzazione QR di A . Dunque: *la fattorizzazione QR non è unica.*

La procedura GS mostra che *qualunque matrice invertibile ammette fattorizzazione QR*. Esistono altre procedure per la ricerca di una fattorizzazione QR di una matrice, più generali di GS e preferibili ad essa da un punto di vista numerico. Queste procedure terminano correttamente in ogni caso e dunque mostrano che *qualunque matrice $n \times n$ ammette fattorizzazione QR*. Scilab ha una funzione predefinita per il calcolo di una fattorizzazione QR di una matrice che utilizza il *metodo di Householder*:²

– `qr`

Questa *funzione predefinita* restituisce un'approssimazione di una fattorizzazione QR di un'assegnata matrice A . Precisamente, se A è una matrice $n \times n$:

$$[U, T] = \text{qr}(A)$$

restituisce la coppia U, T di matrici $n \times n$, T triangolare superiore, che *approssima* una fattorizzazione QR di A . Come già osservato A può non essere invertibile (vedere l'Esercizio E31).

2.32 Osservazione (uso del tipo *numero in virgola mobile*)

Sia `qr` una procedura che per ogni matrice $A \in \mathbb{R}^{n \times n}$ determina una fattorizzazione QR di A . Si ricordi che, assegnata una matrice $A \in \mathbb{R}^{n \times n}$ ed una colonna $b \in \mathbb{R}^n$, la procedura `qr` permette la ricerca della soluzione sistema $Ax = b$ con il seguente procedimento descritto con un linguaggio che consente l'uso del tipo *numero reale*:

```
(U, T) = qr(A);
se esiste k tale che tkk = 0 allora arresta il procedimento e dichiara A non invertibile;
altrimenti
x* = SI(T, UTb)
```

Sostituendo al tipo *numero reale* il tipo *numero in virgola mobile e precisione finita* il procedimento si modifica come segue:

```
(Ũ, T̂) = qr(A');
se esiste k tale che t̂kk = 0 allora arresta il procedimento e dichiara A' non invertibile;
altrimenti
ξ = SI(T̂, ŨT*b̂)
```

dove: A' e b' indicano, rispettivamente, la matrice e la colonna ottenute arrotondando in M ciascuna componente di A e b , `qr` ed `SI` indicano le procedure ottenute sostituendo il tipo *numero reale* rispettivamente nelle procedure `qr` ed `SI` e $\hat{U}^T * \hat{b}$ indica la matrice ottenuta sostituendo le pseudo-operazioni aritmetiche alle operazioni aritmetiche nel prodotto riga per colonna $\hat{U}^T \hat{b}$.

Il vettore finale ξ è utilizzato per approssimare x^* e l'approssimazione è *sempre* accurata quanto consentito dal condizionamento di A . Infatti, anche in questo caso per giudicare l'accuratezza occorre studiare il *condizionamento* del calcolo della soluzione del sistema $Tx = U^T b$, ovvero indagare il *numero di condizionamento* della matrice T .

Scelta in \mathbb{R}^n la norma euclidea N_2 , si studia:

$$c_2(T) = \|T^{-1}\|_2 \|T\|_2$$

Si ha:

$$A = UT \quad \Rightarrow \quad T = U^T A \quad \text{e} \quad A^{-1} = T^{-1} U^T \quad \Rightarrow \quad T^{-1} = A^{-1} U$$

²Si veda, ad esempio: https://en.wikipedia.org/wiki/QR_decomposition#Using_Householder_reflections.

Dunque:

$$\|T\|_2 \leq \|U^T\|_2 \|A\|_2 \quad \text{e} \quad \|T^{-1}\|_2 \leq \|A^{-1}\|_2 \|U\|_2$$

Ma: per ogni matrice M ortogonale si ha $\|M\|_2 = 1$. Perciò:

$$\|T\|_2 \leq \|A\|_2 \quad \text{e} \quad \|T^{-1}\|_2 \leq \|A^{-1}\|_2 \quad \Rightarrow \quad c_2(T) \leq c_2(A)$$

In questo caso la procedura di fattorizzazione QR produce un fattore destro con proprietà di condizionamento non peggiori di quelle della matrice iniziale.

Esercizi

E27 Sia:

$$A = \begin{bmatrix} 0 & 1 & 6 \\ 1 & -1 & 1 \\ 2 & 3 & 2 \end{bmatrix}$$

Determinare $\text{GS}(A)$.

E28 ★ Verificare che, se A è una matrice invertibile, la coppia U, T definita da $\text{GS}(A)$ è una fattorizzazione QR di A .

E29 ★ Sia $A \in \mathbb{R}^{3 \times 3}$. Dimostrare che se $\text{GS}(A)$ termina prematuramente allora A non è invertibile.

E30 Sia:

$$A = \begin{bmatrix} 0 & 2 \\ 1 & -1 \end{bmatrix}$$

Determinare $\text{GS}(A)$ e dedurre dal risultato due diverse fattorizzazioni QR di A .

E31 ♠ Sia $A \in \mathbb{R}^{n \times n}$ la matrice nulla. Per $n = 5$ verificare che la procedura GS applicata ad A termina prematuramente mentre $\text{qr}(A)$ determina una fattorizzazione QR esatta.

E32 ★ Dimostrare, utilizzando la definizione di norma indotta, che: se M è una matrice ortogonale allora $\|M\|_2 = 1$.

2.8 Costo

La nozione di *costo*, già implicitamente adottata in precedenza, è quella “aritmetica.”

2.33 Definizione (costo aritmetico)

Sia ϕ un algoritmo. Si chiama *costo aritmetico* di ϕ il numero $C(\phi)$ di pseudo-operazioni aritmetiche eseguite per calcolare $\phi(x)$.³

Si osservi che il costo di un algoritmo deve essere una quantità almeno approssimativamente proporzionale al *tempo* necessario al calcolatore per portare a termine il calcolo. La definizione adottata riesce nell'intento se:

- La maggior parte del tempo impiegato dal calcolatore per calcolare ϕ è speso nell'esecuzione di pseudo-operazioni aritmetiche.
- Il tempo necessario per eseguire ciascuna pseudo-operazione aritmetica è lo stesso, in particolare *non dipende dagli operandi*.

³Il costo va valutato *nel caso peggiore*: il numero di pseudo-operazioni aritmetiche eseguite per calcolare $\phi(x)$ potrebbe dipendere da x .

Il sussistere della condizione (a) *dipende da ϕ* . Ad esempio, se ϕ è un algoritmo per il calcolo della norma infinito di un vettore il costo aritmetico è *zero* – per questo algoritmo *nessuna* delle funzioni predefinite calcolate è una pseudo-operazione aritmetica – ma non è vero che il calcolatore impiega tempo zero a calcolare $\phi(x)$. Nel seguito la definizione di costo sarà applicata solo ad algoritmi che calcolano *quasi esclusivamente* pseudo-operazioni aritmetiche.

Il sussistere della condizione (b), invece, *non dipende da ϕ* ma *dipende dalla scelta di M* . Si consideri, ad esempio, il calcolo in $F(2, 53)$ dello pseudo-prodotto $2^{b_1} \otimes 2^{b_2} = 2^{b_1+b_2}$. *Non è ragionevole* supporre che il tempo necessario per il calcolo sia indipendente da quali numeri interi b_1 e b_2 occorre sommare (calcolare la somma si due numeri interi *non può* richiedere un tempo *indipendente* dal numero di cifre necessario per rappresentare gli addendi – si pensi al solo tempo necessario per *leggere gli addendi e scrivere la somma*). L'ipotesi (b) è verificata, invece, qualora M sia un insieme di numeri in virgola mobile con *esponente limitato*.

Vediamo alcuni esempi di determinazione del costo di un algoritmo e poi confrontiamo il costo degli algoritmi dedotti dai due procedimenti proposti per la ricerca della soluzione di un sistema di equazioni lineari.

2.34 Esempio

– *Prodotto riga per colonna*

Sia prc_n l'algoritmo ottenuto dal procedimento di calcolo del *prodotto riga per colonna* $a^\top b$, con $a, b \in \mathbb{R}^n$, sostituendo ciascuna operazione aritmetica con la corrispondente pseudo-operazione e specificando l'ordine di composizione delle pseudo-operazioni. La corrispondenza biunivoca tra operazioni aritmetiche e pseudo-operazioni così stabilita rende possibile determinare il costo di $\text{prc}_n(a, b)$ calcolando il numero di operazioni aritmetiche in $a^\top b$. Dette a_i, b_i le componenti di a, b si ha:

$$a^\top b = a_1 b_1 + \cdots + a_n b_n$$

Il calcolo di $a^\top b$ richiede n prodotti e $n - 1$ somme. Allora:

$$C(\text{prc}_n) = 2n - 1$$

– *Prodotto matrice per colonna*

Sia pmc_n l'algoritmo ottenuto dal procedimento di calcolo del *prodotto matrice per colonna* Ab , con $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$. Si ragiona come nel caso del prodotto riga per colonna, ovvero si calcola il numero di operazioni aritmetiche in Ab . Dette r_1, \dots, r_n le righe di A si ha:

$$Ab = \begin{bmatrix} r_1 b \\ \vdots \\ r_n b \end{bmatrix}$$

Ciascuna delle n componenti del vettore Ab è un prodotto riga per colonna. Il calcolo di Ab richiede quindi n^2 prodotti e $n(n - 1)$ somme. Allora:

$$C(\text{pmc}_n) = nC(\text{prc}_n) = 2n^2 - n$$

– *Prodotto matrice triangolare per colonna*

Sia pmtc_n l'algoritmo ottenuto dal procedimento di calcolo del *prodotto matrice triangolare per colonna* Tc , con $T \in \mathbb{R}^{n \times n}$ matrice triangolare, $c \in \mathbb{R}^n$. Si ragiona come nel caso del prodotto matrice per colonna. Dette r_1, \dots, r_n le righe di T si ha:

$$Tc = \begin{bmatrix} r_1 c \\ \vdots \\ r_n c \end{bmatrix}$$

Ciascuna delle n componenti del vettore Tc è un prodotto riga per colonna. Questa volta, però, ciascuna componente ha un costo diverso. Supponendo, ad esempio, T triangolare superiore ed evitando di calcolare operazioni con risultato noto a priori (per ogni $x \in \mathbb{R}$ si ha $0x = 0$ e $0 + x = x$) si ottiene:

$$C(\text{pmtc}_n) = C(\text{prc}_n) + C(\text{prc}_{n-1}) + \cdots + C(\text{prc}_1) = 2(1 + 2 + \cdots + n) - n = n^2$$

Il costo di pmtc_n è circa *metà* del costo di pmc_n .

– *Procedura SI*

Come già sappiamo dal Paragrafo 2.1 il calcolo di $SI(T, c)$ per $T \in \mathbb{R}^{n \times n}$ e $c \in \mathbb{R}^n$ richiede n divisioni e $\frac{1}{2}n(n-1)$ prodotti e somme. Allora:

$$C(SI) = n^2$$

Il costo è uguale a quello di \mathbf{pmc}_n . Gli stessi risultati si ottengono per \mathbf{SA} .

– *Procedura EGPP*

Il calcolo di $EGPP(A)$ per $A \in \mathbb{R}^{n \times n}$ risulta richiedere:

$$\sum_{k=1}^{n-1} k = \frac{1}{2}n(n-1) \text{ divisioni e } \sum_{k=1}^{n-1} k^2 = \frac{1}{6}(n-1)n(2n-1) \text{ prodotti e somme}$$

In totale:

$$C(EGPP) = \frac{2}{3}n^3 - \frac{1}{2}n^2 - \frac{1}{6}n$$

Si osservi che il calcolo di $EGPP(A)$ richiede anche *confronti*, in numero trascurabile rispetto alle pseudo-operazioni aritmetiche.

Il procedimento per per la ricerca della soluzione di un sistema di equazioni lineari che utilizza la fattorizzazione LR ha dunque costo:

$$C(EGPP) + C(\mathbf{SA}) + C(SI) = \frac{2}{3}n^3 + \text{termini di grado inferiore in } n$$

Un calcolo analogo per il procedimento per per la ricerca della soluzione di un sistema di equazioni lineari che utilizza la fattorizzazione QR porta ad un costo:⁴

$$C(\mathbf{qr}) + C(\mathbf{pmc}_n) + C(SI) = \frac{4}{3}n^3 + \text{termini di grado inferiore in } n$$

che è circa *il doppio* del precedente. Si osservi che si è scelto di esprimere il costo dei procedimenti mostrando esplicitamente solo *il termine dominante* al crescere di n . In entrambi i casi *il termine dominante è generato dalla procedura di ricerca della fattorizzazione*.

2.35 Osservazione (calcolo della matrice inversa)

La funzione predefinita `inv` di *Scilab* cerca un'approssimazione della matrice inversa di una data matrice usando la procedura EGPP. Sia $A \in \mathbb{R}^{n \times n}$ una matrice invertibile. Dette e_1, \dots, e_n le colonne della matrice identica, il calcolo di $Y = A^{-1}$ avviene in questo modo:

$$[S, D, P] = EGPP(A)$$

se $d_{kk} = 0$ per qualche k allora arresta la procedura e dichiara A non invertibile;

altrimenti

per $k = 1, \dots, n$ ripeti:

$$c = \mathbf{SA}(S, Pe_k);$$

$$y_k = \mathbf{SI}(D, c);$$

$$Y = (y_1 \dots, y_n).$$

Dunque si calcolano le n colonne della matrice inversa come soluzione degli n sistemi lineari $Ay = e_1, \dots, Ay = e_n$. Tutti i sistemi hanno *la stessa matrice* e per la soluzione degli n sistemi è sufficiente calcolare *una sola volta* la fattorizzazione di A . Questo fa sì che il termine dominante del costo del procedimento sia ancora un multiplo di n^3 .

Esercizi

E33 ★ Verificare che il calcolo della matrice H_k nel passo k -esimo di $EGPP(A)$ richiede $n - k$ divisioni e che il calcolo del prodotto $H_k P_k A^{(k)}$ richiede $(n - k)^2$ prodotti e somme.

⁴Si veda la pagina di Wikipedia citata in precedenza. Si osservi che il calcolo della fattorizzazione QR richiede anche $n - 1$ radici quadrate.

- E34* Verificare che il calcolo di $\text{EGPP}(A)$ richiede non più di $\frac{1}{2}n(n-1)$ confronti. Determinare l'errore relativo commesso approssimando il numero di pseudo-operazioni aritmetiche e confronti richiesto dal calcolo di $\text{EGPP}(A)$ con il numero delle sole pseudo-operazioni aritmetiche.
- E35* Sia ata_n l'algoritmo ottenuto dal procedimento di calcolo del prodotto $A^\top A$ con $A \in \mathbb{R}^{n \times n}$. Determinare $C(\text{ata}_n)$. Si tenga conto che la matrice $A^\top A$ è *simmetrica*.
-