Matrici elementari e fattorizzazioni

Dario A. Bini, Università di Pisa

1 febbraio 2020

Sommario

Questo modulo didattico introduce ed analizza la classe delle matrici elementari. Tale classe verrà usata per costruire algoritmi di fattorizzazione di matrici e risolvere sistemi lineari.

1 Introduzione

Si introduce la classe delle matrici elementari che hanno proprietà computazionali interessanti e permettono di costruire algoritmi efficienti per calcolare le principali fattorizzazioni di matrici. Poi si considerano due sottoclassi particolari di matrici elementari: le matrici elementari di Gauss che sono triangolari inferiori, le matrici elementari di Householder che sono unitarie e hermitiane. Successivamente mostriamo come le matrici elementari possono essere usate per calcolare una generica fattorizzazione del tipo $A = SU$ dove S è un prodotto di matrici elementari e U è triangolare superiore. La specializzazione nella scelta delle matrici elementari alle matrici di Gauss e di Householder conduce ai metodi di Gauss e di Householder per la fattorizzazione LU e QR di una matrice.

2 Matrici elementari

Siano $u, v \in \mathbb{C}^n$. Per ragioni di chiarezza osserviamo che, con le nostre notazioni, l'espressione $v^H u = \sum_{i=1}^n \bar{v}_i u_i$ fornisce un numero complesso; mentre l'espressione uv^H, prodotto righe per colonne di un vettore colonna e di un vettore riga, fornisce una matrice $n \times n$ di elementi $u_i \bar{v}_j$. Ciò premesso, siamo pronti per dare la definizione di matrice elementare

Definizione 1 Una matrice del tipo

$$M = I - \sigma u v^H,$$

dove I è la matrice identica $n \times n$, σ un numero complesso e $u, v \in \mathbb{C}^n$, è detta *matrice elementare.*

Si osservi che nella definizione data c'è ridondanza di parametri, infatti il parametro scalare σ potrebbe essere inglobato in uno dei due vettori, inoltre uno dei due vettori potrebbe essere normalizzato a piacimento. Questa ridondanza ci permette una maggior facilità nell'analizzare le proprietà computazionali di questa classe di matrici.

Si osservi ancora che se $x \in \mathbb{C}^n$ allora $Mx = x - \sigma u(v^H x)$, per cui tutti i vettori dello spazio ortogonale a v, cioè tali che $v^H x = 0$, vengono trasformati da M in se stessi. Se d'altro canto $x = u$ allora $Mu = (1 - \sigma v^H u)u$, cioè $1 - \sigma v^H u$ è autovalore di M corrispondente all'autovettore u. In particolare se $\sigma v^H u = 1$ allora M è singolare. Viceversa, se M è singolare, esiste $x \neq 0$ tale che $Mx = 0$, cioè deve essere $x = \sigma(v^H x)u$, $v^H x \neq 0$. Per cui a meno di una costante moltiplicativa vale $x = u$, e $\sigma v^H u = 1$.

Studiamo ora alcune proprietà computazionali della classe di matrici elementari. Per semplicità assumiamo che $\sigma uv^H \neq 0$. Questa condizione esclude il caso in cui $M = I$ che non ha bisogno di ulteriore analisi.

Poiché M lascia l'intero sottospazio ortogonale a v invariato, e ha u come autovettore, così fa l'inversa di M se $\det M \neq 0$. Quindi viene naturale cercare l'inversa di M nella classe delle matrici elementari dello stesso tipo di M. Proviamo allora a cercare un numero complesso τ tale che $M^{-1} = (I - \tau uv^H)$. Cioè imponiamo la condizione $(I - \tau uv^H)M = I$, dove naturalmente supponiamo che $\sigma v^H u \neq 1$, condizione necessaria e sufficiente di invertibilità. Sviluppando i calcoli si ottiene

$$(\tau\sigma(v^H u) - \tau - \sigma)uv^H = 0$$

che è verificata se e solo se

$$\tau(1 - \sigma v^H u) = -\sigma.$$

Se $\sigma v^H u = 1$ abbiamo già osservato che la matrice non è invertibile, infatti l'equazione di sopra non ha soluzione. Se $\sigma v^H u \neq 1$ allora l'equazione ha soluzione

$$\tau = \frac{-\sigma}{1 - \sigma v^H u}.$$

Si può sintetizzare questo risultato col seguente

Teorema 1 *La matrice $M = I - \sigma uv^H$ è non singolare se e solo se $\sigma v^H u \neq 1$ e la sua inversa è data da*

$$I - \tau uv^H, \qquad \tau = \frac{-\sigma}{1 - \sigma v^H u}.$$

Il calcolo di τ richiede l'esecuzione di $n+1$ moltiplicazioni, n addizioni e una divisione. Un'altra proprietà interessante è riportata nel seguente

Teorema 2 *Data $M = I - \sigma uv^H$ matrice elementare e dato un vettore b, il calcolo del prodotto $y = Mb$ costa $2n + 1$ moltiplicazioni e $2n - 1$ addizioni. La risoluzione del sistema $Mx = b$ costa $3n + 2$ moltiplicazioni $3n - 1$ addizioni e una divisione.*

Una proprietà importante delle matrici elementari è che sono in grado di trasformare un qualsiasi vettore non nullo in un qualsiasi altro vettore non nullo come è precisato nel seguente

Teorema 3 *Per ogni $x, y \in \mathbb{C}^n \backslash 0$ esiste una matrice elementare $M = I - \sigma uv^H$ non singolare tale che $Mx = y$.*

Dim. Si procede in modo costruttivo. La condizione $(I - \sigma uv^H)x = y$ si riscrive come $\sigma u(v^H x) = x - y$. Basta quindi scegliere v non ortogonale a x e porre $\sigma u = (x - y)/(v^H x)$. Per avere la non singolarità di M basta imporre la condizione $\sigma v^H u \neq 1$, cioè $v^H(x - y)/(v^H x) \neq 1$, che è equivalente a $v^H y \neq 0$. Basta allora scegliere v in modo che non sia ortogonale né a x né a y. $\square$

3 Matrici elementari di Gauss

Una matrice elementare di Gauss è ottenuta ponendo $v = e^{(1)}$, primo versore della base canonica, $\sigma = 1$ e u tale che $u_1 = 0$. Cioè per una matrice elementare di Gauss M vale

$$M = I - uv^H = \begin{bmatrix} 1 & & & 0 \\ -u_2 & 1 & & \\ \vdots & 0 & \ddots & \\ -u_n & & & 1 \end{bmatrix}$$

È facile verificare che $M^{-1} = I + uv^H$ cioè

$$M^{-1} = \begin{bmatrix} 1 & & & 0 \\ u_2 & 1 & & \\ \vdots & 0 & \ddots & \\ u_n & & & 1 \end{bmatrix}.$$

Quindi invertire una matrice elementare di Gauss non richiede alcuna operazione aritmetica. Basta cambiare segno agli elementi della prima colonna escluso l'elemento diagonale.

Si osservi che se $x = (x_i)$ è tale che $x_1 \neq 0$ allora esiste una matrice elementare di Gauss M tale che $Mx = x_1 e^{(1)}$. Infatti basta porre $u_i = x_i/x_1$, cioè

$$M = \begin{bmatrix} 1 & & & 0 \\ -x_2/x_1 & 1 & & \\ \vdots & 0 & 1 & \\ -x_n/x_1 & & & 1 \end{bmatrix}.$$

Si osservi che le matrici M e M^{-1} hanno entrambe norma infinito uguale a $1 + \max_i |x_i|/|x_1|$ quindi il numero di condizionamento in norma infinito di M

è $(1+\max_i |x_i|/|x_1|)^2$. Se poi x è tale che $|x_1| = \max_i |x_i|$ allora il numero di condizionamento di M in norma infinito è al più 4. È quindi indipendente dalla dimensione n.

Un'altra classe di matrici elementari con numero di condizionamento indipendente da n è quella delle matrici di Householder.

4 Matrici elementari di Householder

Una matrice elementare di Householder è una matrice elementare hermitiana e unitaria. Vale quindi $M = I - \beta uu^H$ con $\beta = 0$ oppure $\beta = 2/(u^H u)$ e $u \neq 0$. Infatti, risulta

$$(I - \beta uu^H)(I - \beta uu^H)^H = (I - \beta uu^H)^2 = I - 2\beta uu^H + \beta^2(u^H u)uu^H = I.$$

Evidentemente l'inversa di una matrice di Householder M è M stessa. Inoltre in norma 2 il condizionamento di M è 1 poiché le matrici unitarie hanno norma 2 unitaria.

Non è difficile costruire una matrice di Householder che trasformi un vettore x non nullo in un vettore del tipo $\alpha e^{(1)}$. Infatti, si osserva subito che, essendo M unitaria essa trasforma i vettori lasciando invariata la norma 2, risulta quindi $|\alpha| = \|x\|_2$. Inoltre, poiché M è hermitiana, il valore di $x^H M x$ è reale qualunque sia x. Questo implica che $x^H \alpha e^{(1)} = \bar{x}_1 \alpha$ è reale. Ciò permette di determinare il valore di α. Infatti di α conosciamo il modulo, quindi α è determinato a meno di un fattore complesso di modulo 1. Vale cioè $\alpha = \theta \|x\|_2$, con $|\theta| = 1$. Si verifica facilmente che se poniamo

$$\theta = \begin{cases} \pm x_1/|x_1| & \text{se } x_1 \neq 0, \\ \pm 1 & \text{se } x_1 = 0, \end{cases}$$

nel caso $x_1 \neq 0$ risulta

$$\bar{x}_1 \alpha = \pm \bar{x}_1 x_1/|x_1| \, \|x\|_2 = \pm |x_1| \, \|x\|_2 \in \mathbb{R},$$

se invece $x_1 = 0$ si ha $\bar{x}_1 \alpha = 0 \in \mathbb{R}$.

In teoria tutte e due i segni vanno bene, ma per ragioni di stabilità numerica vedremo tra poco che la scelta obbligata è il segno meno.

A questo punto siamo pronti per determinare il vettore u e conseguentemente lo scalare $\beta = 2/(u^H u)$. Infatti, dalla condizione $Mx = \alpha e^{(1)}$ si deduce che $(I - \beta uu^H)x = \alpha e^{(1)}$ e quindi

$$\beta(u^H x)u = x - \alpha e^{(1)}.$$

Ciò permette di determinare il vettore u a meno della sua lunghezza. Ma la lunghezza di u non è rilevante poichè questa informazione viene inglobata nel parametro β. Infatti basta porre

$$u = x - \alpha e^{(1)}$$

e ricavare β dalla condizione $\beta = 2/(u^H u)$. Infatti con questa scelta di β si verifica facilmente che $\beta(u^H x) = 1$. Si osserva che il vettore u ha tutte le componenti uguali a quelle di x tranne la prima che, nel caso $x_1 \neq 0$ è

$$u_1 = x_1 - \alpha = x_1 \mp \theta \|x\|_2 = x_1(1 \mp \|x\|_2/|x_1|).$$

È evidente che nella determinazione di θ conviene optare per il segno meno in modo che nella formula precedente non ci sia cancellazione numerica. Chiaramente, nel caso $x_1 = 0$ risulta $u_1 = -\alpha = \mp\|x\|_2$. In questo caso il segno non è influente ai fini della stabilità numerica. Per semplicità scegliamo ancora il segno meno. In questo modo si ha

$$u_1 = \begin{cases} x_1(1 + \frac{1}{|x_1|}\|x\|_2) & \text{se } x_1 \neq 0, \\ \|x\|_2 & \text{se } x_1 = 0, \end{cases}$$

e vale

$$\beta = 1/(\|x\|_2^2 + |x_1|\,\|x\|_2). \tag{1}$$

Il costo computazionale del calcolo di u e di β è dominato dal calcolo di $\|x\|_2$, cioè n moltiplicazioni e $n-1$ addizioni più una estrazione di radice. Si osservi inoltre che da (1) segue che $\beta(u^H x) = 1$ come richiesto.

5 Fattorizzazione mediante matrici elementari

Mostriamo come sia possibile utilizzare le matrici elementari per realizzare metodi per fattorizzare una matrice A nel prodotto $A = SU$ dove

$$S = M_1^{-1} M_2^{-1} \cdots M_{n-1}^{-1}$$

è un prodotto di matrici elementari ed U è una matrice triangolare superiore. La disponibilità di tale fattorizzazione ci permette di risolvere agevolmente il sistema $Ax = b$ attraverso la risoluzione dei due sistemi $Sy = b$ e $Ux = y$. Il secondo, essendo con matrice triangolare superiore, si risolve con n^2 operazioni aritmetiche. Il primo permette di esprimere la soluzione mediante la formula

$$y = M_{n-1} \cdots M_1 b$$

e quindi, se le matrici elementari $M_1, \ldots, M_{n-1}$ sono disponibili, permette di calcolare la soluzione in circa $3n^2$ moltiplicazioni e altrettante addizioni grazie al teorema 2. Cioè, data la fattorizzazione di A, il sistema $Ax = b$ è risolubile in $O(n^2)$ operazioni aritmetiche. Quindi tutto è ricondotto al calcolo della fattorizzazione $A = SU$. Vediamo ora come si realizza.

Posto $A_1 = A$ andiamo a generare una successione di matrici $A_k = (a_{i,j}^{(k)})$, $k = 1, 2 \ldots, n$ tali che $A_{k+1} = M_k A_k$ dove M_k è una matrice elementare e dove A_k ha le prime $k-1$ colonne in forma triangolare, cioè

$$A_k = \left[\begin{array}{ccc|ccc} a_{1,1}^{(k)} & \dots & a_{1,k-1}^{(k)} & a_{1,k}^{(k)} & \dots & a_{1,n}^{(k)} \\ & \ddots & \vdots & \vdots & & \vdots \\ 0 & & a_{k-1,k-1}^{(k)} & a_{k-1,k}^{(k)} & \dots & a_{k,n}^{(k)} \\ \hline 0 & \dots & 0 & a_{k,k}^{(k)} & \dots & a_{k+1,n}^{(k)} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & a_{n,k}^{(k)} & \dots & a_{n,n}^{(k)} \end{array}\right] =: \left[\begin{array}{c|c} T_k & V_k \\ \hline 0 & W_k \end{array}\right] \tag{2}$$

dove T_k ha dimensione $(k-1)\times(k-1)$, W_k ha dimensione $(n-k+1)\times(n-k+1)$ e V_k ha dimensione $(k-1)\times(n-k+1)$.

Vediamo il primo passo. Consideriamo la matrice $A = A_1$, denotiamo $a^{(1)}$ la sua prima colonna. Per il teorema 3 esiste una matrice elementare M_1 che trasforma $a^{(1)}$ in un vettore proporzionale al primo vettore $e^{(1)}$ della base canonica di $\mathbb{C}^n$, cioè M_1 è tale che $M_1 a^{(1)} = a_{1,1}^{(2)} e^{(1)}$. Vale allora

$$M_1 A_1 = \left[\begin{array}{c|c} a_{1,1}^{(2)} & V_2 \\ \hline 0 & W_2 \end{array}\right] =: A_2$$

dove W_2 è matrice $(n-1)\times(n-1)$ mentre V_2 ha dimensioni $1\times(n-1)$. Cioè mediante la moltiplicazione per M_1 abbiamo iniziato il primo passo del processo di triangolarizzazione di A. Adesso ripetiamo lo stesso procedimento sulla matrice W_2 costruendo una matrice elementare che trasformi la prima colonna di W_2 in un vettore proporzionale al primo versore della base canonica di $\mathbb{C}^{n-1}$. Descriviamo questo processo nella sua generalità.

Supponiamo di avere la matrice A_k e di voler costruire la matrice M_k tale che $A_{k+1} = M_k A_k$ abbia le prime k colonne in forma triangolare. Per questo si consideri una matrice elementare $\widehat{M}_k$ di dimensione $(n-k+1)\times(n-k+1)$ che trasformi la prima colonna di W_k in un vettore proporzionale al primo versore della base canonica di $\mathbb{C}^{n-k+1}$. Tale matrice esiste per il teorema 3. Vale allora

$$\widehat{M}_k W_k = \left[\begin{array}{c|c} a_{k,k}^{(k+1)} & z^T \\ \hline 0 & W_{k+1} \end{array}\right]$$

con W_{k+1} di dimensioni $(n-k)\times(n-k)$.

Quindi ponendo

$$M_k = \left[\begin{array}{c|c} I_{k-1} & 0 \\ \hline 0 & \widehat{M}_k \end{array}\right] \tag{3}$$

dove I_{k-1} è una matrice identica di ordine $k-1$, si ha

$$M_k A_k = \left[\begin{array}{c|c} T_k & V_k \\ \hline 0 & \widehat{M}_k W_k \end{array}\right] = \left[\begin{array}{c|c} T_k & V_k \\ \hline 0 & \begin{array}{c|c} a_{k,k}^{(k+1)} & z^T \\ \hline 0 & W_{k+1} \end{array} \end{array}\right] := A_{k+1}$$

dove A_{k+1} ha la stessa struttura in (2) con k sostituito da $k+1$.

Inoltre la matrice M_k è ancora una matrice elementare essendo $M_k = I - \sigma u v^H$ dove

$$u = \begin{bmatrix} 0 \\ \hat{u} \end{bmatrix}, \qquad v = \begin{bmatrix} 0 \\ \hat{v} \end{bmatrix}.$$

con $\widehat{M}_k = I_{n-k+1} - \sigma \hat{u}\hat{v}^H$.

In questo modo si è realizzato il generico passo del processo di triangolarizzazione di A.

Dopo $n-1$ passi si ottiene la fattorizzazione $A_n = M_{n-1} \cdots M_1 A$, dove A_n è triangolare superiore, da cui si ottiene

$$A = M_1^{-1} \cdots M_{n-1}^{-1} A_n.$$

Si osservi che per risolvere il sistema lineare $Ax = b$, la tecnica di fattorizzazione appena introdotta può essere utilizzata in due modi diversi ma equivalenti.

- Calcolare e memorizzare ogni singola matrice elementare M_k assieme alla matrice A_n triangolare superiore e alla fine calcolare $y = M_{n-1} \cdots M_1 b$ mediante prodotti successivi, e poi risolvere il sistema triangolare $A_n x = y$.

- Costruire la successione di sistemi equivalenti $A_k x = b^{(k)}$, dove $b^{(k+1)} = M_k b^{(k)}$ e $b^{(1)} = b$, e alla fine risolvere il sistema $A_n x = b^{(n)}$.

La differenza tra i due approcci riguarda solo la tempistica delle operazioni.

Nel secondo caso si applica una strategia usa e getta nel calcolo delle matrici M_k che comporta un ingombro di memoria più basso rispetto al primo approccio.

Computazionalmente i due metodi sono equivalenti visto che entrambi calcolano, anche se in tempi diversi, gli $n-1$ prodotti matrice-vettore $y = b^{(n)} = M_{n-1} \cdots M_1 b$.

6 Metodi di Gauss e di Householder

Se ad ogni passo del metodo descritto sopra si sceglie come M_k una matrice del tipo (3) dove $\widehat{M}_k = I - \hat{\beta}_k \hat{u}^{(k)} \hat{u}^{(k)H}$ è matrice di Householder, allora anche M_k è matrice di Householder essendo $M_k = I - \beta_k u^{(k)} u^{(k)H}$ con

$$u^{(k)} = \begin{bmatrix} 0 \\ \hat{u}^{(k)} \end{bmatrix}$$

e $\beta_k = \hat{\beta}_k$. La fattorizzazione che si ottiene in questo modo è una fattorizzazione QR dove Q è un prodotto di matrici di Householder e quindi è unitaria.

Poiché esiste sempre una matrice di Householder che trasforma un arbitrario vettore in un vettore proporzionale al primo versore della base canonica, la costruzione mostrata può essere sempre portata a termine senza interruzioni.

Questo dimostra in modo costruttivo che la fattorizzazione QR di una matrice esiste sempre.

Scegliendo invece ad ogni passo come matrice elementare la matrice M_k descritta in (3) dove $\widehat{M}_k$ è una matrice di Gauss si arriva alla fattorizzazione LU.

Si osserva che la matrice $\widehat{M}_k$ ha elementi

$$\widehat{M}_k = \begin{bmatrix} 1 & & & 0 \\ -a_{2,k}^{(k)}/a_{k,k}^{(k)} & 1 & & \\ \vdots & & \ddots & \\ -a_{n,k}^{(k)}/a_{k,k}^{(k)} & & & 1 \end{bmatrix}. \tag{4}$$

Diversamente dalle matrici di Householder, l'esistenza della matrice elementare di Gauss è legata alla condizione $a_{k,k}^{(k)} \neq 0$. Infatti se tale condizione non fosse verificata la matrice di Gauss (4) che realizza un passo del processo di triangolarizzazione non esisterebbe in generale. L'elemento $a_{k,k}^{(k)}$ viene chiamato *elemento pivot*.

È possibile verificare agevolmente che se tutte le sottomatrici principali di testa di A di dimensione $k \times k$ sono non singolari per $k = 1, \ldots, n-1$, per cui esiste ed è unica la fattorizzazione LU, allora tutti gli elementi pivot sono diversi da zero e quindi il processo di triangolarizzazione può essere portato a termine senza interruzioni.

Infatti, supponiamo per assurdo che il metodo di fattorizzazione LU appena visto si interrompa al passo k-esimo poiché $a_{k,k}^{(k)} = 0$.

Allora, dalla relazione $A_k = M_{k-1} \cdots M_1 A$, poiché $L_k = M_{k-1} \cdots M_1$è triangolare inferiore si ha che la sottomatrice principale di testa di A_k di dimensione $k \times k$ è uguale alla sottomatrice principale di testa di L_k per la sottomatrice principale di testa di A che è non singolare per ipotesi.

Questo implica che la sottomatrice principale di testa $k \times k$ di A_k è non singolare. Ma questo è assurdo essendo tale matrice triangolare superiore con elementi diagonali $a_{i,i}^{(k)}$ ed essendo $a_{k,k}^{(k)} = 0$.

I metodi per il calcolo della fattorizzazione LU e QR che si ottengono nel modo descritto sono detti rispettivamente metodo di Gauss, o metodo di eliminazione Gaussiana, e metodo di Householder. Una loro analisi computazionale viene svolta nel prossimo articolo Aspetti computazionali dei metodi di Gauss e Householder.

7 Il complemento di Schur

Si partizioni la matrice A nel seguente modo

$$A = \left[\begin{array}{c|c} A_{1,1} & A_{1,2} \\ \hline A_{2,1} & A_{2,2} \end{array}\right]$$

dove $A_{1,1}$ è $k \times k$ e non singolare. La matrice

$$S = A_{2,2} - A_{2,1} A_{1,1}^{-1} A_{1,2}$$

è definita il *complemento di Schur* di $A_{2,2}$ in A. Il complemento di Schur è legato alla fattorizzazione LU di A. Infatti si verifica facilmente che vale la fattorizzazione a blocchi

$$A = \begin{bmatrix} I & 0 \\ A_{2,1}A_{1,1}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & S \end{bmatrix}. \tag{5}$$

Inoltre S coincide con la matrice W_k che compare in (2).

Il complemento di Schur ha proprietà interessanti. Ad esempio, da (5) segue che $\det A = \det A_{1,1} \det S$. Quindi se A è invertibile anche S lo è. Inoltre S^{-1} coincide con la sottomatrice principale di A^{-1} formata dagli indici $i, j > k$.

Riferimenti bibliografici

[1] D. Bini, M. Capovani, O. Menchi. Metodi Numerici per l'Algebra Lineare. Zanichelli, Bologna 1988.