Note del Corso di Calcolo Numerico

versione del 8 aprile 2018 ore 23:59

Alessandro Russo

Dipartimento di Matematica e Applicazioni, Università di Milano-Bicocca
E-mail address: alessandro.russo@unimib.it

Indice

Capitolo 5

Decomposizione di Cholesky

5.1. Matrici simmetriche e definite positive

Una classe molto importante di matrici sono le cosiddette matrici *simmetriche e definite positive*. Queste matrici hanno molte belle proprietà e possiamo dire che sono le matrici più facilmente "trattabili" numericamente.

Una matrice A si dice simmetrica e definita positiva se:

(1) $A^{\mathrm{T}} = A$ (simmetria);

(2) $(A\boldsymbol{x}, \boldsymbol{x}) \geq 0$, e $(A\boldsymbol{x}, \boldsymbol{x}) = 0$ se e solo se $\boldsymbol{x} = 0$ (definita positività)

dove $(\cdot, \cdot)$ è il prodotto scalare euclideo: $(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} x_i y_i$.

Iniziamo a vederne alcune proprietà.

- A è invertibile. Infatti $A\boldsymbol{x} = 0$ implica $(A\boldsymbol{x}, \boldsymbol{x}) = 0$ e quindi $\boldsymbol{x} = 0$ per la (2). Pertanto $\ker(A) = \{0\}$ e A è invertibile.

- Gli elementi sulla diagonale di A sono strettamente positivi. Infatti $a_{ii} = (A\boldsymbol{e}_i, \boldsymbol{e}_i) > 0$ per la (2). Più in generale, si può dimostrare che vale il seguente Criterio di Sylvester:

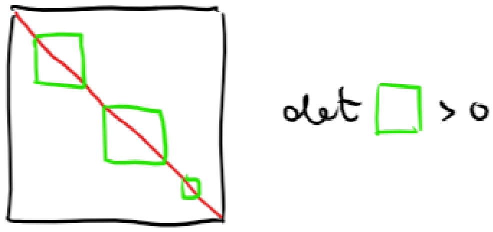
  Teorema 5.1. *Una matrice A simmetrica è definita positiva se e solo se vale la seguente proprietà: detta A_k, $k = 1, \ldots, n$ la matrice ottenuta ritagliando da A le righe e le colonne dalla k alla n, risulta* $\det A_k > 0$.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad A_1 = A \;, \quad A_2 = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}, \quad A_3 = [a_{33}]$$

  Per $k = n$ otteniamo $\det[a_{nn}] = a_{nn} > 0$. Dal Criterio di Sylvester si deduce che gli elementi sulla diagonale di A sono positivi in questo modo: sia $P_{i,n}$ la matrice di permutazione che scambia la riga i con la riga n, Consideriamo la matrice $P_{i,n}AP_{i,n}$ che è la matrice A con la riga i-esima scambiata con la riga n-esima e con la colonna i-esima scambiata con la colonna n-esima. Risulta ovviamente $P_{i,n}AP_{i,n}$ simmetrica e

$$(P_{i,n}AP_{i,n}\boldsymbol{x}, \boldsymbol{x}) = (AP_{i,n}\boldsymbol{x}, P_{i,n}\boldsymbol{x}) = (A(P_{i,n}\boldsymbol{x}), (P_{i,n}\boldsymbol{x})) \geq 0$$

per la positività di A. Quindi anche $P_{i,n}AP_{i,n}$ è simmetrica e definita positiva e per il criterio di Sylvester l'elemento di posto nn è positivo; ma l'elemento di posto nn della matrice $P_{i,n}AP_{i,n}$ è proprio a_{ii} che risulta quindi positivo. Ragionando in modo analogo, si può dedurre che tutte le sottomatrici di A aventi la diagonale sulla diagonale di A hanno determinante strettamente positivo:



- gli autovalori di A sono reali e positivi, e gli autovettori generano lo spazio $\mathbb{R}^n$.

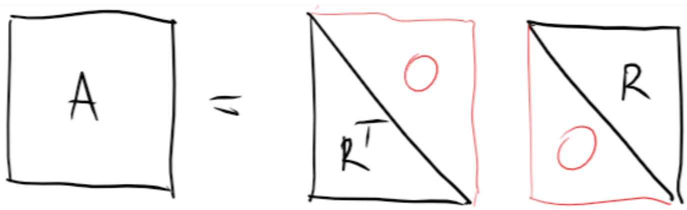
Non è quindi inaspettato che la decomposizione $PA = LU$ nel caso di matrici simmetriche e definite positive diventi in un certo senso più semplice.

5.2. Decomposizione di Cholesky

Teorema 5.1. *(Decomposizione di Cholesky). Sia A una matrice simmetrica e definita positiva. Allora esiste una matrice R triangolare superiore tale che*

$$A = R^T R$$

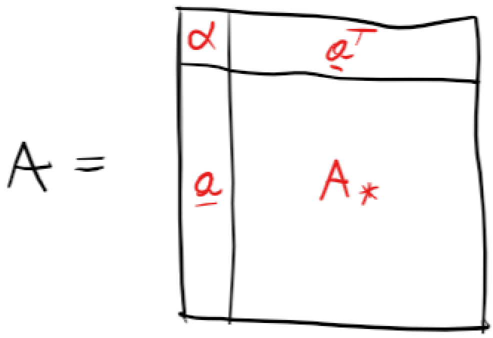
La matrice R è unica se si richiede che gli elementi sulla diagonale siano positivi.



Dimostrazione. La dimostrazione è costruttiva, nel senso che fornisce anche direttamente l'algoritmo per calcolare la matrice R. Prima di procedere facciamo due osservazioni importanti. A differenza della decomposizione $PA = LU$:

- non occorreranno permutazioni per definire la R;
- non occorreranno permutazioni per rendere stabile l'algoritmo.

Sia $\alpha = a_{11}$ $\boldsymbol{a} = [a_{12}, \ldots, a_{1n}]^{\mathrm{T}}$ e A_* la matrice ottenuta prendendo le righe e le colonne di A dalla 2 in poi:



in modo che A si possa decomporre a blocchi come segue:

$$A = \begin{bmatrix} \alpha & \boldsymbol{a}^{\mathrm{T}} \\ \boldsymbol{a} & A_* \end{bmatrix}.$$

Analogamente decomponiamo R:

$$R = \begin{bmatrix} \rho & \boldsymbol{r}^{\mathrm{T}} \\ 0 & R_*^{\mathrm{T}} \end{bmatrix}$$

e scriviamo l'equazione $A = R^{\mathrm{T}} R$:

$$\begin{bmatrix} \alpha & \boldsymbol{a}^{\mathrm{T}} \\ \boldsymbol{a} & A_* \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ \boldsymbol{r} & R_*^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \rho & \boldsymbol{r}^{\mathrm{T}} \\ 0 & R_* \end{bmatrix} = \begin{bmatrix} \rho^2 & \rho \boldsymbol{r}^{\mathrm{T}} \\ \rho \boldsymbol{r} & \boldsymbol{r}\boldsymbol{r}^{\mathrm{T}} + R_*^{\mathrm{T}} R_* \end{bmatrix}.$$

Eguagliamo i blocchi:

$$\alpha = \rho^2, \quad \rho \boldsymbol{r}^{\mathrm{T}} = \boldsymbol{a}^{\mathrm{T}}, \quad \rho \boldsymbol{r} = \boldsymbol{a}, \quad A_* = \boldsymbol{r}\boldsymbol{r}^{\mathrm{T}} + R_*^{\mathrm{T}} R_*.$$

Dato che $a_{11} = \alpha > 0$, possiamo risolvere la prima equazione che ci dà

$$\rho = \sqrt{\alpha}$$

(scegliamo il segno $+$); la seconda e la terza equazione sono una la trasposta dell'altra e possiamo ricavare $\boldsymbol{r}$:

$$\boldsymbol{r} = \frac{1}{\rho} \boldsymbol{a} = \frac{1}{\sqrt{\alpha}} \boldsymbol{a}.$$

La quarta equazione diventa

$$R_*^{\mathrm{T}} R_* = A_* - \frac{1}{\alpha} \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}.$$

Ci siamo ricondotti a dover calcolare una decomposizione di Cholesky di ordine $(n-1)$ della matrice modificata

$$\hat{A}_* = A_* - \frac{1}{\alpha} \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}.$$

Per poter procedere con un altro passo occorre solo controllare che la matrice $\hat{A}_*$ è a sua volta simmetrica e definita positiva.

- $\hat{A}_*$ simmetrica:

$$\hat{A}_*^{\mathrm{T}} = \left(A_* - \frac{1}{\alpha} \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}} \right)^{\mathrm{T}} = A_*^{\mathrm{T}} - \frac{1}{\alpha} (\boldsymbol{a}^{\mathrm{T}})^{\mathrm{T}} \boldsymbol{a}^{\mathrm{T}} = A_* - \frac{1}{\alpha} \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}} = \hat{A}_*.$$

- $\hat{A}_*$ definita positiva: dobbiamo mostrare che per ogni vettore $\boldsymbol{y} \in \mathbb{R}^{n-1}$ diverso da zero si ha

$$0 < \boldsymbol{y}^{\mathrm{T}} \hat{A}_* \boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}} A_* \boldsymbol{y} - \frac{1}{\alpha} \boldsymbol{y}^{\mathrm{T}} (\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}) \boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}} A_* \boldsymbol{y} - \frac{1}{\alpha} (\boldsymbol{y}^{\mathrm{T}} \boldsymbol{a})^2.$$

versione del 8 aprile 2018 ore 23:59

Sia η uno scalare. Dalla definita positività di A abbiamo

$$0 < [\eta \;\; \boldsymbol{y}]^{\mathrm{T}} A \begin{bmatrix} \eta \\ \boldsymbol{y} \end{bmatrix} = [\eta \;\; \boldsymbol{y}^{\mathrm{T}}] \begin{bmatrix} \alpha & \boldsymbol{a}^{\mathrm{T}} \\ \boldsymbol{a} & A_* \end{bmatrix} \begin{bmatrix} \eta \\ \boldsymbol{y} \end{bmatrix} =$$

$$[\eta \;\; \boldsymbol{y}^{\mathrm{T}}] \begin{bmatrix} \alpha\eta + \boldsymbol{a}^{\mathrm{T}}\boldsymbol{y} \\ \eta\boldsymbol{a} + A_*\boldsymbol{y} \end{bmatrix} = \alpha\eta^2 + \eta(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a}) + \eta(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{y}) + \boldsymbol{y}^{\mathrm{T}}A_*\boldsymbol{y} =$$

$$\alpha\eta^2 + 2\eta(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a}) + \boldsymbol{y}^{\mathrm{T}}A_*\boldsymbol{y}.$$

Riassumendo: dobbiamo dimostrare che per ogni $\boldsymbol{y} \in \mathbb{R}^{n-1}$

$$\boldsymbol{y}^{\mathrm{T}}A_*\boldsymbol{y} \underbrace{-\frac{1}{\alpha}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a})^2}_{(*)} > 0$$

Sappiamo che per ogni $\eta \in \mathbb{R}$

$$\underbrace{\alpha\eta^2 + 2\eta(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a})}_{(**)} + \boldsymbol{y}^{\mathrm{T}}A_*\boldsymbol{y} > 0$$

Basta quindi mostrare che possiamo scegliere un particolare η in modo che $(*) = (**)$:

$$-\frac{1}{\alpha}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a})^2 = \alpha\eta^2 + 2\eta(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a}).$$

Moltiplicando per α abbiamo l'equazione

$$0 = \alpha^2\eta^2 + 2\alpha\eta(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a}) + (\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a})^2 = [\alpha\eta + (\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a})]^2$$

che possiamo risolvere ponendo

$$\eta = -\frac{1}{\alpha}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{a}).$$

A questo punto possiamo procedere per tutti i passi fino a ottenere la matrice R. □

La dimostrazione si può agevolmente tradurre in un codice, osservando che al passo k calcoliamo la riga k di R (oppure la colonna k di R^{T}). Esplicitiamo il passo 1:

```
%
% calcolo \rho = \sqrt(\alpha)
%
R(1,1) = sqrt(A((1,1));
%
% calcolo r = 1/sqrt(\alpha)*a
%
for i=2:n
    R(1,i) = A(1,i)/R(1,1);
end
%
% calcolo la matrice A cappello *:
%
for i=2:n
    for j=2:n
        A(i,j) = A(i,j) = -1/A(1,1) * A(i,1)*A(1,j);
    end
end
%
```

e questo è il programma completo:

```
function R=mychol(A)
%
n = size(A,1);
%
R = zeros(n,n);
%
for k=1:n
    %
    R(k,k) = sqrt(A(k,k));
    %
    for i=k+1:n
        R(k,i) = A(k,i)/R(k,k);
    end
    %
    for i=k+1:n
        for j=k+1:n
            A(i,j) = A(i,j) - 1/A(k,k)*A(i,k)*A(k,j);
        end
    end
end
```

5.2.1. Stabilità dell'algoritmo di Cholesky. Si possono dimostrare i seguenti risultati riguardo alla stabilità della decomposizione di Cholesky:

- Dalla decomposizione $A = R^{\mathrm{T}}R$ di Cholesky si può facilmente arrivare alla decomposizione $PA = LU$ descritta nel capitolo procedente con $P = I$ (non c'è bisogno di scambi per portarla a termine);
- la decomposizione di Cholesky però NON corrisponde al pivot parziale, cioè non è sempre vero che il pivot che mi capita è il più grande in modulo della sua colonna;
- tuttavia la decomposizione di Cholesky risulta sempre stabile, ovvero il fattore γ è dell'ordine di 1.

5.3. Applicazione alla soluzione di un sistema lineare

Sia $A\boldsymbol{x} = \boldsymbol{b}$ un sistema lineare la cui matrice dei coefficienti A è una matrice simmetrica e definita positiva e supponiamo di avere calcolato la decomposizione di Cholesky di A:

$$A = R^{\mathrm{T}}R.$$

Dalla decomposizione di A possiamo ricavare la $\boldsymbol{x}$ risolvendo in sequenza due sistemi lineari con matrice triangolare:

(1) dato $\boldsymbol{b}$, trovo $\boldsymbol{y}$ tale che $R^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{b}$ (forward substitution);

(2) dato $\boldsymbol{y}$, trovo $\boldsymbol{x}$ tale che $R\boldsymbol{x} = \boldsymbol{y}$ (backward substitution).

Sostituendo $\boldsymbol{y} = R\boldsymbol{x}$ nell'equazione del punto (1), trovo

$$R^{\mathrm{T}}R\boldsymbol{x} = \boldsymbol{b} \implies A\boldsymbol{x} = \boldsymbol{b}.$$