MA580H Matrix Computations

Lecture 8: Cholesky Factorization

Rafikul Alam
Department of Mathematics
IIT Guwahati

Outline

- Characterization of positive definite matrices
- Cholesky factorization

Quadratic forms

A pure quadratic $f(x, y)$ comes directly from a symmetric 2 by 2 matrix!

$$\mathbf{x}^\top A\mathbf{x} \text{ in } \mathbb{R}^2 \quad ax^2 + 2bxy + cy^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

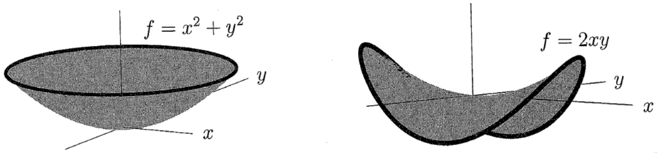

Figure 6.1: A bowl and a saddle: Definite $A = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ and indefinite $A = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$.

For any symmetric matrix A, the product $\mathbf{x}^\top A\mathbf{x}$ is a pure quadratic form $f(x_1, \ldots, x_n)$:

$$\mathbf{x}^\top A\mathbf{x} \text{ in } \mathbb{R}^n \quad \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_n \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Positive definite matrices

A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is said to be

- positive semidefinte if $x^\top Ax \geq 0$ for all $x \in \mathbb{R}^n$ (written as $A \succeq 0$)
- positive definite if $x^\top Ax > 0$ for all nonzero $x \in \mathbb{R}^n$ (written as $A \succ 0$)

A matrix $A \in \mathbb{C}^{n\times n}$ is said to be

- positive semidefinte if $x^* Ax \geq 0$ for all $x \in \mathbb{C}^n$ (written as $A \succeq 0$)
- positive definite if $x^* Ax > 0$ for all nonzero $x \in \mathbb{C}^n$ (written as $A \succ 0$)

A real positive definite matrix is also referred to as a symmetric positive definite (SPD) matrix.

Remark: Let $A \in \mathbb{C}^{n\times n}$. Then $x^*Ax \in \mathbb{R}$ for all $x \in \mathbb{C}^n \iff A = A^*$.
But $A \in \mathbb{R}^{n\times n}$ and $x^\top Ax \in \mathbb{R}$ for all $x \in \mathbb{R}^n \not\Longrightarrow A = A^\top$.

Indeed, if $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ then $x^\top Ax = (x_1 + x_2)^2 \geq 0$ for all $x \in \mathbb{R}^2$ but $A \neq A^\top$.

Positive definite matrices

If $A \in \mathbb{R}^{n\times n}$ is partitioned in the form

$$A = \left[\begin{array}{c|c} A_m & B \\ \hline C & D \end{array}\right], \quad A_m \in \mathbb{R}^{m\times m},$$

then A_m is called a principal submatrix of A. Note that

$$A^\top = A \iff A_m^\top = A_m, \quad C = B^\top, \quad D^\top = D.$$

It follows that if A is SPD then so is A_m. Indeed, for any nonzero $x \in \mathbb{R}^m$, we have

$$x^\top A_m x = \left[\begin{array}{c} x \\ \hline 0 \end{array}\right]^\top \left[\begin{array}{c|c} A_m & B \\ \hline C & D \end{array}\right] \left[\begin{array}{c} x \\ \hline 0 \end{array}\right] > 0.$$

In particular, if A is SPD then $a_{jj} = e_j^\top A e_j > 0$ for $j = 1 : n$. Also, A is nonsingular (why?).

Positive definite matrices

Facts: Let $A \in \mathbb{R}^{n\times n}$ be an SPD matrix. Then the following results hold:

1. If $X \in \mathbb{R}^{n\times p}$ with $\operatorname{rank}(X) = p$ then $X^\top AX$ is SPD. Indeed, for all nonzero $y \in \mathbb{R}^p$,

$$Xy \neq 0 \quad \text{(why?)} \quad \textit{and} \quad y^\top(X^\top AX)y = (Xy)^\top A(Xy) > 0 \implies X^\top AX \text{ is SPD.}$$

2. Leading principal submatrices of A are SPD, that is, $\mathtt{A(1:j,1:j)}$ is SPD for $j = 1:n$.
3. Let $A = \left[\begin{array}{c|c} A_m & B^\top \\ \hline B & D \end{array}\right]$. Then $S := D - BA_m^{-1}B^\top$ is the Schur complement of A_m. Now

$$\left[\begin{array}{c|c} A_m & B^\top \\ \hline B & D \end{array}\right] = \left[\begin{array}{c|c} I & 0 \\ \hline BA_m^{-1} & I \end{array}\right] \left[\begin{array}{c|c} A_m & 0 \\ \hline 0 & D - BA_m^{-1}B^\top \end{array}\right] \left[\begin{array}{c|c} I & 0 \\ \hline BA_m^{-1} & I \end{array}\right]^\top$$

shows that

$$A \text{ is SPD} \iff A_m \text{ and } S := D - BA_m^{-1}B^\top \text{ are SPD.}$$

LDV factorization

Theorem: Suppose that all leading principal submatrices $A \in \mathbb{R}^{n\times n}$ are nonsingular. Then $A = LDV$ is a unique decomposition of A, where L is unit lower triangular, D is diagonal, and V is unit upper triangular.

Proof: By assumption, A has a unique LU factorization $A = LU$. Let $D := \mathrm{diag}(u_{11}, \ldots, u_{nn})$, where $u_{11}, \ldots, u_{nn}$ are diagonal entries of U. Then $V := D^{-1}U$ is unit upper triangular and $A = LDV$. ■

Corollary: If $A \in \mathbb{R}^{n\times n}$ is symmetric and all leading principal submatrices of A are nonsingular then $A = LDL^\top$ is a unique factorization of A, where L is unit lower triangular and D is a diagonal matrix.

Corollary: If A is SPD then $A = LDL^\top$ is a unique factorization of A, where L is unit lower triangular and D is a diagonal SPD matrix.

Cholesky factorization

Theorem: Let $A \in \mathbb{R}^{n\times n}$ be nonsingular. Then A is SPD $\iff$ $A = GG^\top$, where G is a unique lower triangular matrix with positive diagonal entries.

Proof: $A = GG^\top \Rightarrow x^\top Ax = x^\top GG^\top x = (G^\top x)^\top G^\top x = \|G^\top x\|_2 > 0$ for $x \neq 0 \Rightarrow A$ is SPD.

A is SPD $\Rightarrow A = LDL^\top$ is a unique factorization, where L is unit lower triangular and D is diagonal SPD matrix. Let D be given by $D = \mathrm{diag}(d_{11}, \ldots, d_{nn})$. Since $d_{jj} > 0$, define $\sqrt{D} := \mathrm{diag}(\sqrt{d_{11}}, \ldots, \sqrt{d_{nn}})$ and $G := L\sqrt{D}$. Then $A = L\sqrt{D}(L\sqrt{D})^\top = GG^\top$. ■

Definition: If A is SPD then $A = GG^\top$, where G lower triangular with positive diagonals, is called the Cholesky factorization of A and G is called the Cholesky factor of A.

Example:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}^\top .$$

Algorithm (inner product)

Let $A := \begin{bmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{bmatrix}$ and $G := \begin{bmatrix} g_{11} & \\ g_{21} & g_{22} \end{bmatrix}$. Then $A = GG^\top$ yields

$$\begin{bmatrix} a_{11} & a_{21} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} g_{11} & \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} g_{11} & g_{21} \\ & g_{22} \end{bmatrix} = \begin{bmatrix} g_{11}^2 & g_{11}g_{21} \\ g_{11}g_{21} & g_{21}^2 + g_{22}^2 \end{bmatrix}.$$

Equating the columns, we have

$$\begin{aligned} a_{11} &= g_{11}^2 \\ a_{21} &= g_{11}g_{21} \\ a_{22} &= g_{21}^2 + g_{22}^2 \end{aligned} \quad \Longrightarrow \quad \begin{aligned} g_{11} &= \sqrt{a_{11}} \\ g_{21} &= a_{21}/g_{11} \\ g_{22} &= \sqrt{a_{22} - g_{21}^2} \end{aligned}$$

Remark: The factorization is possible if $a_{11} > 0$ and $a_{22} - g_{21}^2 > 0$.

Algorithm (inner product)

More generally, equating columns on both sides of $A = GG^\top$, we have

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix} = g_{11} \begin{bmatrix} g_{11} \\ \vdots \\ g_{n1} \end{bmatrix}, \quad \begin{bmatrix} a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} = g_{21} \begin{bmatrix} g_{21} \\ \vdots \\ g_{n1} \end{bmatrix} + g_{22} \begin{bmatrix} g_{22} \\ \vdots \\ g_{n2} \end{bmatrix}$$

$$\begin{bmatrix} a_{jj} \\ \vdots \\ a_{nj} \end{bmatrix} = g_{j1} \begin{bmatrix} g_{j1} \\ \vdots \\ g_{n1} \end{bmatrix} + g_{j2} \begin{bmatrix} g_{j2} \\ \vdots \\ g_{n2} \end{bmatrix} + \cdots + g_{jj} \begin{bmatrix} g_{jj} \\ \vdots \\ g_{nj} \end{bmatrix}, \; j = 1 : n$$

Algorithm (Inner product):

For $j = 1 : n$

$$g_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} g_{jk}^2}$$

$$g_{ij} = \left(a_{ij} - \sum_{k=1}^{j-1} g_{ik} g_{jk} \right) / g_{jj}, \; i = j+1 : n$$

end

Cost: $n^3/3$ flops - half the cost of GE.

Algorithm (inner product)

Example: Consider $\begin{bmatrix} 16 & -16 & 0 \\ -16 & 41 & -5 \\ 0 & -5 & 5 \end{bmatrix}$. Then

$$\begin{aligned} g_{11} &= \sqrt{a_{11}} = \sqrt{16} = 4, \quad g_{21} = \frac{a_{21}}{g_{11}} = \frac{-16}{4} = -4, \quad g_{31} = \frac{a_{31}}{g_{11}} = \frac{0}{4} = 0 \\ g_{22} &= \sqrt{a_{22} - g_{21}^2} = \sqrt{41 - 16} = 5, \quad g_{32} = \frac{a_{32} - g_{31}g_{21}}{g_{22}} = \frac{-5 - 0 \times (-4)}{5} = -1 \\ g_{33} &= \sqrt{a_{33} - g_{31}^2 - g_{32}^2} = \sqrt{5 - 0 - 1} = 2. \end{aligned}$$

Hence

$$\begin{bmatrix} 16 & -16 & 0 \\ -16 & 41 & -5 \\ 0 & -5 & 5 \end{bmatrix} = \begin{bmatrix} 4 & & \\ -4 & 5 & \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 4 & & \\ -4 & 5 & \\ 0 & -1 & 2 \end{bmatrix}^\top .$$

Algorithm (outer product)

Let $A \in \mathbb{R}^{n\times n}$ be SPD. Then $A = GG^\top$ can be written as

$$\left[\begin{array}{c|c} a_{11} & h^\top \\ \hline h & \widehat{A} \end{array}\right] = \left[\begin{array}{c|c} g_{11} & 0 \\ \hline g & \widehat{G} \end{array}\right] \left[\begin{array}{c|c} g_{11} & g^\top \\ \hline 0 & \widehat{G}^\top \end{array}\right].$$

Equating the blocks, we have

$$\begin{aligned} a_{11} = g_{11}^2 &\implies g_{11} = \sqrt{a_{11}} \\ h = g_{11} g &\implies g = h/g_{11} \\ \widehat{A} = gg^\top + \widehat{G}\widehat{G}^\top &\implies \widehat{A} - gg^\top = \widehat{G}\widehat{G}^\top \end{aligned}$$

```
For k = 1:n
  A(k,k) = sqrt(A(k,k));
  g = A(k+1:n,k)/A(k,k); A(k+1:n,k) = g;
  A(k+1:n, k+1:n) = A(k+1:n, k+1:n)- g*g';
end
```

Cost: $n^3/3$ flops - half the cost of GE.

Example

$$\left[\begin{array}{c|cc} 25 & 15 & -5 \\ \hline 15 & 18 & 0 \\ -5 & 0 & 11 \end{array}\right] = \left[\begin{array}{c|cc} g_{11} & 0 & 0 \\ \hline g_{21} & g_{22} & 0 \\ g_{31} & g_{32} & g_{33} \end{array}\right] \left[\begin{array}{c|cc} g_{11} & g_{21} & g_{31} \\ \hline 0 & g_{22} & g_{32} \\ 0 & 0 & g_{33} \end{array}\right]$$

$$= \left[\begin{array}{c|cc} 5 & 0 & 0 \\ \hline 3 & g_{22} & 0 \\ -1 & g_{32} & g_{33} \end{array}\right] \left[\begin{array}{c|cc} 5 & 3 & -1 \\ \hline 0 & g_{22} & g_{32} \\ 0 & 0 & g_{33} \end{array}\right]$$

Equating (2, 2) blocks, we have

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} g_{22} & 0 \\ g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} g_{22} & g_{32} \\ 0 & g_{33} \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 \\ 1 & g_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & g_{33} \end{bmatrix}$$

Equating (2, 2) entry, we have $10 - 1 = g_{33}^2 \Longrightarrow g_{33} = 3$.

Solving SPD system

Let $A \in \mathbb{R}^{n\times n}$ be SPD and $b \in \mathbb{R}^n$. Then the system $Ax = b$ can be solved using Cholesky factorization as follows.

- Compute Cholesky factorization $A = GG^\top$. Cost: $n^3/3$ flops.
- Solve the lower triangular system $Gy = b$. Cost: n^2 flops.
- Solve the upper triangular system $G^\top x = y$. Cost: n^2 flops.

The MATLAB command `chol` computes Cholesky factorization of a positive definite matrix A. More specifically, the commands

$$\mathtt{R} = \mathtt{chol(A)} \text{ and } \mathtt{L} = \mathtt{chol(A,'lower')}$$

compute an upper triangular matrix R and a lower triangular matrix L such that

$$\mathtt{A} = \mathtt{R}^\top\mathtt{R} \text{ and } \mathtt{A} = \mathtt{L}\mathtt{L}^\top$$

A direct proof of Cholesky factorization

Problem: Let $A = \left[\begin{array}{c|c} a_{11} & h^\top \\ \hline h & D \end{array}\right]$, where $h \in \mathbb{R}^{n-1}$, be SPD. Then the Schur complement $S := D - hh^\top/a_{11}$ is SPD. Now use

$$\begin{aligned}
\left[\begin{array}{c|c} a_{11} & h^\top \\ \hline h & D \end{array}\right] &= \left[\begin{array}{c|c} 1 & 0 \\ \hline h/a_{11} & I_{n-1} \end{array}\right] \left[\begin{array}{c|c} a_{11} & h^\top \\ \hline 0 & D - hh^\top/a_{11} \end{array}\right] \\
&= \left[\begin{array}{c|c} 1 & 0 \\ \hline h/a_{11} & I_{n-1} \end{array}\right] \left[\begin{array}{c|c} a_{11} & 0 \\ \hline 0 & D - hh^\top/a_{11} \end{array}\right] \left[\begin{array}{c|c} 1 & 0 \\ \hline h/a_{11} & I_{n-1} \end{array}\right]^\top \\
&= \left[\begin{array}{c|c} \sqrt{a_{11}} & 0 \\ \hline h/\sqrt{a_{11}} & I_{n-1} \end{array}\right] \left[\begin{array}{c|c} 1 & 0 \\ \hline 0 & D - hh^\top/a_{11} \end{array}\right] \left[\begin{array}{c|c} \sqrt{a_{11}} & 0 \\ \hline h/\sqrt{a_{11}} & I_{n-1} \end{array}\right]^\top
\end{aligned}$$

and induction on n to prove that Cholesky factorization $A = GG^\top$ exists and is unique.

***