Lecture 12. Math 609 600

The Choleski factorization.

We want to write

$$A = LL^t \tag{0.1}$$

where A is an $n \times n$ matrix and L is a lower triangular $n \times n$ matrix with non-negative diagonal entries. Such a factorization is called a Choleski factorization. Let $(\cdot, \cdot)$ denote the Euclidean inner product on $\mathbb{R}^n$. Clearly, for the Choleski factorization to exist, it is necessary that A be symmetric and positive semi-definite since

$$(Ax, x) = (L^t x, L^t x) \geq 0, \quad \text{for all } x \in \mathbb{R}^n$$

and

$$(Ax, y) = (L^t x, L^t y) = (x, Ay), \qquad \text{for all } x, y \in \mathbb{R}^n.$$

A positive semidefinite matrix is one satisfying

$$(Ax, x) \geq 0, \quad \text{for all } x \in \mathbb{R}^n \tag{0.2}$$

while a positive definite matrix is one satisfying (0.2) and

$$(Ax, x) = 0, \quad \text{only if } x = \mathbf{0}. \tag{0.3}$$

If A is nonsingular and has a Choleski factorization, then the diagonal entries of L have to be strickly positive implying that A is symmetirc and positive definite (SPD).

If A is SPD then every principle minor is also SPD. That the principle minors are symmetric follows from the symmetry of A and their positive definiteness follows from

$$(A^k x_k, x_k) = (A\tilde{x}_k, \tilde{x}_k), \quad \text{for all } x_k \in \mathbb{R}^k$$

where $\tilde{x}_k = (x_k, 0, 0, \ldots, 0)^t \in \mathbb{R}^n$ (with $n - k$ zeros). It is also easily follows from Theorem 2.1 of the 609d notes that a SPD matrix is always nonsingular.

We then have the following theorem:

Theorem 12.1. *A non-singular square matrix A has a Choleski factorization if and only if it is SPD. Moreover, the Choleski factorization is unique when it exists.*

Proof. We have already observed that a non-singular matrix A having a Choleski factorization is SPD and in this case, the diagonal entries of L are all positive. Thus, we need only show that a SPD matrix A has a unique Choleski factorization.

The remainder of the proof is along the lines of that of the theorem in the previous lecture, We assume that A is a SPD $n \times n$ matrix. We show, by mathematical induction, that for $j = 1, 2, \ldots, n$, A^j has a unique Choleski factorization $A^j = L^j(L^j)^t$ with the L^j having positive diagonal entries. The theorem follows from the case of $j = n$. The case of $j = 1$ is obvious as A^1 SPD implies that $a_{1,1} > 0$ so that $A^1 = (\sqrt{a_{1,1}})(\sqrt{a_{1,1}})$ is its unique Choleski factorization (taking the positve square root).

We assume $1 < j \leq n$, A^{j-1} has a unique Choleski factorization $A^{j-1} = L^{j-1}(L^{j-1})^t$ with the L^{j-1} having positive diagonal entries. We again consider the block partitioning of A^j, namely,

$$A^j = \left(\begin{array}{c|c} A^{j-1} & (\alpha^j)^t \\ \hline \alpha^j & a_{j,j} \end{array}\right)$$

where $\alpha^j = (a_{j,1}, a_{j,2}, \ldots, a_{j,j-1})$ in this case.

Let $\mathbf{0}$ denote the column vector of length $j - 1$. We look for a factorization of the form

$$A^j = L^j(L^j)^t \tag{0.4}$$

with L^j lower triangular, i.e.,

$$L^j = \left(\begin{array}{c|c} L^{j-1} & \mathbf{0} \\ \hline \ell^j & \ell_{j,j} \end{array}\right)$$

where $\ell^j = (\ell_{j,1}, \ell_{j,2}, \ldots, \ell_{j,j-1})$ is a row vector, L^{j-1} is lower triangular and $\ell_{j,j}$ is a number. Because of the way that block multiplication works, we must have

$$\begin{aligned} L^j(L^j)^t &= \left(\begin{array}{c|c} L^{j-1} & \mathbf{0} \\ \hline \ell^j & \ell_{j,j} \end{array}\right) \left(\begin{array}{c|c} (L^{j-1})^t & (\ell^j)^t \\ \hline \mathbf{0}^t & \ell_{j,j} \end{array}\right) \\ &= \left(\begin{array}{c|c} L^{j-1}(L^{j-1})^t & L^{j-1}(\ell^j)^t \\ \hline \ell^j(L^{j-1})^t & \ell^j(\ell^j)^t + \ell_{j,j}^2 \end{array}\right) \end{aligned} \tag{0.5}$$

For (0.4) to hold, we must have:

$$L^{j-1}(L^{j-1})^t = A^{j-1}, \tag{0.6}$$

$$L^{j-1}(\ell^j)^t = (\alpha^j)^t, \tag{0.7}$$

$$\ell^j(\ell^j)^t + \ell_{j,j}^2 = a_{j,j}. \tag{0.8}$$

For L^j to have positive diagonal, so must L^{j-1}, i.e., (0.6) is the unique Choleski factorization of A^{j-1}. (0.7) uniquely defines ℓ^j and the theorem will follow if we can show that

$$\gamma := a_{j,j} - \ell^j(\ell^j)^t > 0 \tag{0.9}$$

as $\ell_{j,j}$ is then defined to be the positive square root of γ. Notice, that (0.6) and (0.7) implies that

$$A^j = \left(\begin{array}{c|c} L^{j-1}(L^{j-1})^t & L^{j-1}(\ell^j)^t \\ \hline \ell^j(L^{j-1})^t & a_{j,j} \end{array}\right) \tag{0.10}$$

To show (0.9), we start by defining

$$\widetilde{L}^j = \left(\begin{array}{c|c} L^{j-1} & \mathbf{0} \\ \hline \ell^j & 0 \end{array}\right).$$

As

$$(\widetilde{L}^j)^t = \left(\begin{array}{c|c} (L^{j-1})^t & (\ell^j)^t \\ \hline \mathbf{0}^t & 0 \end{array}\right).$$

is singular, there is a nonzero vector $y \in \mathbb{R}^j$ satisfying $(\widetilde{L}^j)^t y = 0$. As $(L^{j-1})^t$ is nonsingular, $y_j \neq 0$ so we may assume without loss of generality that $y_j = 1$. Finally,

$$A^j = \widetilde{L}^j(\widetilde{L}^j)^t + \gamma \left(\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & 1 \end{array}\right) \tag{0.11}$$

with γ given by (0.9). This can be seen by expanding out the product involving $\widetilde{L}_j$ in (0.11) and comparing it with (0.10). Finally,

$$0 < (A^j y, y) = (\widetilde{L}^j(\widetilde{L}^j)^t y, y) + \gamma = \gamma, \tag{0.12}$$

i.e., (0.9) holds completes the proof of the theorem. $\square$

Applications LU or Choleski factorizations are particularly useful the systems coming from parabolic initial value problem, especially those with time independent coefficients. Such a problem is touched upon in Lecture 27 of the 639 notes. Implementations with a fixed time step require many solutions with a fixed matrix A_τ which depends on the time step size τ. Thus, the matrix is initially factored and the factors are back and forward solved at each step.

Another important application of the Choleski factorization involves the generalized eigenvalue problem. The generalized eigenvalue problem is to find the generalized eigenpairs (ψ_i, λ_i) with $\psi \neq 0$ satisfying

$$A\psi_i = \lambda_i B\psi_i$$

where A and B are $n \times n$ matrices and B is SPD. Using the Choleski factorization of $B = LL^t$, the generalized eigenvalue problem can be reduced to the standard eigenvalue problem. Setting $y_i = L^t\psi_i$, we find that (y_i, λ_i) satisfies the standard eigenvalue problem

$$\widetilde{A}y_i = \lambda_i y_i \qquad \text{for } \widetilde{A} = L^{-1}A(L^t)^{-1}.$$

Once the standard eigenvectors $\{y_i\}$ are computed, the generalized eigevalues are given by $\psi_i = (L^t)^{-1}y_i$ and share the same eigenvalue λ_i.