Lecture 3: Cholesky and QR

Michael S. Floater

September 19, 2018

These notes are based on Sections 2.4–2.5 of the book.

1 Cholesky factorization

Let $A = [A_{i,j}]$ be an $n \times n$ symmetric matrix, i.e., such that $A_{i,j} = A_{j,i}$ for $i, j = 1, \ldots, n$. A Cholesky factorization of A is a factorization $A = LDL^T$, where L is lower triangular with ones on its diagonal and D is a diagonal matrix. Expressing L using its columns this means that

$$A = \begin{bmatrix} \mathbf{l}_1 & \cdots & \mathbf{l}_n \end{bmatrix} \begin{bmatrix} D_{1,1} & 0 & \cdots & 0 \\ 0 & D_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & D_{n,n} \end{bmatrix} \begin{bmatrix} \mathbf{l}_1^T \\ \vdots \\ \mathbf{l}_n^T \end{bmatrix} = \sum_{k=1}^{n} D_{k,k} \mathbf{l}_k \mathbf{l}_k^T.$$

This requires roughly half the storage required by an LU factorization. The algorithm is similar to that of LU:

- Initialize $A_0 = A$.
- For $k = 1, \ldots, n$, let $\mathbf{l}_k$ be the k-th column of A_{k-1} scaled so that $L_{k,k} = 1$. Let $D_{k,k} = (A_{k-1})_{k,k}$ and calculate $A_k = A_{k-1} - D_{k,k} \mathbf{l}_k \mathbf{l}_k^T$.

As we saw with LU, this algorithm is not always valid since it assumes that $(A_{k-1})_{k,k} \neq 0$ for each k.

As an example, we find the Cholesky factorization of

$$A = \begin{bmatrix} 3 & 4 \\ 4 & 6 \end{bmatrix}.$$

We initialize $A_0 = A$. In the first step we use the first column of A_0:

$$\mathbf{l}_1 = \frac{1}{3}\begin{bmatrix}3\\4\end{bmatrix} = \begin{bmatrix}1\\4/3\end{bmatrix},$$

and the first diagonal element of A_0:

$$D_{1,1} = 3.$$

Then we find

$$\mathbf{l}_1\mathbf{l}_1^T = \begin{bmatrix}1 & 4/3\\4/3 & 16/9\end{bmatrix},$$

and so

$$D_{1,1}\mathbf{l}_1\mathbf{l}_1^T = \begin{bmatrix}3 & 4\\4 & 16/3\end{bmatrix},$$

and we let

$$A_1 = A_0 - D_{1,1}\mathbf{l}_1\mathbf{l}_1^T = \begin{bmatrix}0 & 0\\0 & 2/3\end{bmatrix}.$$

In the second step we use the second column of A_1:

$$\mathbf{l}_2 = \frac{1}{2/3}\begin{bmatrix}0\\2/3\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix},$$

and the second diagonal element of A_1:

$$D_{2,2} = 2/3.$$

This completes the factorization:

$$\begin{bmatrix}3 & 4\\4 & 6\end{bmatrix} = \begin{bmatrix}\mathbf{l}_1 & \mathbf{l}_2\end{bmatrix}\begin{bmatrix}D_{1,1} & 0\\0 & D_{2,2}\end{bmatrix}\begin{bmatrix}\mathbf{l}_1^T\\\mathbf{l}_2^T\end{bmatrix} = \begin{bmatrix}1 & 0\\4/3 & 1\end{bmatrix}\begin{bmatrix}3 & 0\\0 & 2/3\end{bmatrix}\begin{bmatrix}1 & 4/3\\0 & 1\end{bmatrix}.$$

Two more examples are

$$\begin{bmatrix}-3 & 4\\4 & 6\end{bmatrix} = \begin{bmatrix}1 & 0\\-4/3 & 1\end{bmatrix}\begin{bmatrix}-3 & 0\\0 & 34/3\end{bmatrix}\begin{bmatrix}1 & -4/3\\0 & 1\end{bmatrix},$$

and

$$\begin{bmatrix}3 & 2\\2 & 0\end{bmatrix} = \begin{bmatrix}1 & 0\\2/3 & 1\end{bmatrix}\begin{bmatrix}3 & 0\\0 & -4/3\end{bmatrix}\begin{bmatrix}1 & 2/3\\0 & 1\end{bmatrix}.$$

Recall that A is positive definite if $\mathbf{x}^T A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

Theorem 1 *Let A be a real $n \times n$ symmetric matrix. It is positive definite if and only if an LDL^T factorization exists where the diagonal elements of D are all positive.*

Proof. Suppose first that $A = LDL^T$ with $D_{i,i} > 0$ for $i = 1, \dots, n$ and let $\mathbf{x} \neq 0$. Since L has ones on the diagonal, it is nonsingular and so $\mathbf{y} = L^T\mathbf{x}$ is nonzero. Hence

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \sum_{i=1}^{n} D_{i,i} y_i^2 > 0,$$

and A is positive definite.

Conversely, suppose A is positive definite. Let $\mathbf{e}_k \in \mathbb{R}^n$ denote the k-th unit vector. Firstly, $A_{1,1} = \mathbf{e}_1^T A\mathbf{e}_1 > 0$ and so $\mathbf{l}_1$ and $D_{1,1}$ are well-defined with $D_{1,1} > 0$. To complete the proof we will show that $(A_{k-1})_{k,k} > 0$ for all $k = 1, \dots, n$ and hence that $\mathbf{l}_k$ and $D_{k,k}$ are well-defined with $D_{k,k} > 0$. We proceed by induction on k, and assume that

$$A_{k-1} = A - \sum_{i=1}^{k-1} D_{i,i}\mathbf{l}_i\mathbf{l}_i^T$$

has been computed successfully, with $D_{1,1}, \dots, D_{k-1,k-1} > 0$. Let $\mathbf{x} \in \mathbb{R}^n$ be such that $x_{k+1} = x_{k+2} = \cdots = x_n = 0$, $x_k = 1$, and $x_j = -\sum_{i=j+1}^{k} L_{i,j}x_i$ for $j = k-1, k-2, \dots, 1$. With this choice of $\mathbf{x}$,

$$\mathbf{l}_j^T\mathbf{x} = \sum_{i=1}^{n} L_{i,j}x_i = \sum_{i=j}^{k} L_{i,j}x_i = x_j + \sum_{i=j+1}^{k} L_{i,j}x_i = 0, \qquad j = 1, \dots, k-1.$$

Since the first $k-1$ rows and columns of A_{k-1} and the last $n-k$ components of $\mathbf{x}$ vanish and $x_k = 1$, we have

$$\begin{aligned}(A_{k-1})_{k,k} &= \mathbf{x}^T A_{k-1}\mathbf{x} = \mathbf{x}^T \left(A - \sum_{i=1}^{k-1} D_{i,i}\mathbf{l}_i\mathbf{l}_i^T \right) \mathbf{x} \\ &= \mathbf{x}^T A\mathbf{x} - \sum_{i=1}^{k-1} D_{j,j}(\mathbf{l}_j^T\mathbf{x})^2 = \mathbf{x}^T A\mathbf{x} > 0.\end{aligned}$$

□

The conclusion from the theorem is that we can check whether a symmetric matrix is positive definite by attempting to calculate its LDL^T factorization. Moreover, if A is indeed positive definite, we can can define $D^{1/2}$ as the diagonal matrix where $(D^{1/2})_{k,k} = \sqrt{D_{k,k}}$. Then $D^{1/2}D^{1/2} = D$ and so we obtain the alternative Cholesky factorization

$$A = RR^T,$$

where

$$R = LD^{1/2}.$$

2 QR factorization

Now we consider a different way to factorize a matrix which applies both to square matrices and to rectangle matrices where the number of rows is greater than the number of columns.

The *scalar product* of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle = \sum_{i=1}^{n} x_i y_i = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}.$$

The scalar product is a linear function of both $\mathbf{x}$ and $\mathbf{y}$: for any $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in $\mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$,

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle.$$

The *norm* or *Euclidean length* of $\mathbf{x} \in \mathbb{R}^n$ is

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} \geq 0.$$

The norm of $\mathbf{x}$ is zero if and only if $\mathbf{x}$ is the zero vector.

Two vectors $\mathbf{x}, \mathbf{y}$ are *orthogonal to each other* if

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0.$$

A set of vectors $\mathbf{q}_1, \ldots, \mathbf{q}_m \in \mathbb{R}^n$ is called *orthonormal* if

$$\langle \mathbf{q}_k, \mathbf{q}_l \rangle = \begin{cases} 1, & k = l, \\ 0, & k \neq l, \end{cases} \qquad k, l = 1, \ldots, m.$$

Let $Q = [\mathbf{q}_1 \cdots \mathbf{q}_n]$ be an $n \times n$ real matrix. It is called *orthogonal* if its columns are orthonormal. It follows from $(Q^T Q)_{k,l} = \langle \mathbf{q}_k, \mathbf{q}_l \rangle$ that $Q^T Q = I$ where I is the identity matrix. Thus Q is nonsingular and the inverse exists, $Q^{-1} = Q^T$. Furthermore,

$$QQ^T = Q^T Q = I.$$

Therefore the rows of an orthogonal matrix are also orthonormal and Q^T is also an orthogonal matrix. Further,

$$1 = \det I = \det(QQ^T) = \det Q \det(Q^T) = (\det Q)^2$$

and we deduce

$$\det Q = \pm 1.$$

Lemma 1 *If P, Q are orthogonal, then so is PQ.*

Proof. Since $P^T P = Q^T Q = I$, we have

$$(PQ)^T (PQ) = (Q^T P^T)(PQ) = Q^T (P^T P) Q = Q^T Q = I.$$

□

Definition 1 *(QR factorization) Let A be a real $n \times m$ matrix with m linearly independent columns, $n \geq m$. The QR factorization of A has the form $A = QR$, where Q is an $n \times n$ orthogonal matrix and R is an upper triangular $n \times m$ matrix (i.e., $R_{i,j} = 0$ for $i > j$).*

Denoting the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_m$ and the columns of Q by $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^n$, we can write the factorization as

$$\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} R_{1,1} & R_{1,2} & \cdots & R_{1,m} \\ 0 & R_{2,2} & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & R_{m,m} \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}.$$

Then

$$\mathbf{a}_k = \sum_{i=1}^{k} R_{i,k}\mathbf{q}_i, \qquad k = 1, \dots, m, \tag{1}$$

so the k-th column of A is a linear combination of the first k columns of Q.

The QR factorization can be used to solve $A\mathbf{x} = \mathbf{b}$ when A is $n \times n$. We first solve $Q\mathbf{y} = \mathbf{b}$ by letting $\mathbf{y} = Q^T\mathbf{b}$. We then find $\mathbf{x}$ that solves $R\mathbf{x} = \mathbf{y}$ by back substitution.

3 Exercises

Exercise 2.6 *Calculate the Cholesky factorization of the matrix*

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 1 & \lambda \end{bmatrix}.$$

Deduce from the factorization the value of λ which makes the matrix singular.

Exercise 2.7 *Let $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ be the columns of the matrix*

$$\begin{bmatrix} 3 & 6 & -1 \\ -6 & -6 & 1 \\ 2 & 1 & -1 \end{bmatrix}.$$

Using the Gram-Schmidt procedure, generate orthonormal vectors $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$ such that $\mathbf{a}_k = \sum_{i=1}^{k} R_{i,k}\mathbf{q}_i$, $k = 1, 2, 3$. Thus express A as the product $A = QR$ where Q is orthogonal and R is upper triangular.