The Cholesky Factorization

What if, instead of factoring A into two matrices L and U, we could factor it into the product of one triangular matrix? That is, when does $A = R^T R$ for some upper-triangular matrix R? Such a factorization is called a *Cholesky* factorization.

Definition 1. An $n \times n$ matrix A is *symmetric* if $A^T = A$.

Question 2. Suppose $A = R^T R$. Is A symmetric?

Yes, and here's the proof.

Proof. Suppose there exists a matrix R where $A = R^T R$. Then $A^T = (R^T R)^T$. By the properties of matrix transpose, $A^T = (R^T)(R^T)^T = R^T R = A$. Thus A is symmetric. $\square$

So in order to have a Cholesky factorization, it is *necessary* that A be symmetric. This is not sufficient; here's an example that we will later see does not have one:

$$\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}.$$

There's a second property of matrices that, combined with symmetry, will guarantee the existence of a Cholesky factorization.

Definition 3. An $n \times n$ matrix A is *positive definite* if $\mathbf{x}^T A\mathbf{x} > 0$ for all vectors $\mathbf{x} \neq 0$.

Example 1. Let $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$, $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{x}_2 = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$. Compute $\mathbf{x}^T A\mathbf{x}$ for each $\mathbf{x}$. Is this enough information to determine if A is or is not positive definite?

$$\mathbf{x}_1^T A\mathbf{x}_1 = \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 10 \end{bmatrix} = 18 + 10 = 28$$

$$\mathbf{x}_2^T A\mathbf{x}_2 = \begin{bmatrix} -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \end{bmatrix} \right) = \begin{bmatrix} -3 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ -8 \end{bmatrix} = -8$$

This is enough information to determine that A is NOT positive definite, because we've found an $\mathbf{x}$ where $\mathbf{x}^T A\mathbf{x} \leq 0$. However, testing vectors will never be enough to show that a matrix is positive definite.

Example 2. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$. Find a formula for $\mathbf{x}^T A\mathbf{x}$.

$$\mathbf{x}^T A\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \left(\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1 + 3x_2 \\ 3x_1 + x_2 \end{bmatrix} = x_1^2 + 3x_1x_2 + 3x_1x_2 + x_2^2 = x_1^2 + 6x_1x_2 + x_2^2.$$

To determine positive definiteness, we keep going until we have the sum or difference of squares only, because we know that if something is squared then it's nonnegative for all input values. This requires completing the square.

$x_1^2+6x_1x_2+x_2^2 = x_1^2+x_1(6x_2)+x_2^2 = x_1^2+6x_1x_2+9x_2^2-9x_2^2+x_2^2 = (x_1+3x_2)^2-8x_2^2$

At this point, we have a negative sign on a term so that there should exist example vectors where this is negative. Then it's a matter of finding one; $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -4 \\ 1 \end{bmatrix}$ work.

Essentially, the process of showing a matrix is positive definite is exactly the same. If you can write the expression entirely as the (positive) sum of squares, the result is guaranteed to be positive. If you can't, then start experimenting with some numbers that make the expression negative. If you can find even one pair, then the claim that A is positive definite is false.

Example 3. Show that each matrix is positive definite. Are they also symmetric positive definite?

Example 4. Show that $A_1 = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$, $A_2 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 5 & 4 \\ 0 & 4 & 8 \end{bmatrix}$, and $A_3 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 5 & 6 \\ 0 & 2 & 8 \end{bmatrix}$ are positive definite. Which ones are symmetric positive definite?

$\mathbf{x}^T A_1 \mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2x_1+2x_2 \\ 2x_1+5x_2 \end{bmatrix} = 2x_1^2+2x_1x_2+2x_1x_2+5x_2^2 = 2x_1^2+4x_1x_2+5x_2^2 = 2(x_1^2+2x_1x_2+x_2^2-x_2^2)+5x_2^2 = 2(x_1+x_2)^2+3x_2^2 > 0$ for all $\mathbf{x} \neq 0$. So A_1 is positive definite, and since A_1 is symmetric, A_1 is symmetric positive definite.

$\mathbf{x}^T A_2 \mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 5 & 4 \\ 0 & 4 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} x_1-x_2 \\ -x_1+5x_2+4x_3 \\ 4x_2+8x_3 \end{bmatrix} = x_1^2-x_1x_2-x_1x_2+5x_2^2+4x_2x_3+4x_2x_3+8x_3^2 = x_1^2-2x_1x_2+5x_2^2+8x_2x_3+8x_3^2 = (x_1^2-2x_1x_2+x_2^2)-x_2^2+5x_2^2+8x_2x_3+8x_3^2 = (x_1-x_2)^2+4(x_2^2+2x_2x_3+x_3^2)-4x_3^2+8x_3^2 = (x_1-x_2)^2+4(x_2+x_3)^2+4x_3^2$. Since all terms are positive squares, this is greater than zero for all $\mathbf{x} \neq 0$. Since A_2 is symmetric, A_2 is symmetric positive definite.

Finally, $\mathbf{x}^T A_3 \mathbf{x}$ also equals $x_1^2-2x_1x_2+5x_2^2+8x_2x_3+8x_3^2 = (x_1-x_2)^2+4(x_2+x_3)^2+4x_3^2$, so A_3 is positive definite. However, A_3 is not symmetric.

Some theory regarding positive definite matrices.

Theorem 4. *All symmetric positive definite matrices are invertible.*

This theorem also means that $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b}$, and that A has linearly independent columns, 0 is not an eigenvalue, etc. etc.

Theorem 5. *If the $n \times n$ matrix A is symmetric, then A is positive definite if and only if all of its eigenvalues are positive.*

We haven't touched on eigenvalues yet, other than mentioning that they also can only be found iteratively. That means that while this theorem is nice theoretically, it's not particularly helpful if a computer is trying to determine if A is symmetric positive definite.

Definition 6. A *principal submatrix* of a square matrix A is a square submatrix whose diagonal entries are diagonal entries of A.

Example 5. List the principal submatrices of $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 5 & 4 \\ 0 & 4 & 8 \end{bmatrix}$.

$$\begin{bmatrix} 1 \end{bmatrix}, \begin{bmatrix} 5 \end{bmatrix}, \begin{bmatrix} 8 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ -1 & 5 \end{bmatrix}, \begin{bmatrix} 5 & 4 \\ 4 & 8 \end{bmatrix}, A$$

Theorem 7. *If A is symmetric positive definite, then so are all its principle submatrices.*

Theorem 8. *If A is a square symmetric positive definite matrix, then there exists an upper triangular matrix such that $A = R^T R$.*

Cholesky Factorization

We compute the factorization by going along the diagonal. Let A be represented as

$$\left[\begin{array}{c|c} a & \mathbf{b}^T \\ \hline \mathbf{b} & C \end{array}\right]$$

1. First, update the diagonal element a. The square root of a goes into R. **

$$R = \left[\begin{array}{c|c} \sqrt{a} & \mathbf{b}^T \\ \hline \mathbf{0} & C \end{array}\right]$$

2. Next we update the vector $\mathbf{b}$ by also scaling it with $\sqrt{a}$. Let $\mathbf{u} = \frac{1}{\sqrt{a}}\mathbf{b}$. Then we have

$$R = \left[\begin{array}{c|c} \sqrt{a} & \mathbf{u}^T \\ \hline \mathbf{0} & C \end{array}\right]$$

3. Finally we update the submatrix C. The new submatrix D is:

$$D = C - \mathbf{u}\mathbf{u}^T$$

so that R is now

$$R = \left[\begin{array}{c|c} \sqrt{a} & \mathbf{u}^T \\ \hline \mathbf{0} & D \end{array}\right]$$

.

4. Repeat steps 1-3 on the submatrix D until no submatrix remains.

** If at any point the diagonal element a is negative, then the matrix is not symmetric positive definite. This is actually the most efficient way for a computer to check if a matrix is symmetric positive definite - just start computing a Cholesky factorization. If it works, great! The matrix is positive definite. If at any point the first step of the factorization fails, then A is not symmetric positive definite.

Example 6. Find a Cholesky factor of $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 5 & 4 \\ 0 & 4 & 8 \end{bmatrix}$.

First, $a = 1$ so $\sqrt{1} = 1$ and $\mathbf{u} = \mathbf{b}$. Then $\mathbf{u}\mathbf{u}^T = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ so $C - \mathbf{u}\mathbf{u}^T = \begin{bmatrix} 5 & 4 \\ 4 & 8 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 8 \end{bmatrix}$. Repeat the process on this 2×2 matrix.

Now we square root 4, which is 2. We scale the remaining row element, 4, by 2, and get 2. This leaves a submatrix of $[8]$, which we update as $8 - (2)(2) = 4$. Square root this number to get the final element.

So the Cholesky factor R should be

$$R = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix}$$

Let's check: $R^T R = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 5 & 4 \\ 0 & 4 & 8 \end{bmatrix} = A$. Hooray!

Example 7. Find a Cholesky factor of $A = \begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}$.

First, $a = 2$ so $R = \begin{bmatrix} \sqrt{2} & \\ 0 & \end{bmatrix}$. Then $u = \frac{1}{\sqrt{2}}(2) = \sqrt{2}$ and $D = 5 - (\sqrt{2})^2 = 3$. Thus $R = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \end{bmatrix}$. For the final element, don't forget to square root it, so that $R = \begin{bmatrix} \sqrt{2} & \sqrt{2} \\ 0 & \sqrt{3} \end{bmatrix}$

Example 8. Find a Cholesky factor of $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$.

First, $a = 1$ so $\sqrt{1} = 1$ and $\mathbf{u} = \mathbf{b}$. Then $\mathbf{u}\mathbf{u}^T = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ so $C - \mathbf{u}\mathbf{u}^T = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. Repeat the process on this 2×2 submatrix.

Now we square root 1, which is 1. We scale the remaining row element, 1, by 1, and get 1. This leaves a submatrix of $[2]$, which we update as $2 - (1)(1) = 1$. Finally, square root this result to get 1.

So the Cholesky factor R should be

$$R = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Question 9. Compare and contrast an $R^T R$ factorization verses an LU factorization.

Both could be used to solve $A\mathbf{x} = \mathbf{b}$ through back substitution. One advantage is saving storage space - we only have to store one matrix instead of two. A Cholesky factorization is also about half as much work as finding an LU factorization, at about $\frac{n^3}{3}$ operations.