Methods for Computing the Cholesky Factorization

Keith M. Chugg ©
November 19, 1995

1 Problem Statement

The causal simulation and whitening problems require that we factor an $(n \times n)$ positive definite matrix, $\mathbf{K}$, as

$$\mathbf{K} = \mathbf{H}\mathbf{H}^\dagger, \tag{1}$$

where $\mathbf{H}$ represents a causal transformation (i.e. $\mathbf{H}$ is an $(n \times n)$ lower triangular matrix). This factorization is known as the Cholesky Factorization and is guaranteed to exist for $\mathbf{K}$ as described above.

This handout discusses three methods for performing the Cholesky factorization. The *"direct method"* is the most straightforward, but is only useful for small matrices. The *"Eigen method"* is useful for theoretical justification of the existence of the Cholesky factorization. The $\mathbf{LDL}^\dagger$ *method* is the preferred method for programing the procedure and for calculation by hand for larger matrices.

We describe each method in the following sections.

2 The Direct Method

This is the method presented in the supplemental notes by Scholtz. The direct method consists of simply equating the two sides of (1) and then solving for the elements of $\mathbf{H}$. For example, if $n = 3$ (1) expands out to

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{12}^* & k_{22} & k_{23} \\ k_{13}^* & k_{23}^* & k_{33} \end{bmatrix} = \begin{bmatrix} h_{11} & 0 & 0 \\ h_{21} & h_{22} & 0 \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} h_{11}^* & h_{21}^* & h_{31}^* \\ 0 & h_{22}^* & h_{32}^* \\ 0 & 0 & h_{33}^* \end{bmatrix}. \tag{2}$$

Expanding this we get

$$|h_{11}|^2 = k_{11} \tag{3a}$$
$$h_{11}h_{21}^* = k_{12} \tag{3b}$$
$$h_{11}h_{31}^* = k_{13} \tag{3c}$$
$$|h_{21}|^2 + |h_{22}|^2 = k_{22} \tag{3d}$$
$$h_{21}h_{31}^* + h_{22}h_{32}^* = k_{23} \tag{3e}$$
$$|h_{31}|^2 + |h_{32}|^2 + |h_{33}|^2 = k_{33}. \tag{3f}$$

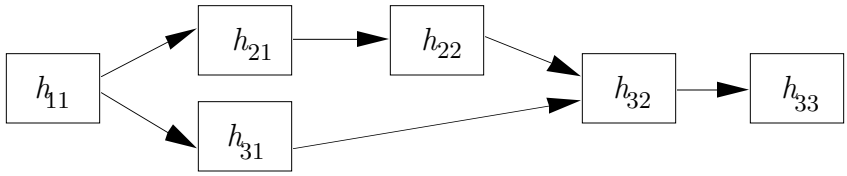


Figure 1: The order for solving the $n = 3$ case with the direct method.

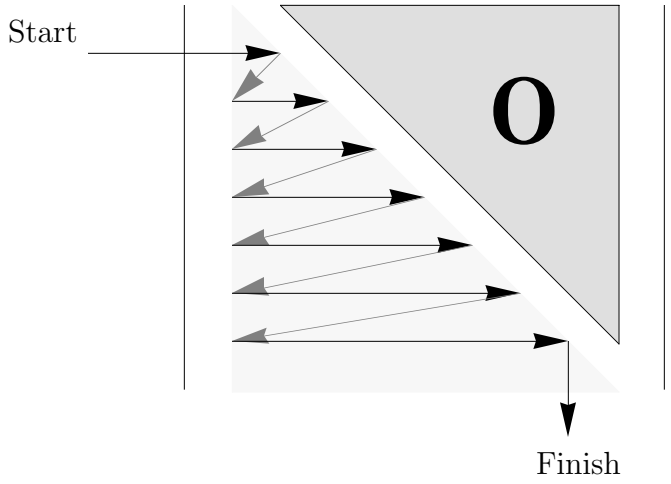


Figure 2: The order for solving for the elements of **H** using the direct method.

The system of equations in (3) should be solved in the order shown in the flow diagram of Figure 1. In general you should solve for the elements of **H** in the order shown in Figure 2. Solving these equations is simpler if **K** is real, in which case **H** can be assumed real. Even for real matrices, the direct method is limited to only small n - imagine solving this for a (10×10) matrix; just obtaining the equivalent of (3) would be quite a task!

It seems that we should try to automate the procedure - the $\mathbf{LDL}^\dagger$ method does so.

3 The $\mathbf{LDL}^\dagger$ Method

The equations in (3) imply that we can always choose **H** so that $h_{ii} > 0$ for $i = 1, 2 \ldots n$ (see problem 1). This means that **H** can be written in the following form

$$\mathbf{H} = \mathbf{L}\mathbf{D}^{1/2}, \tag{4}$$

where $\mathbf{L}$ is lower triangular with 1's down the diagonal (we will refer to this as "unit lower triangular") and $\mathbf{D}$ is a (real) diagonal matrix with positive entries down the diagonal. For $n = 3$ these matrices have the form

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}, \tag{5}$$

where $d_i > 0$ for $i = 1, 2, 3$. In this case the Cholesky factorization in (1) becomes

$$\mathbf{K} = \mathbf{L}\mathbf{D}\mathbf{L}^\dagger, \tag{6}$$

hence the name of this method.

The idea behind this method is to multiply $\mathbf{K}$ by $\mathbf{L}^{-1}$ on the left so that we are left with

$$\mathbf{L}^{-1}\mathbf{K} = \mathbf{D}\mathbf{L}^\dagger. \tag{7}$$

Once we know $\mathbf{D}\mathbf{L}^\dagger$, it is simply to obtain $\mathbf{H}^\dagger = \mathbf{D}^{1/2}\mathbf{L}^\dagger$.

Of course, we don't know what $\mathbf{L}^{-1}$ is; so how do we multiply by it? The answer is the result of two facts:

1. The inverse of $\mathbf{L}$ has the same form as $\mathbf{L}$; it is unit lower triangular (see problem 2).
2. Multiplication on the left by a unit lower triangular matrix correspond to replacing the i^{th} row by itself plus some linear combinations of rows above it.

The second fact is illustrated by denoting $\mathbf{L}^{-1}$ by $\mathbf{M}$, so that by the first fact

$$\mathbf{L}^{-1} = \mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix}. \tag{8}$$

If we let the $\mathbf{k}_i$ be the i^{th} column of $\mathbf{K}$, then for $n = 3$

$$\mathbf{M}\mathbf{K} = \begin{bmatrix} 1 & 0 & 0 \\ m_{21} & 1 & 0 \\ m_{31} & m_{32} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{k}_1^\dagger \\ \mathbf{k}_2^\dagger \\ \mathbf{k}_3^\dagger \end{bmatrix} \tag{9a}$$

$$= \begin{bmatrix} \mathbf{k}_1^\dagger \\ m_{21}\mathbf{k}_1^\dagger + \mathbf{k}_2^\dagger \\ m_{31}\mathbf{k}_1^\dagger + m_{32}\mathbf{k}_2^\dagger + \mathbf{k}_3^\dagger \end{bmatrix}. \tag{9b}$$

So we can multiply by $\mathbf{L}^{-1}$ without explicitly knowing it by performing "unit Gaussian elimination" (i.e. replacing the i^{th} row by itself plus some linear combinations of rows above it). We can "zero" all of the subdiagonal elements of each column in one step; working from left to right as illustrated in Figure 3

The steps of the $\mathbf{L}\mathbf{D}\mathbf{L}^\dagger$ method are:

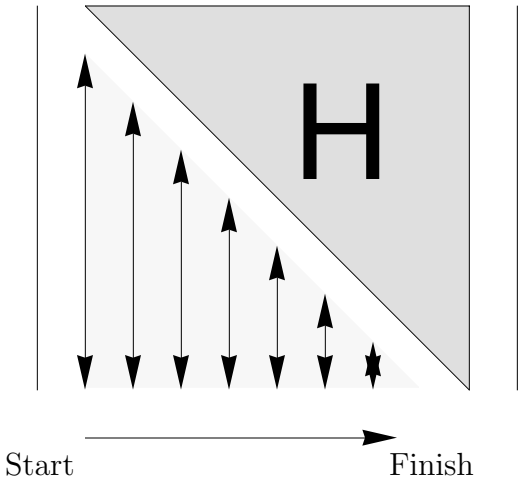


Figure 3: The subdiagonal entries of $\mathbf{K}$ are zeroed one column at a time, from left to right, in the $\mathbf{LDL}^\dagger$ method.

1. Perform unit Gaussian eliminations on $\mathbf{K}$ to zero all of the elements below the diagonal. The result is $\mathbf{DL}^\dagger$.
2. Obtain $\mathbf{H}^\dagger = \mathbf{D}^{1/2}\mathbf{L}^\dagger$ by dividing each i^{th} row of $\mathbf{DL}^\dagger$ by $\sqrt{d_i}$ for $i = 1, 2 \ldots n$.
3. Take the Hermitian transpose to get $\mathbf{H}$.
4. Check your result by multiplying $\mathbf{HH}^\dagger$.

The second step can be easily seen for the case of $n = 3$

$$\mathbf{DL}^\dagger = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} 1 & l_{21}^* & l_{31}^* \\ 0 & 1 & l_{32}^* \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} d_1 \begin{pmatrix} 1 & l_{21}^* & l_{31}^* \end{pmatrix} \\ d_2 \begin{pmatrix} 0 & 1 & l_{32}^* \end{pmatrix} \\ d_3 \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \end{bmatrix}, \tag{10}$$

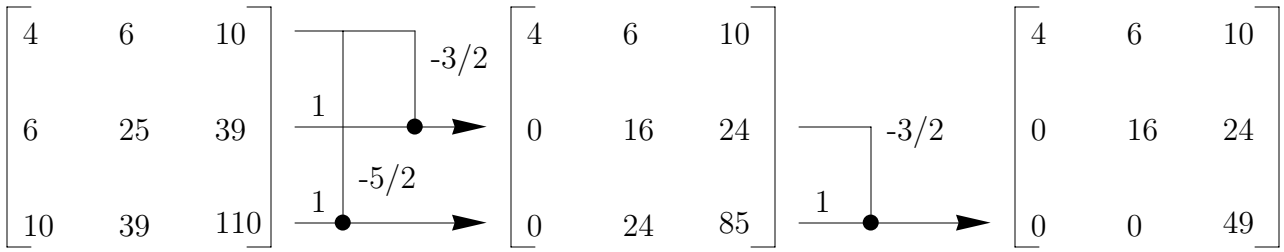
so dividing each row by $\sqrt{d_i}$ is equivalent to factoring out a $\mathbf{D}^{1/2}$ on the left.

3.1 $\mathbf{LDL}^\dagger$ Example

The $\mathbf{LDL}^\dagger$ method is much easier to understand by looking at a specific example. Let's consider a real-valued example

$$\mathbf{K} = \begin{bmatrix} 4 & 6 & 10 \\ 6 & 25 & 39 \\ 10 & 39 & 110 \end{bmatrix}. \tag{11}$$

The steps of the unit Gaussian elimination are



So

$$\mathbf{D}\mathbf{L}^{\mathrm{t}} = \begin{bmatrix} 4 & 6 & 10 \\ 0 & 16 & 24 \\ 0 & 0 & 49 \end{bmatrix}. \tag{12}$$

Factoring out $\mathbf{D}^{1/2}$ on the left (i.e. dividing each row by the square root of the diagonal entry) implies

$$\mathbf{H}^{\mathrm{t}} = \mathbf{D}^{1/2}\mathbf{L}^{\mathrm{t}} = \begin{bmatrix} 2 & 3 & 5 \\ 0 & 4 & 6 \\ 0 & 0 & 7 \end{bmatrix}, \tag{13}$$

which can easily be verified by multiplying $\mathbf{H}\mathbf{H}^{\mathrm{t}}$.

3.2 Programming the $\mathbf{LDL}^{\dagger}$ Factorization

The $\mathbf{LDL}^{\dagger}$ method is easily programmed on a computer - in fact many math packages will perform it for you. Here's the pseudo code for the $\mathbf{LDL}^{\mathrm{t}}$ (real-valued case)

- for$(j = 1, 2, \ldots n)$
  - for$(m = 1, 2, \ldots j-1)$ $r_m := l_{jm} d_m$
  - $d_j := k_{jj} - \sum_{m=1}^{j-1} l_{jm} r_m$
  - for$(i = j+1, \ldots n)$ $l_{ij} := (k_{ij} - \sum_{m=1}^{j-1} l_{im} r_m)/d_j$

Here r_m is just a temporary array introduced to save computation.

4 The Eigen Method

Since $\mathbf{K}$ is assumed to be positive definite symmetric, we know that the (non-causal) factorization given by

$$\mathbf{K} = (\mathbf{E}\mathbf{\Lambda}^{1/2})(\mathbf{E}\mathbf{\Lambda}^{1/2})^{\dagger}, \tag{14}$$

exist, where $\mathbf{E}$ and $\mathbf{L}$ are invertible matrices of eigenvectors and eigenvalues respectively. For any unitary matrix $\mathbf{U}$, (i.e. $\mathbf{U}\mathbf{U}^\dagger = \mathbf{I}$) we have

$$\mathbf{K} = (\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U})(\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U})^\dagger. \tag{15}$$

The idea behind the Eigen method is to design $\mathbf{U}$ so that $\mathbf{H} = (\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U})$ is lower triangular.

The method consists of multiplying $\mathbf{E}\mathbf{\Lambda}^{1/2}$ on the left by a sequence of unitary matrices, each of which "zeros" an element above the diagonal. This uses the fact that the product of unitary matrices is a unitary matrix (see problem 4).

One way to zero elements is to use Given's Rotaions - matrices which zero one element at a time. Consider the (2×2) case

$$\mathbf{E}\mathbf{\Lambda}^{1/2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{16}$$

where we want to zero the 1,2 element, b. To do so we multiply on the right by

$$\mathbf{U}_{1,2} = \frac{1}{\sqrt{|a|^2 + |b|^2}} \begin{bmatrix} a^* & -b \\ b^* & a* \end{bmatrix}. \tag{17}$$

It is straightforward to check that $\mathbf{U}_{1,2}$ is unitary and that

$$\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U}_{1,2} = \begin{bmatrix} \star & 0 \\ \star & \star \end{bmatrix}, \tag{18}$$

where $\star$ represents a non-zero (in general) entry. The important thing is that the 1,2 element has been zeroed.

For larger matrices, we must zero more than one element. This is done by designing $\mathbf{U}_{i,j}$ to zero the i, j element for $j > i$ and for $i = 1, 2 \ldots n$. This is most easily seen by considering $n = 3$. Let's use a, b, c and d as place-holders for the non-zero elements that we use in the rotation (i.e. a, c, d play the same role as in the above example and b is the element that we wish to zero). At each step, the actual values which a, b, c and d represent change, but their role remains the same. Also for given values of a, b, c and d, define the following

$$\alpha = \frac{a^*}{\sqrt{|a|^2 + |b|^2}} \qquad \beta = \frac{b^*}{\sqrt{|a|^2 + |b|^2}}. \tag{19}$$

Initially, all of the elements of $\mathbf{E}\mathbf{\Lambda}^{1/2}$ are non-zero. We zero the 1,2 element first

$$\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U}_{1,2} = \begin{bmatrix} a & b & \star \\ c & d & \star \\ \star & \star & \star \end{bmatrix} \begin{bmatrix} \alpha & -\beta & 0 \\ \beta & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \star & 0 & \star \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix}. \tag{20}$$

The next element to be zeroed is the 1,3 element

$$\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U}_{1,2}\mathbf{U}_{1,3} = \begin{bmatrix} a & 0 & b \\ \star & \star & \star \\ c & \star & d \end{bmatrix} \begin{bmatrix} \alpha & 0 & -\beta \\ 0 & 1 & 0 \\ \beta & 0 & \alpha \end{bmatrix} = \begin{bmatrix} \star & 0 & 0 \\ \star & \star & \star \\ \star & \star & \star \end{bmatrix}. \tag{21}$$

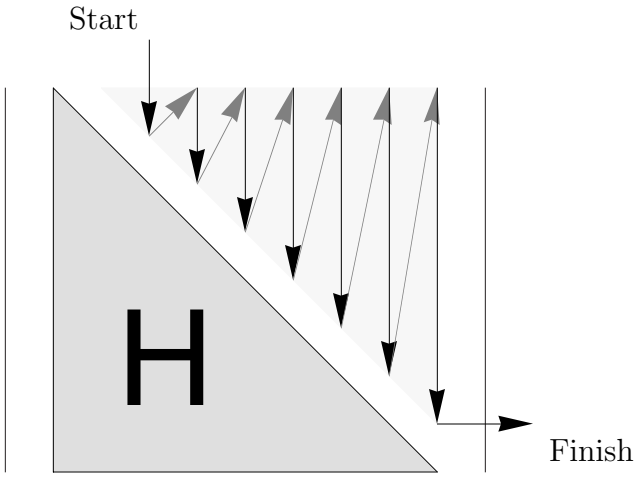


Figure 4: The order for zeroing the above-diagonal elements of the $\mathbf{E}\mathbf{\Lambda}^{1/2}$ matrix using rotation matrices.

The last element to zero is the 2,3 element

$$\mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U}_{1,2}\mathbf{U}_{1,3}\mathbf{U}_{2,3} = \begin{bmatrix} \star & 0 & 0 \\ \star & a & b \\ \star & c & d \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & -\beta \\ 0 & \beta & \alpha \end{bmatrix} = \begin{bmatrix} \star & 0 & 0 \\ \star & \star & 0 \\ \star & \star & \star \end{bmatrix}. \tag{22}$$

So we take $\mathbf{U} = \mathbf{U}_{1,2}\mathbf{U}_{1,3}\mathbf{U}_{2,3}$ and $\mathbf{H} = \mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U}$.

For an n dimensional $\mathbf{K}$ matrix, we choose $\mathbf{U}$ as

$$\mathbf{U} = (\mathbf{U}_{1,2})\,(\mathbf{U}_{1,3}\mathbf{U}_{2,3})\,(\mathbf{U}_{1,4}\mathbf{U}_{2,4}\mathbf{U}_{3,4})\cdots(\mathbf{U}_{1,n}\mathbf{U}_{2,n}\cdots\mathbf{U}_{n-1,n})\,, \tag{23}$$

and take $\mathbf{H} = \mathbf{E}\mathbf{\Lambda}^{1/2}\mathbf{U}$. The order in which we zero elements is important - if the elements are zeroed in another order, we may destroy a zero that we created! The order for performing the zeroing is shown in Figure 4.

This method may be useful if we already have the eigenvectors and eigenvalues of $\mathbf{K}$, but if they are unknown, the additional work to find them makes the Eigen method undesirable.

5 Remarks

Throughout this handout we have assumed that $\mathbf{K}$ was positive definite. If $\mathbf{K}$ is singular, then the methods described may not provide the Cholesky Factor, $\mathbf{H}$. The manifestation of this in the direct method is that when solving for the elements of $\mathbf{H}$, the choices are non-unique. If you choose a value arbitrarily and continue on, it usually works.

In the $\mathbf{LDL}^\dagger$ method, a singular $\mathbf{K}$ means that one of the diagonal elements zero. If this occurs at the last step, then the last row of $\mathbf{DL}^\dagger$ is all zeros. This is not a problem; just divide all of rows by the square-root of the diagonal elements an leave the last row alone (in this case the last column of $\mathbf{H}$ is all zeros). Worse yet, one of the "pivots" (i.e. the diagonal element that you divide by to compute the Gaussian eliminations) is zero. In this case, you have to consider using row permutations. Similar problems arise with the Eigen method.

An approach which can be used when $\mathbf{K}$ is singular is dimension reduction. This is developed in problem 6.

For more on the Cholesky factorization and its numerical implementation, see "Linear Algebra and Its Applications," by G. Strang and "Matrix Computations," by G. Golub and C. Van Loan.

6 Problems

1. Why can we choose h_{ii} real and positive in (3)?

2. Show that when $\mathbf{L}$ is unit lower triangular so is $\mathbf{L}^{-1}$.

3. In the whitening problem we are interested in finding $\mathbf{H}^{-1}$. Describe a method to find $\mathbf{H}^{-1}$ when performing the $\mathbf{LDL}^\dagger$ decomposition. Demonstrate your method on the $\mathbf{K}$ matrix used for the example in Section 3.1.

4. Show that $\mathbf{U} = \mathbf{U}_1\mathbf{U}_2\cdots\mathbf{U}_k$ is unitary when $\mathbf{U}_i$ for $i = 1, 2 \ldots k$ are each unitary.

5. Show that $\mathbf{U}_{1,2}$ is unitary for the two dimensional case. Generalize to n-dimensions and $\mathbf{U}_{i,j}$ for $j > i$ and $i = 1, 2 \ldots n$.

6. This problem deals with what to do when $\mathbf{K}$ is singular. Consider the mean zero (real) random vector

$$\mathbf{x}(u) = \begin{bmatrix} x(u,1) \\ x(u,2) \\ x(u,3) \end{bmatrix},$$

with covariance matrix

$$\mathbf{K}_\mathbf{x} = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

   (a) Show that $x(u,1)+x(u,2)+x(u,3) \overset{\text{as}}{=} 0$. Therefore $x(u,3) \overset{\text{as}}{=} -x(u,1)-x(u,2)$.

   (b) Define the random vector $\mathbf{y}(u)$ by

$$\mathbf{y}(u) = \begin{bmatrix} x(u,1) \\ x(u,2) \end{bmatrix}. \tag{24}$$

   What is $\mathbf{m}_\mathbf{y}$ and $\mathbf{K}_\mathbf{y}$?

(c) Perform the Cholesky decomposition on $\mathbf{K_y}$.

(d) Describe how you would causally simulate $\mathbf{x}(u)$ using $w(u,1)$ and $w(u,2)$, two uncorrelated, mean zero, unit variance random variables.

(e) Use your results from the previous part to write down the Cholesky factorization of $\mathbf{K_x}$.

(f) Describe how you would causally whiten $\mathbf{x}(u)$.

(g) Would the method developed in this problem work if we eliminated $x(u,2)$ instead of $x(u,3)$ in part (6a)?

(h) Try finding the Cholesky factorization of $\mathbf{K_x}$ directly using the $\mathbf{LDL}^\dagger$ method. How is the result related to the solution of part (6d)?