Scientific Computing
Chapter III
Numerical Linear Algebra II, Factorization algorithms

Jonathan Goodman
Courant Institute of Mathemaical Sciences

Last revised February 25, 2002

1 Introduction

As we say earlier, many algorithms of numerical linear algebra may be formulated as ways to calculate matrix factorizations. This point of view gives conceptual insight. Since computing the factorization is usually much more expensive than using it, storing the factors makes it possible, for example, to solve many systems of equations, $Ax = b$, with the same the same A but different b (and therefore different x), faster than if we had started over each time. Finally, when we seek high performance, we might take advantage of alternative ways to organize computations of the factors.

This chapter does not cover the many factorization algorithms in great detail. This material is available, for example, in the book of Golub and van Loan and many other places. My aim is to make the reader aware of what the computer does (roughly), and how long it should take. First I explain how the classical Gaussian elimination algorithm may be viewed as a matrix factorization, the LU factorization. The algorithm presented is not the practical one because it does not include "pivoting". Next, I discuss the Choleski (LL^*) decomposition, which is a natural version of LU for symmetric positive definite matrices. Understanding the details of the Choleski decomposition will be useful later when we study optimization methods and still later when we discuss sampling multivariate normal random variables with correlations. Finally, we show how to compute matrix factorizations, such as the QR decomposition, that lead to orthogonal matrices.

2 Gauss elimination and the LU decomposition

Gauss elimination is a simple systematic way to solve systems of linear equations. For example, suppose we have the system of equations

$$\begin{array}{ll} e_1 : & 2x + y + z = 4 \ , \\ e_2 : & x + 2y + z = 3 \ , \\ e_3 : & x + y + 2z = 4 \ . \end{array}$$

To find the values of x, y, and z, we first try to write equations that contain fewer variables, and eventualy just one. We can "eliminate" x from the equation e_2 by subtracting $\frac{1}{2}$ of both sides of e_1 from e_2.

$$e_2' : \qquad x + 2y + z - \frac{1}{2}(2x + y + z) = 3 - \frac{1}{2} \cdot 4 \ ,$$

Which involves just y and z:

$$e_2' : \qquad \frac{3}{2}y + \frac{1}{2}z = 2 \ .$$

We can do the same to eliminate x from e_3, subtracting $\frac{1}{2}$ of each side of e_1 from the corresponding side of e_3:

$$e_3' : \qquad x + y + 2z - \frac{1}{2}(2x + y + z) = 4 - \frac{1}{2} \cdot 4 \ ,$$

which gives

$$e_3' : \qquad \frac{1}{2}y + \frac{3}{2}z = 2 \ .$$

We now have a pair of equations, e_2', and e_3' that involve only y and z. We can use e_2' to eliminate y from e_3; we subtract $\frac{1}{3}$ of each side of e_2' from the corresponding side of e_3' to get:

$$e_3'' : \qquad \frac{1}{2}y + \frac{3}{2}z - \frac{1}{3}(\frac{3}{2}y + \frac{1}{2}z) = 2 - \frac{1}{3} \cdot 2 \ ,$$

which simplifies to:

$$e_3'' : \qquad \frac{4}{3}z = \frac{5}{3} \ .$$

This completes the elimination phase. In the "back substitution" phase we successively find the values of z, y, and x. First, from e_3'' we immediately find

$$z = \frac{5}{4} \ .$$

Then we use e_2' (not e_3') to get y:

$$\frac{3}{2}y + \frac{1}{2} \cdot \frac{5}{4} = 1 \implies y = \frac{1}{4} \ .$$

Lastly, e_1 yields x:

$$2x + \frac{1}{4} + \frac{5}{4} = 4 \implies x = \frac{9}{4} .$$

The reader can (and should) check that $x = \frac{9}{4}$, $y = \frac{1}{4}$, and $z = \frac{5}{4}$ satisfies the original equations e_1, e_2, and e_3.

The above steps may be formulated in matrix terms. The equations, e_1, e_2, and e_3, may be assembled into a single equation involving a matrix and two vectors:

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 4 \end{pmatrix} .$$

The operation of eliminating x from the second equation may be carried out by multiplying this equation from the left on both sides by the "elementary" matrix

$$E_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

The result is

$$\begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 3 \\ 4 \end{pmatrix} .$$

Doing the matrix multiplication gives

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 1 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 4 \end{pmatrix} .$$

Note that the middle row of the matrix contains the coefficients from e_2'. Similarly, the effect of eliminating x from e_3 comes from multiplying both sides from the left by the elementary matrix

$$E_{31} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} .$$

Which gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 1 & 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 1 \\ 4 \end{pmatrix} ,$$

which multiplies out to become

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix} .$$

This is the matrix form of three equations e_1, e_2', and e_3'. The last elimination step removes y from the last equation using

$$E_{32} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{pmatrix} \; .$$

This gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -\frac{1}{3} & 1 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix} \; ,$$

which multiplies out to be

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & \frac{4}{3} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ \frac{5}{3} \end{pmatrix} \; . \tag{1}$$

Because the matrix in (1) is upper triangular, we may solve for z, then y, then x, as before. The matrix equation (1) is equivalent to the system e_1, e_2', and e_3''.

We can summarize this sequence of multiplications with elementary matrices by saying that we multiplied the original matrix,

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

first by E_{21}, then by E_{31}, then by E_{32} to get the upper triangular matrix

$$U = \begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & \frac{4}{3} \end{pmatrix} \; .$$

This may be written formally as

$$E_{32}E_{31}E_{21}A = U \; .$$

We turn this into a factorization of A by multiplying successively by the inverses of the elementary matrices:

$$A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U \; .$$

It is easy to check that we get the inverse of an elementary matrix, E_{jk} simply by reversing the sign of the number below the diagonal. For example,

$$E_{31}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix}$$

since

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

Also, the product of the elementary matrices just has the nonzero subdiagonal elements of all of them in their respective positions (check this):

$$\begin{aligned} L &= E_{21}^{-1} E_{31}^{-1} E_{32}^{-1} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{3} & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{pmatrix} \end{aligned}$$

Finally, the reader should verify that we actually have $A = LU$:

$$\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & \frac{4}{3} \end{pmatrix} .$$

Now we know that performing Gauss elimination on the three equations e_1, e_2, and e_3 is equivalent to finding an LU factorization of A where the lower triangular factor has ones on its diagonal.

Finally, we can turn this process around and seek the elements of L and U directly from the structure of L and U. In terms of the (supposedly unknown) entries of L and U, the matrix factorization becomes

$$\begin{pmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{pmatrix} \cdot \begin{pmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} . \tag{2}$$

We may find the entries l_{jk} and u_{jk} one by one by multiplying out the product on the left and comparing to the known element on the right. For the $(1,1)$ element, we get

$$1 \cdot u_{11} = 2 ,$$

Which gives $u_{11} = 2$, as we already know. With this, we may calculate either l_{21} from matching the $(2,1)$ entries, or u_{12} from the $(1,2)$ entries. The former gives

$$l_{21} \cdot u_{11} = 1 ,$$

which, given $u_{11} = 2$, gives $l_{21} = \frac{1}{2}$. The latter gives

$$1 \cdot u_{12} = 1 .$$

and then $u_{12} = 1$. These calculations show that the LU factorization, if it exists, is unique (remembering to put ones on the diagonal of L). They also show that there is some freedom in the order in which we compute the l_{jk} and u_{jk}.

We may compute the LU factors of A without knowing the right hand side

$$\begin{pmatrix} 4 \\ 3 \\ 4 \end{pmatrix} .$$

If we know L and U and then learn the right hand side, we may find x, y, and z in a two stage process that begins with "forward" substitution and concludes with the familiar back substitution. In matrix terms, we have

$$L \cdot U \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 4 \end{pmatrix} .$$

We first find an intermediate vector, x', y', and z' by solving

$$L \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 4 \end{pmatrix} , \tag{3}$$

and then solving

$$U \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} . \tag{4}$$

Multiplying (4) by L, and using (3) and $LU = A$ gives

$$A \cdot \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \\ 4 \end{pmatrix} .$$

Since the equations (3) are lower triangular, we may find the elements x', y', and z' one by one. First, we get $1 \cdot x\prime = 4$ (because $l_{11} = 1$), so $x\prime = 4$ Next we get $l_{21}x' + 1 \cdot y' = 3$. Since $l_{21} = \frac{1}{2}$, this gives $y' = 1$. In a similar way, we get $z' = \frac{5}{3}$. With this information, and the known values of the u_{jk}, (4) becomes

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & \frac{4}{3} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ \frac{5}{3} \end{pmatrix} ,$$

which is the same system of equations we arrived at by applying elementary matrices to A and the right hand side above. Recall that the elements of L contain all the numbers we used in Gauss elimination. We have just seen that the forward elimination step, which uses L, is equivalent to applying the same elementary operations to the right hand side. Our LU approach to solving linear systems of equations is entirely equivalent to the Gauss elimination approach.

The general LU algorithm for solving linear systems should be clear from this example. We have an $n \times n$ matrix A, and a column vector $b \in R^n$, and we wish to find another column vector x so that $Ax = b$. This is equivalent to n

linear equations in the n unknowns $x_1, \ldots, x_n$. We first compute the LU decomposition of A, in one of several related ways. Then we solve a "lower triangular" system of equations $Ly = b$ using forward elimination. The intermediate vector entries, $y_1, \ldots, y_n$, are what we would have gotten had we applied Gauss elimination to the right hand side and A at the same time. Finally, we perform back substition, finding x with $Ux = y$. Multiplying this by L and using $LU = a$ and $Ly = b$, we see that this x solves our problem.

Some of the motivation for the factorization approach come from the work involved. Solving a system of equations requires on the order of n^3 operations. This is also the work required to compute the LU factors. The forward and back substitutions require only $O(n^2)$ operations, which is less than $O(n^3)$ by a factor of n. If we have many systems of equations to solve, we should factor A once and use the factors repeatedly.

The elimination and factorization algorithms just described may fail or be numerically unstable even when A is well conditioned. To get a stable algorithm, we need to introduce "pivoting". In the present context[1] this means adaptively reordering the equations or the unknowns so that the elements of L do not grow. Details are in the references.

3 Choleski factorization

A real matrix, A, is symmetric if $A^* = A$. A real matrix is positive definite if $x^*Ax > 0$ for any column vector $x \in R^n$ with $x \neq 0$. We write SPD for "real positive definite". It is possible for a matrix to be positive definite without being symmetric, but it is rare that we are interested in that fact. If we are asking whether a matrix is positive definite, we probably already know the matrix is symmetric. For example, a function of n variables is strictly convex at a point if it's Hessian matrix (the matrix of all second partial derivatives) is positive definite. The Hessian is symmetric no matter what because "mixed partial derivatives commute": $\partial^2 f/\partial x_j \partial x_k = \partial^2 f/\partial x_k \partial x_j$. Another example would be a statistical estimate of a covariance matrix for several random variables. It is very unlikely that we would ever see a nonsymmetric matrix, no matter how silly the estimator. If A has complex entries, the anologue of symmetric is Hermitian, also written $A^* = A$. The definition of positive definiteness is also the same, but now $x \in C^n$ rather than R^n. We focus on real SPD matrices here, as they are more common in computational applications. It might be hardto tell whether a given A is SPD, but there are some checks to rule it out. For example, since $e_k^* A e_k = a_{kk}$, the diagonal entries in an SPD matrix must be positive. Also, a diagonal $k \times k$ block of any size of an SPD matrix must be SPD because you can choose "trial vectors", x with only these k entries not zero.

The Choleski, or LL^* factorization is a special version of the LU factorization adapted to SPD matrices. A theorem of linear algebra (see a reference) states that A is SPD if and only if $A = LL^*$ for some nonsingular lower triangular matrix, L. In view of this theorem, a computational algorithm for computing L

[1] The term "pivot" means something different, for example, in linear programming.

can be used as a test of positive definiteness. Run the algorithm. If it produces L with no zeros on the diagonal, your original A was positive definite. If it shows there is no such L, then A was not positive definite. We do not prove the theorem here, but one part is easy: if $A = LL^*$ then $x^*Ax = x^*LL^*x = (L^*x)^* \cdot (L^*x) = \|L^*x\|_{l^2}^2$. The length of L^*x cannot be negative and can be zero only if $x = 0$ or L is singular.

As with the LU factorization, we can find the entries of L from the equations for the entries of $LL^* = A$ one at a time, in a certain order.

$$\begin{pmatrix} l_{11} & 0 & 0 & \cdots & 0 \\ l_{21} & l_{22} & 0 & \cdots & \vdots \\ l_{31} & l_{32} & l_{33} & \ddots & \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \cdots & & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} l_{11} & l_{21} & l_{31} & \cdots & l_{n1} \\ 0 & l_{22} & l_{32} & \cdots & l_{n2} \\ 0 & 0 & l_{33} & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & & l_{nn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{n1} \\ a_{21} & a_{22} & a_{32} & \cdots & a_{n2} \\ a_{31} & a_{32} & a_{33} & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \\ a_{n1} & a_{n2} & \cdots & & a_{nn} \end{pmatrix}$$

Notice that we have written, for example, a_{32} for the $(2,3)$ entry because A is symmetric. We start with the top left corner. Doing the matrix multiplication gives

$$l_{11}^2 = a_{11} \implies l_{11} = \sqrt{a_{11}} \; .$$

The square root is real because $a_{11} > 0$ because A is positive definite. Next we match the $(2,1)$ entry of A. The matrix multiplication gives:

$$l_{21}l_{11} = a_{21} \implies l_{21} = \frac{1}{l_{11}} a_{21} \quad .$$

The denominator is not zero because $l_{11} > 0$ because $a_{11} > 0$. We could continue in this way, to get the whole first column of L. Alternatively, we could match $(2,2)$ entries to get l_{22}:

$$l_{21}^2 + l_{22}^2 = a_{22} \implies l_{22} = \sqrt{a_{22} - l_{21}^2} \quad .$$

It is possible to show (see references) that if the square root on the right is not real, then A was not positive definite. Given l_{22}, we can now compute the rest of the second column of L. For example, matching $(3,2)$ entries gives:

$$l_{31} \cdot l_{21} + l_{32} \cdot l_{22} = a_{32} \implies l_{32} = \frac{1}{l_{22}} \left(a_{32} - l_{31} \cdot l_{21} \right) \quad .$$

Continuing in this way, we can find all the entries of L. It is clear that if L exists and if we always use the positive square root, then all the entries of L are uniquely determined.

Once we have the Choleski decomposition of A, we can solve systems of equations $Ax = b$ using forward and back substitution, as we did for the LU factorization.