STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2019
LECTURE 11

1. Orthogonalization using Givens rotations

- we illustrate the process in the case where A is a 2×2 matrix
- in Gaussian elimination, we compute $L^{-1}A = U$ where L^{-1} is unit lower triangular and U is upper triangular, specifically,
$$\begin{bmatrix} 1 & 0 \\ m_{21} & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} \\ 0 & a_{22}^{(2)} \end{bmatrix}, \quad m_{21} = -\frac{a_{21}}{a_{11}}$$
- by contrast, the QR decomposition takes the form
$$\begin{bmatrix} \gamma & \sigma \\ -\sigma & \gamma \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{bmatrix}$$
where $\gamma^2 + \sigma^2 = 1$
- from the relationship $-\sigma a_{11} + \gamma a_{21} = 0$ we obtain
$$\begin{aligned} \gamma a_{21} &= \sigma a_{11} \\ \gamma^2 a_{21}^2 &= \sigma^2 a_{11}^2 = (1-\gamma^2) a_{11}^2 \end{aligned}$$
which yields
$$\gamma = \pm \frac{a_{11}}{\sqrt{a_{21}^2 + a_{11}^2}}$$
- it is conventional to choose the + sign
- then, we obtain
$$\sigma^2 = 1 - \gamma^2 = 1 - \frac{a_{11}^2}{a_{21}^2 + a_{11}^2} = \frac{a_{21}^2}{a_{21}^2 + a_{11}^2},$$
or
$$\sigma = \pm \frac{a_{21}}{\sqrt{a_{21}^2 + a_{11}^2}}$$
- again, we choose the + sign
- as a result, we have
$$r_{11} = a_{11} \frac{a_{11}}{\sqrt{a_{21}^2 + a_{11}^2}} + a_{21} \frac{a_{21}}{\sqrt{a_{21}^2 + a_{11}^2}} = \sqrt{a_{21}^2 + a_{11}^2}$$
- the matrix
$$Q^{\mathsf{T}} = \begin{bmatrix} \gamma & \sigma \\ -\sigma & \gamma \end{bmatrix}$$
is called a *Givens rotation*
- it is called a rotation because it is orthogonal, and therefore length-preserving, and also because there is an angle θ such that $\sin\theta = \sigma$ and $\cos\theta = \gamma$, and its effect is to rotate a vector through the angle θ

Date: November 4, 2019, version 1.0. Comments, bug reports: `lekheng@galton.uchicago.edu`.

- in particular,

$$\begin{bmatrix} \gamma & \sigma \\ -\sigma & \gamma \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \rho \\ 0 \end{bmatrix}$$

  where $\rho = \sqrt{\alpha^2 + \beta^2}$, $\alpha = \rho \cos\theta$ and $\beta = \rho \sin\theta$
- it is easy to verify that the product of two rotations is itself a rotation
- now, in the case where A is an $n \times n$ matrix, suppose that we are given the vector

$$\begin{bmatrix} \times \\ \vdots \\ \times \\ \alpha \\ \times \\ \vdots \\ \times \\ \beta \\ \times \\ \vdots \\ \times \end{bmatrix} \in \mathbb{R}^n,$$

  then

$$\begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & & & \\ & & & \gamma & & & & \sigma & & & \\ & & & & 1 & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & 1 & & & & \\ & & & -\sigma & & & & \gamma & & & \\ & & & & & & & & 1 & & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & 1 \end{bmatrix} \begin{bmatrix} \times \\ \vdots \\ \times \\ \alpha \\ \times \\ \vdots \\ \times \\ \beta \\ \times \\ \vdots \\ \times \end{bmatrix} = \begin{bmatrix} \times \\ \vdots \\ \times \\ \rho \\ \times \\ \vdots \\ \times \\ 0 \\ \times \\ \vdots \\ \times \end{bmatrix}$$

- so, in order to transform A into an upper triangular matrix R, we can find a product of rotations Q such that $Q^{\mathsf{T}} A = R$
- it is easy to see that $O(n^2)$ rotations are required

2. Givens rotations versus Householder reflections

- we showed how to construct Givens rotations in order to rotate two elements of a column vector so that one element would be zero, and that approximately $n^2/2$ such rotations could be used to transform A into an upper triangular matrix R
- because each rotation only modifies two rows of A, it is possible to interchange the order of rotations that affect different rows, and thus apply sets of rotations in parallel
- this is the main reason why Givens rotations can be preferable to Householder reflections
- other reasons are that they are easy to use when the QR factorization needs to be updated as a result of adding a row to A or deleting a column of A
- Givens rotations are also more efficient when A is sparse

3. COMPUTING THE COMPLETE ORTHOGONAL FACTORIZATION

- we first seek a decomposition of the form $A = QR\Pi$ where the permutation matrix Π is chosen so that the diagonal elements of R are maximized at each stage
- specifically, suppose

$$H_1 A = \begin{bmatrix} r_{11} & \times & \cdots & \times \\ 0 & \times & \cdots & \times \\ \vdots & \vdots & & \vdots \\ 0 & \times & \cdots & \times \end{bmatrix}, \quad r_{11} = \|\mathbf{a}_1\|_2$$

- so, we choose Π_1 so that $\|\mathbf{a}_1\|_2 \geq \|\mathbf{a}_j\|_2$ for $j \geq 2$
- for Π_2, look at the lengths of the columns of the submatrix; we don't need to recompute the lengths each time, because we can update by subtracting the square of the first component from the square of the total length
- eventually, we get

$$Q \begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix} \Pi_1 \cdots \Pi_r = A$$

  where R is upper triangular
- suppose

$$A = Q \begin{bmatrix} R & S \\ 0 & 0 \end{bmatrix} \Pi$$

  where R is upper triangular, then

$$A^{\mathsf{T}} = \Pi^{\mathsf{T}} \begin{bmatrix} R^{\mathsf{T}} & 0 \\ S^{\mathsf{T}} & 0 \end{bmatrix} Q^{\mathsf{T}}$$

  where R^{T} is lower triangular
- we apply Householder reflections so that

$$H_k \cdots H_2 H_1 \begin{bmatrix} R^{\mathsf{T}} & 0 \\ S^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} U & 0 \\ 0 & 0 \end{bmatrix}$$

- then

$$A^{\mathsf{T}} = Z^{\mathsf{T}} \begin{bmatrix} U & 0 \\ 0 & 0 \end{bmatrix} Q^{\mathsf{T}}$$

  where $Z = H_k \cdots H_1 \Pi$
- we look at LU factorization and some of its variants: condensed LU, LDU, LDL^{T}, and Cholesky factorizations

4. EXISTENCE OF LU FACTORIZATION

- the solution method for a linear system $A\mathbf{x} = \mathbf{b}$ depends on the structure of A: A may be a sparse or dense matrix, or it may have one of many well-known structures, such as being a banded matrix, or a Hankel matrix
- for the general case of a dense, unstructured matrix A, the most common method is to obtain a decomposition $A = LU$, where L is lower triangular and U is upper triangular
- this decomposition is called the *LU factorization* or *LU decomposition*
- we deduce its existence via a constructive proof, namely, *Gaussian elimination*

- the motivation for this is something you learnt in middle school, i.e., solving $Ax = b$ by eliminating variables

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & \vdots & & \vdots \\ a_{n1}x_1 & + & \cdots & + & a_{nn}x_n & = & b_n \end{array}$$

- we proceed by multiplying the first equation by $-a_{21}/a_{11}$ and adding it to the second equation, and in general multiplying the first equation by $-a_{i1}/a_{11}$ and adding it to equation i and this leaves you with the equivalent system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ 0x_1 & + & a'_{22}x_2 & + & \cdots & + & a'_{2n}x_n & = & b'_2 \\ \vdots & & & & & & \vdots & & \vdots \\ 0x_1 & + & a'_{n2}x_2 & + & \cdots & + & a'_{nn}x_n & = & b'_n \end{array}$$

- continuing in this fashion, adding multiples of the second equation to each subsequent equation to make all elements below the diagonal equal to zero, you obtain an upper triangular system and may then solve for all $x_n, x_{n-1}, \ldots, x_1$ by back substitution
- getting the LU factorization $A = LU$ is very similar, the main difference is that you want not just the final upper triangular matrix (which is your U) but also to keep track of all the elimination steps (which is your L)

5. Gaussian elimination revisited

- we are going to look at Gaussian elimination in a slightly different light from what you learnt in your undergraduate linear algebra class
- we think of Gaussian elimination as the process of transforming A to an upper triangular matrix U is equivalent to multiplying A by a sequence of matrices to obtain U
- but instead of elementary matrices, we consider again a rank-1 change to I, i.e., a matrix of the form

$$I - \mathbf{u}\mathbf{v}^{\mathsf{T}}$$

where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$
- in Householder QR, we used Householder reflection matrices of the form

$$H = I - 2\mathbf{u}\mathbf{u}^{\mathsf{T}}$$

- in Gaussian elimination, we use so-called *Gauss transformation* or *elimination matrices* of the form

$$M = I - \mathbf{m}\mathbf{e}_i^{\mathsf{T}}$$

where $\mathbf{e}_i = [0, \ldots, 1, \ldots 0]^{\mathsf{T}}$ is the ith standard basis vector
- the same trick that led us to the appropriate $\mathbf{u}$ in Householder matrix can be applied to find the appropriate $\mathbf{m}$ too: suppose we want $M_1 = I - \mathbf{m}_1\mathbf{e}_1^{\mathsf{T}}$ to 'zero out' all the entries beneath the first in a vector $\mathbf{a}$, i.e.,

$$M_1 \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \gamma \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

i.e.,

$$\begin{aligned}
(I - \mathbf{m}_1 \mathbf{e}_1^{\mathsf{T}})\mathbf{a} &= \gamma \mathbf{e}_1 \\
\mathbf{a} - (\mathbf{e}_1^{\mathsf{T}}\mathbf{a})\mathbf{m}_1 &= \gamma \mathbf{e}_1 \\
a_1 \mathbf{m}_1 &= \mathbf{a} - \gamma \mathbf{e}_1
\end{aligned}$$

and if $a_1 \neq 0$, then we may set

$$\gamma = a_1, \quad \mathbf{m}_1 = \begin{bmatrix} 0 \\ a_2/a_1 \\ \vdots \\ a_n/a_1 \end{bmatrix}$$

- so we get

$$M_1 = I - \mathbf{m}_1 \mathbf{e}_1^{\mathsf{T}} = \begin{bmatrix} 1 & & & 0 \\ -a_2/a_1 & 1 & & \\ \vdots & 0 & \ddots & \\ -a_n/a_1 & & & 1 \end{bmatrix}$$

and, as required,

$$M_1 \mathbf{a} = \begin{bmatrix} 1 & & & 0 \\ -a_2/a_1 & 1 & & \\ \vdots & 0 & \ddots & \\ -a_n/a_1 & & & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} a_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = a_1 \mathbf{e}_1$$

- applying this to zero out the entries beneath a_{11} in the first column of a matrix A, we get $M_1 A = A_2$ where

$$A_2 = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} \end{bmatrix}$$

where the superscript in parenthesis denote that the entries have changed

- we will write

$$M_1 = \begin{bmatrix} 1 & & & 0 \\ -\ell_{21} & 1 & & \\ \vdots & 0 & \ddots & \\ -\ell_{n1} & & & 1 \end{bmatrix}, \quad \ell_{i1} = \frac{a_{i1}}{a_{11}}$$

for $i = 2, \dots, n$

- if we do this recursively, defining M_2 by

$$M_2 = \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & -\ell_{32} & 1 & & \\ \vdots & \vdots & & \ddots & \\ 0 & -\ell_{n2} & & & 1 \end{bmatrix}, \quad \ell_{i2} = \frac{a_{i2}^{(2)}}{a_{22}^{(2)}}$$

for $i = 3, \ldots, n$, then

$$M_2 A_2 = A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a_{n3}^{(3)} & \cdots & a_{nn}^{(3)} \end{bmatrix}$$

- in general, we have

$$M_k = \begin{bmatrix} 1 & & & & & \\ 0 & \ddots & & & & \\ \vdots & \ddots & 1 & & & \\ \vdots & & -\ell_{k+1,k} & 1 & & \\ \vdots & & \vdots & & \ddots & \\ 0 & & -\ell_{nk} & & & 1 \end{bmatrix}, \quad \ell_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$$

for $i = k+1, \ldots, n$, and

$$M_{n-1} M_{n-2} \cdots M_1 A = A_n \equiv \begin{bmatrix} u_{11} & \cdots & & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & u_{nn} \end{bmatrix}$$

or, equivalently,

$$A = M_1^{-1} M_2^{-1} \cdots M_{n-1}^{-1} U$$

- it turns out that M_j^{-1} is very easy to compute, we claim that

$$M_1^{-1} = \begin{bmatrix} 1 & & & 0 \\ \ell_{21} & 1 & & \\ \vdots & 0 & \ddots & \\ \ell_{n1} & & & 1 \end{bmatrix} \tag{5.1}$$

- to see this, consider the product

$$M_1 M_1^{-1} = \begin{bmatrix} 1 & & & 0 \\ -\ell_{21} & 1 & & \\ \vdots & 0 & \ddots & \\ -\ell_{n1} & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ \ell_{21} & 1 & & \\ \vdots & 0 & \ddots & \\ \ell_{n1} & & & 1 \end{bmatrix}$$

which can easily be verified to be equal to the identity matrix
- in general, we have

$$M_k^{-1} = \begin{bmatrix} 1 & & & & & \\ 0 & \ddots & & & & \\ \vdots & \ddots & 1 & & & \\ \vdots & & \ell_{k+1,k} & 1 & & \\ \vdots & & \vdots & & \ddots & \\ 0 & & \ell_{nk} & & & 1 \end{bmatrix} \tag{5.2}$$

- now, consider the product

$$M_1^{-1}M_2^{-1} = \begin{bmatrix} 1 & & & & \\ \ell_{21} & 1 & & & \\ \ell_{31} & 0 & 1 & & \\ \vdots & \vdots & & \ddots & \\ \ell_{n1} & 0 & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & \ell_{32} & 1 & & \\ \vdots & \vdots & & \ddots & \\ 0 & \ell_{n2} & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & & & & \\ \ell_{21} & 1 & & & \\ \ell_{31} & \ell_{32} & 1 & & \\ \vdots & \vdots & & \ddots & \\ \ell_{n1} & \ell_{n2} & & & 1 \end{bmatrix}$$

- so inductively we get

$$M_1^{-1}M_2^{-1}\cdots M_{n-1}^{-1} = \begin{bmatrix} 1 & & & & \\ \ell_{21} & \ddots & & & \\ \vdots & \ell_{32} & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ \ell_{n1} & \ell_{n2} & \cdots & \ell_{n,n-1} & 1 \end{bmatrix}$$

- it follows that under proper circumstances, we can write $A = LU$ where

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \ell_{21} & 1 & 0 & \cdots & 0 \\ \ell_{31} & \ell_{32} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \ell_{n1} & \ell_{n2} & \cdots & \ell_{n,n-1} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} u_{11} & u_{12} & \cdots & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & \cdots & u_{2n} \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & u_{nn} \end{bmatrix}$$

- what exactly are proper circumstances?
- we will discuss them in the next section and also introduce *pivoting* to ensure that they are always satisfied

6. NEED FOR PIVOTING

- we must have $a_{kk}^{(k)} \neq 0$, or we cannot proceed with the decomposition
- for example, if

$$A = \begin{bmatrix} 0 & 1 & 11 \\ 3 & 7 & 2 \\ 2 & 9 & 3 \end{bmatrix} \quad \text{or} \quad A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 4 \\ 7 & 1 & 2 \end{bmatrix}$$

  Gaussian elimination will fail; note that both matrices are nonsingular
- in the first case, it fails immediately; in the second case, it fails after the subdiagonal entries in the first column are zeroed, and we find that $a_{22}^{(k)} = 0$
- in general, we must have $\det A_{ii} \neq 0$ for $i = 1, \ldots, n$ where

$$A_{ii} = \begin{bmatrix} a_{11} & \cdots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{ii} \end{bmatrix}$$

  for the LU factorization to exist

- the existence of LU factorization (without pivoting) can be guaranteed by several conditions, one example is *column[1] diagonal dominance*: if a nonsingular $A \in \mathbb{R}^{n\times n}$ satisfies

$$|a_{jj}| \geq \sum_{\substack{i=1\\ i\neq j}}^{n} |a_{ij}|, \quad j = 1, \dots, n,$$

  then one can guarantee that Gaussian elimination as described above produces $A = LU$ with $|\ell_{ij}| \leq 1$
- there are necessary and sufficient conditions guaranteeing the existence of LU decomposition but those are difficult to check in practice and we do not state them here
- how can we obtain the LU factorization for a general nonsingular matrix?
- if A is nonsingular, then *some* element of the first column must be nonzero
- if $a_{i1} \neq 0$, then we can interchange row i with row 1 and proceed
- this is equivalent to multiplying A by a permutation matrix Π_1 that interchanges row 1 and row i:

$$\Pi_1 = \begin{bmatrix} 0 & \cdots & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ 1 & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 0 \\ & & & & & 1 & & \\ & & & & & & \ddots & \\ & & & & & & & 1 \end{bmatrix}$$

- thus $M_1\Pi_1 A = A_2$ (refer to earlier lecture notes for more information about permutation matrices)
- then, since A_2 is nonsingular, some element of column 2 of A_2 below the diagonal must be nonzero
- proceeding as before, we compute $M_2\Pi_2 A_2 = A_3$, where Π_2 is another permutation matrix
- continuing, we obtain

$$A = (M_{n-1}\Pi_{n-1}\cdots M_1\Pi_1)^{-1}U$$

- it can easily be shown that $\Pi A = LU$ where Π is a permutation matrix — easy but a bit of a pain because notation is cumbersome
- so we will be informal but you'll get the idea
- for example if after two steps we get (recall that permutation matrices or orthogonal matrices),

$$\begin{aligned} A &= (M_2\Pi_2 M_1\Pi_1)^{-1}A_2 \\ &= \Pi_1^{\mathsf{T}} M_1^{-1}\Pi_2^{\mathsf{T}} M_2^{-1}A_2 \\ &= \Pi_1^{\mathsf{T}}\Pi_2^{\mathsf{T}}(\Pi_2 M_1^{-1}\Pi_2^{\mathsf{T}})M_2^{-1}A_2 \\ &= \Pi^{\mathsf{T}} L_1 L_2 A_2 \end{aligned}$$

  then
  - $\Pi = \Pi_2\Pi_1$ is a permutation matrix
  - $L_2 = M_2^{-1}$ is a unit lower triangular matrix

[1] the usual type of diagonal dominance, i.e., $|a_{ii}| \geq \sum_{j=1, j\neq i}^{n} |a_{ij}|$, $i = 1, \dots, n$, is called row diagonal dominance

- $L_1 = \Pi_2 M_1^{-1} \Pi_2^\intercal$ will always be a unit lower triangular matrix because M_1^{-1} is of the form in

$$M_1^{-1} = \begin{bmatrix} 1 & \\ \boldsymbol{\ell} & I \end{bmatrix} = \begin{bmatrix} 1 & & & 0 \\ \ell_{21} & 1 & & \\ \vdots & 0 & \ddots & \\ \ell_{n1} & & & 1 \end{bmatrix} \tag{6.1}$$

whereas Π_2 must be of the form

$$\Pi_2 = \begin{bmatrix} 1 & \\ & \widehat{\Pi}_2 \end{bmatrix}$$

for some $(n-1) \times (n-1)$ permutation matrix $\widehat{\Pi}_2$ and so

$$\Pi_2 M_1^{-1} \Pi_2^\intercal = \begin{bmatrix} 1 & 0 \\ \widehat{\Pi}_2 \boldsymbol{\ell} & I \end{bmatrix}$$

in other words $\Pi_2 M_1^{-1} \Pi_2^\intercal$ also has the form in (6.1)

- if we do one more steps we get

$$\begin{aligned} A &= (M_3 \Pi_3 M_2 \Pi_2 M_1 \Pi_1)^{-1} A_3 \\ &= \Pi_1^\intercal M_1^{-1} \Pi_2^\intercal M_2^{-1} \Pi_3^\intercal M_3^{-1} A_3 \\ &= \Pi_1^\intercal \Pi_2^\intercal \Pi_3^\intercal (\Pi_3 \Pi_2 M_1^{-1} \Pi_2^\intercal \Pi_3^\intercal)(\Pi_3 M_2^{-1} \Pi_3^\intercal) M_3^{-1} A_3 \\ &= \Pi^\intercal L_1 L_2 L_3 A_3 \end{aligned}$$

where

- $\Pi = \Pi_3 \Pi_2 \Pi_1$ is a permutation matrix
- $L_3 = M_3^{-1}$ is a unit lower triangular matrix
- $L_2 = \Pi_3 M_2^{-1} \Pi_3^\intercal$ will always be a unit lower triangular matrix because M_2^{-1} is of the form

$$M_2^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & \boldsymbol{\ell} & I \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ 0 & 1 & & & \\ 0 & -\ell_{32} & 1 & & \\ \vdots & \vdots & & \ddots & \\ 0 & -\ell_{n2} & & & 1 \end{bmatrix} \tag{6.2}$$

whereas Π_3 must be of the form

$$\Pi_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & \widehat{\Pi}_3 \end{bmatrix}$$

for some $(n-2) \times (n-2)$ permutation matrix $\widehat{\Pi}_3$ and so

$$\Pi_3 M_2^{-1} \Pi_3^\intercal = \begin{bmatrix} 1 & & \\ & 1 & 0 \\ & \widehat{\Pi}_3 \boldsymbol{\ell} & I \end{bmatrix}$$

in other words $\Pi_3 M_2^{-1} \Pi_3^\intercal$ also has the form in (6.2)

- $L_1 = \Pi_3 \Pi_2 M_1^{-1} \Pi_2^\intercal \Pi_3^\intercal$ will always be a unit lower triangular matrix for the same reason above because $\Pi_3 \Pi_2$ must have the form

$$\Pi_3 \Pi_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & \widehat{\Pi}_3 \end{bmatrix} \begin{bmatrix} 1 & \\ & \widehat{\Pi}_2 \end{bmatrix} = \begin{bmatrix} 1 & \\ & \Pi_{32} \end{bmatrix}$$

for some $(n-1)\times(n-1)$ permutation matrix

$$\Pi_{32} = \begin{bmatrix} 1 & \\ & \widehat{\Pi}_3 \end{bmatrix} \widehat{\Pi}_2$$

- more generally if we keep doing this, then

$$A = \Pi^{\mathsf{T}} L_1 L_2 \cdots L_{n-1} A_{n-1}$$

where
  - $\Pi = \Pi_{n-1}\Pi_{n-2}\cdots\Pi_1$ is a permutation matrix
  - $L_{n-1} = M_{n-1}^{-1}$ is a unit lower triangular matrix
  - $L_k = \Pi_{n-1}\cdots\Pi_{k+1}M_k^{-1}\Pi_{k+1}^{\mathsf{T}}\cdots\Pi_{n-1}^{\mathsf{T}}$ is a unit lower triangular matrix for all $k = 1,\dots,n-2$
  - $A_{n-1} = U$ is an upper triangular matrix
  - $L = L_1 L_2 \cdots L_{n-1}$ is a unit lower triangular matrix
- this algorithm with the row permutations is called *Gaussian elimination with partial pivoting* or GEPP for short; we will say more in the next section

7. Pivoting strategies

- the (k,k) entry at step k during Gaussian elimination is called the *pivoting entry* or just *pivot* for short
- in the preceding section, we said that if the pivoting entry is zero, i.e., $a_{kk}^{(k)} = 0$, then we just need to find an entry below it in the same column, i.e., $a_{ik}^{(k)}$ for some $i > k$, and then permute this entry into the pivoting position, before carrying on with the algorithm
- but it is really better to choose the *largest* entry below the pivot, and not just any nonzero entry
- that is, the permutation Π_k is chosen so that row k is interchanged with row i, where $|a_{ik}^{(k)}| = \max_{i=k,k+1,\dots,n}|a_{ik}^{(k)}|$, i.e., upon this permutation, we are guaranteed

$$|a_{kk}^{(k)}| = \max_{i=k,k+1,\dots,n} |a_{ik}^{(k)}|$$

- this guarantees that $|\ell_{kj}| \le 1$ for all k and j
- this strategy is known as *partial pivoting*, which is guaranteed to produce an LU factorization if $A \in \mathbb{R}^{m\times n}$ has full column-rank, i.e., $\operatorname{rank}(A) = n \ge m$ (it can fail if A doesn't have full column-rank, think of what happens when A has a column of zeros)
- another common strategy, *complete pivoting*, which uses both row and column interchanges to ensure that at step k of the algorithm, the element $a_{kk}^{(k)}$ is the largest element in absolute value from the entire submatrix obtained by deleting the first $k-1$ rows and columns, i.e.,

$$|a_{kk}^{(k)}| = \max_{\substack{i=k,k+1,\dots,n\\ j=k,k+1,\dots,n}} |a_{ij}^{(k)}|$$

- in this case we need both row and column permutation matrices, i.e., we get

$$\Pi_1 A \Pi_2 = LU$$

when we do complete pivoting
- complete pivoting is necessary when $\operatorname{rank}(A) < \min\{m,n\}$
- there are yet other pivoting strategies due to considerations such as preserving sparsity (if you're interested, look up *minimum degree algorithm* or *Markowitz algorithm*) or a tradeoff between partial and complete pivoting (e.g., *rook pivoting*)

8. uniqueness of the LU factorization

- the LU decomposition of a nonsingular matrix, if it exists (i.e., without row or column permutations), is unique
- if A has two LU decompositions, $A = L_1U_1$ and $A = L_2U_2$
- from $L_1U_1 = L_2U_2$ we obtain $L_2^{-1}L_1 = U_2U_1^{-1}$
- the inverse of a unit lower triangular matrix is a unit lower triangular matrix, and the product of two unit lower triangular matrices is a unit lower triangular matrix, so $L_2^{-1}L_1$ must be a unit lower triangular matrix
- similarly, $U_2U_1^{-1}$ is an upper triangular matrix
- the only matrix that is both upper triangular and unit lower triangular is the identity matrix I, so we must have $L_1 = L_2$ and $U_1 = U_2$

9. Gauss–Jordan elimination

- a variant of Gaussian elimination is called *Gauss–Jordan elimination*
- it entails zeroing elements above the diagonal as well as below, transforming an $m \times n$ matrix into *reduced row echelon form*, i.e., a form where all pivoting entries in U are 1 and all entries above the pivots are zeros
- this is what you probably learnt in your undergraduate linear algebra class, e.g.,

$$A = \begin{bmatrix} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 5 & 35 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 2 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -2 & -3 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

- the main drawback is that the elimination process can be numerically unstable, since the multipliers can be large
- furthermore the way it is done in undergraduate linear algebra courses is that the elimination matrices (i.e., the L and Π) are not stored

10. condensed LU factorization

- just like QR and svd, LU factorization with complete pivoting has a condensed form too
- let $A \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(A) = r \le \min\{m, n\}$, recall that gecp yields

$$\begin{aligned} \Pi_1 A \Pi_2 &= LU \\ &= \begin{bmatrix} L_{11} & 0 \\ L_{21} & I_{m-r} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} L_{11} \\ L_{21} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \end{bmatrix} =: \widetilde{L}\widetilde{U} \end{aligned}$$

  where $L_{11} \in \mathbb{R}^{r\times r}$ is unit lower triangular (thus nonsingular) and $U_{11} \in \mathbb{R}^{r\times r}$ is also nonsingular
- note that $\widetilde{L} \in \mathbb{R}^{m\times r}$ and $\widetilde{U} \in \mathbb{R}^{r\times n}$ and so

$$A = (\Pi_1^{\mathsf{T}}\widetilde{L})(\widetilde{U}\Pi_2^{\mathsf{T}})$$

  is a rank-retaining factorization

11. LDU and LDL^{T} factorizations

- if $A \in \mathbb{R}^{n\times n}$ has nonsingular principal submatrices $A_{1:k,1:k}$ for $k = 1, \dots, n$, then there exists a unit lower triangular matrix $L \in \mathbb{R}^{n\times n}$, a unit upper triangular matrix $U \in \mathbb{R}^{n\times n}$, and a diagonal matrix $D = \operatorname{diag}(d_{11}, \dots, d_{nn}) \in \mathbb{R}^{n\times n}$ such that

$$A = LDU = \begin{bmatrix} 1 & & & 0 \\ \ell_{21} & 1 & & \\ \vdots & & \ddots & \\ \ell_{n1} & \ell_{n2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} d_{11} & & & \\ & d_{22} & & \\ & & \ddots & \\ & & & d_{nn} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & \cdots & u_{1n} \\ & 1 & & u_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

- this is called the *LDU factorization* of A
- if A is furthermore symmetric, then $L = U^{\mathsf{T}}$ and this called the *LDL^{T} factorization*
- if they exist, then both LDU and LDL^{T} factorizations are unique (exercise)
- if a symmetric A has an LDL^{T} factorization and if $d_{ii} > 0$ for all $i = 1, \dots, n$, then A is positive definite
- in fact, even though $d_{11}, \dots, d_{nn}$ are not the eigenvalues of A (why not?), they must have the same signs as the eigenvalues of A, i.e., if A has p positive eigenvalues, q negative eigenvalues, and z zero eigenvalues, then there are exactly p, q, and z positive, negative, and zero entries in $d_{11}, \dots, d_{nn}$ — a consequence of the Sylvester law of inertia
- unfortunately, both LDU and LDL^{T} factorizations are difficult to compute because
  - the condition on the principal submatrices is difficult to check in advance
  - algorithms for computing them are invariably unstable because size of multipliers cannot be bounded in terms of the entries of A
- for example, the LDL^{T} factorization of a 2×2 symmetric matrix is

$$\begin{bmatrix} a & c \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & c \\ 0 & d - (c/a)c \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ c/a & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & d - (c/a)c \end{bmatrix} \begin{bmatrix} 1 & c/a \\ 0 & 1 \end{bmatrix}$$

- so

$$\begin{bmatrix} \varepsilon & 1 \\ 1 & \varepsilon \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/\varepsilon & 1 \end{bmatrix} \begin{bmatrix} \varepsilon & 0 \\ 0 & \varepsilon - 1/\varepsilon \end{bmatrix} \begin{bmatrix} 1 & 1/\varepsilon \\ 0 & 1 \end{bmatrix}$$

  the elements of L and D are arbitrarily large when $|\varepsilon|$ is small
- note that you can't do partial or complete pivoting in LDL^{T} factorization since those could destroy the symmetry in A
- nonetheless there is one special case when LDL^{T} factorization not only exists but can be computed in an efficient and stable way — when A is positive definite