STAT 309: MATHEMATICAL COMPUTATIONS I
FALL 2023
LECTURE 12

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 rotation in the plane $\mathbb{R}^2$
- 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 10, 2023, version 1.0. Comments, bug reports: `lekheng@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 $\beta = \rho \sin\theta$
- the representation $\alpha = \rho\cos\theta$, $\beta = \rho\sin\theta$ is a purely theoretical device

$$\boxed{\text{anyone who stores } \theta \text{ instead of } \alpha \text{ and } \beta \text{ fails this class instantly}}$$

- 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 $Q^\intercal A = R$
- it is easy to see that $O(n^2)$ rotations are required

- the matrix above is a rotation in the plane $\operatorname{span}\{\mathbf{e}_i, \mathbf{e}_j\}$ and is called either *Givens rotations* or *plane rotations*

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccc} & & & i & & & & j & & & \end{array} \\
\begin{array}{c} \\ \\ \\ i \\ \\ \\ \\ j \\ \\ \\ \\ \end{array} &
\begin{bmatrix}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
 & & & \gamma & & & & \sigma & & & \\
 & & & & 1 & & & & & & \\
 & & & & & \ddots & & & & & \\
 & & & & & & 1 & & & & \\
 & & & -\sigma & & & & \gamma & & & \\
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{bmatrix}
\end{array}
=: G_{ij}(\theta)
$$

where $\alpha = \rho \cos\theta$, $\beta = \rho \sin\theta$

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
- a Givens rotation $G_{ij}(\theta) \in \mathbb{R}^{n \times n}$ should be stored as $(i, j, \alpha, \beta) \in \{1, \ldots, n\}^2 \times \mathbb{R}^2$, just as a Householder reflection $H_{\mathbf{v}} = I - 2\mathbf{v}\mathbf{v}^{\mathsf{T}} \in \mathbb{R}^{n \times n}$ should be stored as a vector $\mathbf{v} \in \mathbb{R}^n$, just as a permutation matrix $P_\sigma \in \mathbb{R}^{n \times n}$ should be stored as a permutation $\sigma \in \mathfrak{S}_n$

anyone who stores a Givens, Householder, or permutation matrix as a full $n \times n$ matrix fails this class instantly

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 $Z = H_k \cdots H_1 \Pi$

4. Existence of LU factorization

- we next look at LU factorization and some of its variants: condensed LU, LDU, LDL^T, and Cholesky factorizations
- 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 $\mathbf{e}_i = [0, \dots, 1, \dots 0]^\mathsf{T}$ 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 $M_1A = A_2$ 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 $i = 3, \dots, n$, then

$$M_2A_2 = A_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & \cdots & a_{3n}^{(3)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & a_{n3}^{(3)} & \cdots & a_{nn}^{(3)} \end{bmatrix}$$

- 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

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

- 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 $i = 1, \dots, n$ 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

[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

- 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
  - $L_1 = \Pi_2 M_1^{-1}\Pi_2^{\mathsf{T}}$ 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

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

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

- if we do one more steps we get

$$\begin{aligned}
A &= (M_3\Pi_3M_2\Pi_2M_1\Pi_1)^{-1}A_3\\
&= \Pi_1^{\mathsf{T}}M_1^{-1}\Pi_2^{\mathsf{T}}M_2^{-1}\Pi_3^{\mathsf{T}}M_3^{-1}A_3\\
&= \Pi_1^{\mathsf{T}}\Pi_2^{\mathsf{T}}\Pi_3^{\mathsf{T}}(\Pi_3\Pi_2M_1^{-1}\Pi_2^{\mathsf{T}}\Pi_3^{\mathsf{T}})(\Pi_3M_2^{-1}\Pi_3^{\mathsf{T}})M_3^{-1}A_3\\
&= \Pi^{\mathsf{T}}L_1L_2L_3A_3
\end{aligned}$$

  where
  - $\Pi = \Pi_3\Pi_2\Pi_1$ is a permutation matrix
  - $L_3 = M_3^{-1}$ is a unit lower triangular matrix
  - $L_2 = \Pi_3M_2^{-1}\Pi_3^{\mathsf{T}}$ 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

$$\Pi_3M_2^{-1}\Pi_3^{\mathsf{T}} = \begin{bmatrix} 1 & & \\ & 1 & 0 \\ & \widehat{\Pi}_3\boldsymbol{\ell} & I \end{bmatrix}$$

    in other words $\Pi_3M_2^{-1}\Pi_3^{\mathsf{T}}$ also has the form in (6.2)
  - $L_1 = \Pi_3\Pi_2M_1^{-1}\Pi_2^{\mathsf{T}}\Pi_3^{\mathsf{T}}$ 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

$$A = \Pi^{\mathsf{T}}L_1L_2\cdots L_{n-1}A_{n-1}$$

  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
  - $L = L_1L_2\cdots L_{n-1}$ 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