矩阵代数

Matrix Algebra

Baobin Li

Email:libb@ucas.ac.cn

School of Computer Science and Technology, UCAS

Introduction

- Throughout $\Re$ will denote the set of real numbers, and $\mathcal{C}$ will denote the complex numbers.
- The set of all n -tuples of real numbers will be denoted by $\Re^n$, and the set of all complex n -tuples will be denoted by $\mathcal{C}^n$.
- Analogously, $\Re^{m\times n}$ and $\mathcal{C}^{m\times n}$denote the $m \times n$ matrices containing real numbers and complex numbers, respectively.
- Matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ are defined to be **equal matrices** when $\mathbf{A}$ and $\mathbf{B}$ have the same shape and corresponding entries are equal.
- That is, $a_{ij} = b_{ij}$ for each $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$.
- An array (or matrix) consisting of a single column, such as $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, is called a column vector, while an array consisting of a single row, such as $\mathbf{v} = (1\ 2\ 3)$, is called a row vector.

Addition and Transposition

Addition of Matrices

If $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices, the ***sum*** of $\mathbf{A}$ and $\mathbf{B}$ is defined to be the $m \times n$ matrix $\mathbf{A}+\mathbf{B}$ obtained by adding corresponding entries. That is,

$$[\mathbf{A}+\mathbf{B}]_{ij} = [\mathbf{A}]_{ij} + [\mathbf{B}]_{ij} \quad \text{for each } i \text{ and } j.$$

- The matrix $(-\mathbf{A})$ called the **additive inverse** of $\mathbf{A}$, is defined to be the matrix obtained by negating each entry of $\mathbf{A}$.
- That is, if $\mathbf{A} = [a_{ij}]$, then $-\mathbf{A} = [-a_{ij}]$. This allows matrix subtraction to be defined in the natural way.
- For two matrices of the same shape, the difference $\mathbf{A} - \mathbf{B}$ is defined to be the matrix $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$ so that

$$[\mathbf{A}-\mathbf{B}]_{ij} = [\mathbf{A}]_{ij} - [\mathbf{B}]_{ij} \quad for \quad each\ i\ and\ j.$$

Properties of Matrix Addition

For $m \times n$ matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$, the following properties hold.

Closure property:	$\mathbf{A}+\mathbf{B}$ is again an $m \times n$ matrix.
Associative property:	$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+(\mathbf{B}+\mathbf{C})$.
Commutative property:	$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$.
Additive identity:	The $m \times n$ matrix $\mathbf{0}$ consisting of all zeros has the property that $\mathbf{A}+\mathbf{0}=\mathbf{A}$.
Additive inverse:	The $m \times n$ matrix $(-\mathbf{A})$ has the property that $\mathbf{A}+(-\mathbf{A})=\mathbf{0}$.

Scalar Multiplication

The product of a scalar α times a matrix $\mathbf{A}$, denoted by $\alpha\mathbf{A}$, is defined to be the matrix obtained by multiplying each entry of $\mathbf{A}$ by α. That is, $[\alpha\mathbf{A}]_{ij}=\alpha[\mathbf{A}]_{ij}$ for each i and j.

Properties of Scalar Multiplication

For $m \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ and for scalars α and β, the following properties hold.

Closure property: $\alpha\mathbf{A}$ is again an $m \times n$ matrix.

Associative property: $(\alpha\beta)\mathbf{A} = \alpha(\beta\mathbf{A})$.

Distributive property: $\alpha(\mathbf{A}+\mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B}$. Scalar multiplication is distributed over matrix addition.

Distributive property: $(\alpha+\beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A}$. Scalar multiplication is distributed over scalar addition.

Identity property: $1\mathbf{A} = \mathbf{A}$. The number 1 is an identity element under scalar multiplication.

Transpose

The ***transpose*** of $\mathbf{A}_{m\times n}$ is defined to be the $n \times m$ matrix $\mathbf{A}^T$ obtained by interchanging rows and columns in $\mathbf{A}$. More precisely, if $\mathbf{A} = [a_{ij}]$, then $[\mathbf{A}^T]_{ij} = a_{ji}$. For example,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

It should be evident that for all matrices, $(\mathbf{A}^T)^T = \mathbf{A}$.

Conjugate Transpose

For $\mathbf{A} = [a_{ij}]$, the ***conjugate matrix*** is defined to be $\overline{\mathbf{A}} = [\overline{a}_{ij}]$, and the ***conjugate transpose*** of $\mathbf{A}$ is defined to be $\bar{\mathbf{A}}^T = \overline{\mathbf{A}^T}$. From now on, $\bar{\mathbf{A}}^T$ will be denoted by $\mathbf{A}^*$, so $[\mathbf{A}^*]_{ij} = \overline{a}_{ji}$. For example,

$$\begin{pmatrix} 1-4\mathrm{i} & \mathrm{i} & 2 \\ 3 & 2+\mathrm{i} & 0 \end{pmatrix}^* = \begin{pmatrix} 1+4\mathrm{i} & 3 \\ -\mathrm{i} & 2-\mathrm{i} \\ 2 & 0 \end{pmatrix}.$$

$(\mathbf{A}^*)^* = \mathbf{A}$ for all matrices, and $\mathbf{A}^* = \mathbf{A}^T$ whenever $\mathbf{A}$ contains only real entries. Sometimes the matrix $\mathbf{A}^*$ is called the ***adjoint*** of $\mathbf{A}$.

Properties of the Transpose

If $\mathbf{A}$ and $\mathbf{B}$ are two matrices of the same shape, and if α is a scalar, then each of the following statements is true.

$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T \quad \text{and} \quad (\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*.$$

$$(\alpha\mathbf{A})^T = \alpha\mathbf{A}^T \quad \text{and} \quad (\alpha\mathbf{A})^* = \overline{\alpha}\mathbf{A}^*.$$

Sometimes transposition doesn't change anything. For example, if

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}, \quad \text{then} \quad \mathbf{A}^T = \mathbf{A}.$$

This is because the entries in $\mathbf{A}$ are symmetrically located about the ***main diagonal***—the line from the upper-left-hand corner to the lower-right-hand corner.

Matrices of the form $\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$ are called ***diagonal matrices***, and they are clearly symmetric in the sense that $\mathbf{D} = \mathbf{D}^T$. This is one of several kinds of symmetries described below.

Symmetries

Let $\mathbf{A} = [a_{ij}]$ be a square matrix.

- $\mathbf{A}$ is said to be a ***symmetric matrix*** whenever $\mathbf{A} = \mathbf{A}^T$, i.e., whenever $a_{ij} = a_{ji}$.
- $\mathbf{A}$ is said to be a ***skew-symmetric matrix*** whenever $\mathbf{A} = -\mathbf{A}^T$, i.e., whenever $a_{ij} = -a_{ji}$.
- $\mathbf{A}$ is said to be a ***hermitian matrix*** whenever $\mathbf{A} = \mathbf{A}^*$, i.e., whenever $a_{ij} = \overline{a}_{ji}$. This is the complex analog of symmetry.
- $\mathbf{A}$ is said to be a ***skew-hermitian matrix*** when $\mathbf{A} = -\mathbf{A}^*$, i.e., whenever $a_{ij} = -\overline{a}_{ji}$. This is the complex analog of skew symmetry.

Matrix Multiplication

- The concept of linearity is the underlying theme of our subject.
- In elementary mathematics "linear function" refers to straight lines.
- In higher mathematics linearity means something much more general.

Linear Functions

Suppose that $\mathcal{D}$ and $\mathcal{R}$ are sets that possess an addition operation as well as a scalar multiplication operation—i.e., a multiplication between scalars and set members. A function f that maps points in $\mathcal{D}$ to points in $\mathcal{R}$ is said to be a ***linear function*** whenever f satisfies the conditions that

$$f(x+y) = f(x) + f(y)$$

and

$$f(\alpha x) = \alpha f(x)$$

for every x and y in $\mathcal{D}$ and for all scalars α. These two conditions may be combined by saying that f is a linear function whenever

$$f(\alpha x + y) = \alpha f(x) + f(y)$$

for all scalars α and for all $x, y \in \mathcal{D}$.

- One of the simplest linear functions is $f(x) = \alpha x$, whose graph in $\Re^2$ is a straight line through the origin.
- However, $f(x) = \alpha x + \beta$ does not qualify the title linear function—it is a linear function that has been translated by a constant β.
- Translations of linear functions are referred to as **affine functions**.
- In $\Re^3$, the surface $f(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2$ is a plane through the origin, and it is easy to verify that f is a linear function.
- For $\beta \neq 0$, $f(x_1, x_2) = \alpha_1 x_1 + \alpha_2 x_2 + \beta$ is no longer a linear function—an affine function.
- Although we cannot visualize higher dimensions with our eyes, it seems reasonable to suggest that a general linear function of the form

$$f(x_1, x_2, \cdots, x_n) = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n.$$

- For scalars α_j and matrices $\mathbf{X}_j$, the expression

$$\alpha_1 \mathbf{X}_1 + \alpha_2 \mathbf{X}_2 + \cdots + \alpha_n \mathbf{X}_n = \sum_{j=1}^{n} \alpha_j \mathbf{X}_j$$

is called a linear combination of the $\mathbf{X}_j$'s

- If you are given the task of formulating a definition for composing two matrices $\mathbf{A}$ and $\mathbf{B}$ in some sort of "natural" multiplicative fashion, your first attempt would probably be to compose $\mathbf{A}$ and $\mathbf{B}$ by multiplying corresponding entries–much the same way matrix addition is defined.
- Asked then to defend the usefulness of such a definition, you might be hard pressed to provide a truly satisfying response.
- Unless a person is in the right frame of mind, the issue of deciding how to best define matrix multiplication is not all transparent.
- The world had to wait for Arthur Cayley to come to this proper frame.
- Around 1855, Cayley became interested in composing linear functions.
- In particular, he was investigating linear functions of the following type

$$f(\mathbf{x}) = f\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} \quad \text{and} \quad g(\mathbf{x}) = g\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} Ax_1 + Bx_2 \\ Cx_1 + Dx_2 \end{pmatrix}.$$

Consider, as Cayley did, composing f and g to create another linear function

$$h(\mathbf{x}) = f\big(g(\mathbf{x})\big) = f\begin{pmatrix} Ax_1 + Bx_2 \\ Cx_1 + Dx_2 \end{pmatrix} = \begin{pmatrix} (aA + bC)x_1 + (aB + bD)x_2 \\ (cA + dC)x_1 + (cB + dD)x_2 \end{pmatrix}.$$

It was Cayley's idea to use matrices of coefficients to represent these linear functions. That is, f, g, and h are represented by

$$\mathbf{F} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} aA+bC & aB+bD \\ cA+dC & cB+dD \end{pmatrix}.$$

After making this association, it was only natural for Cayley to call $\mathbf{H}$ the *composition* (or *product*) of $\mathbf{F}$ and $\mathbf{G}$, and to write

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} aA+bC & aB+bD \\ cA+dC & cB+dD \end{pmatrix}.$$

In other words, the product of two matrices represents the composition of the two associated linear functions. By means of this observation, Cayley brought to life the subjects of matrix analysis and linear algebra.

Matrix Multiplication

- Matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be ***conformable*** for multiplication in the order $\mathbf{AB}$ whenever $\mathbf{A}$ has exactly as many columns as $\mathbf{B}$ has rows—i.e., $\mathbf{A}$ is $m \times p$ and $\mathbf{B}$ is $p \times n$.

- For conformable matrices $\mathbf{A}_{m\times p} = [a_{ij}]$ and $\mathbf{B}_{p\times n} = [b_{ij}]$, the ***matrix product*** $\mathbf{AB}$ is defined to be the $m \times n$ matrix whose (i, j)-entry is the inner product of the i^{th} row of $\mathbf{A}$ with the j^{th} column in $\mathbf{B}$. That is,

$$[\mathbf{AB}]_{ij} = \mathbf{A}_{i*}\mathbf{B}_{*j} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}.$$

- In case $\mathbf{A}$ and $\mathbf{B}$ fail to be conformable—i.e., $\mathbf{A}$ is $m \times p$ and $\mathbf{B}$ is $q \times n$ with $p \neq q$—then no product $\mathbf{AB}$ is defined.

Matrix Multiplication Is Not Commutative

Matrix multiplication is a noncommutative operation—i.e., it is possible for $\mathbf{AB} \neq \mathbf{BA}$, even when both products exist and have the same shape.

For example, if

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}_{2\times 3} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{pmatrix}_{3\times 4}$$

inside ones match

shape of the product

then the product **AB** exists and has shape 2×4. Consider a typical entry of this product, say, the (2,3)-entry. The definition says $[\mathbf{AB}]_{23}$ is obtained by forming the inner product of the second row of **A** with the third column of **B**

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ \boxed{a_{21} \quad a_{22} \quad a_{23}} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \boxed{\begin{matrix} b_{13} \\ b_{23} \\ b_{33} \end{matrix}} & b_{14} \\ b_{21} & b_{22} & & b_{24} \\ b_{31} & b_{32} & & b_{34} \end{pmatrix},$$

so

$$[\mathbf{AB}]_{23} = \mathbf{A}_{2*}\mathbf{B}_{*3} = a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} = \sum_{k=1}^{3} a_{2k}b_{k3}.$$

There are various ways to express the individual rows and columns of a matrix product. For example, the i^{th} row of $\mathbf{AB}$ is

$$[\mathbf{AB}]_{i*} = \left[\mathbf{A}_{i*}\mathbf{B}_{*1} \,|\, \mathbf{A}_{i*}\mathbf{B}_{*2} \,|\, \cdots \,|\, \mathbf{A}_{i*}\mathbf{B}_{*n}\right] = \mathbf{A}_{i*}\mathbf{B}$$

$$= \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{ip} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{1*} \\ \mathbf{B}_{2*} \\ \vdots \\ \mathbf{B}_{p*} \end{pmatrix} = a_{i1}\mathbf{B}_{1*} + a_{i2}\mathbf{B}_{2*} + \cdots + a_{ip}\mathbf{B}_{p*}.$$

As shown below, there are similar representations for the individual columns.

Rows and Columns of a Product

Suppose that $\mathbf{A} = [a_{ij}]$ is $m \times p$ and $\mathbf{B} = [b_{ij}]$ is $p \times n$.

- $[\mathbf{AB}]_{i*} = \mathbf{A}_{i*}\mathbf{B}$ $\left[(\, i^{th} \text{ row of } \mathbf{AB}\,) = (\, i^{th} \text{ row of } \mathbf{A}\,) \times \mathbf{B}\right]$.
- $[\mathbf{AB}]_{*j} = \mathbf{AB}_{*j}$ $\left[(\, j^{th} \text{ col of } \mathbf{AB}\,) = \mathbf{A} \times (\, j^{th} \text{ col of } \mathbf{B}\,)\right]$.
- $[\mathbf{AB}]_{i*} = a_{i1}\mathbf{B}_{1*} + a_{i2}\mathbf{B}_{2*} + \cdots + a_{ip}\mathbf{B}_{p*} = \sum_{k=1}^{p} a_{ik}\mathbf{B}_{k*}$.
- $[\mathbf{AB}]_{*j} = \mathbf{A}_{*1}b_{1j} + \mathbf{A}_{*2}b_{2j} + \cdots + \mathbf{A}_{*p}b_{pj} = \sum_{k=1}^{p} \mathbf{A}_{*k}b_{kj}$.

These last two equations show that rows of $\mathbf{AB}$ are combinations of rows of $\mathbf{B}$, while columns of $\mathbf{AB}$ are combinations of columns of $\mathbf{A}$.

- Matrix multiplication provides a convenient representation for a linear system of equations.

Linear Systems

Every linear system of m equations in n unknowns

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1, \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2, \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m, \end{array}$$

can be written as a single matrix equation $\mathbf{Ax} = \mathbf{b}$ in which

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Conversely, every matrix equation of the form $\mathbf{A}_{m\times n}\mathbf{x}_{n\times 1} = \mathbf{b}_{m\times 1}$ represents a system of m linear equations in n unknowns.

Properties of Matrix Multiplication

- There are some differences between scalar and matrix algebra—-most notable is the fact that matrix multiplication is not commutative.
- But there are also some important similarities.

Distributive and Associative Laws

For conformable matrices each of the following is true.

- $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$ (left-hand distributive law).
- $(\mathbf{D}+\mathbf{E})\mathbf{F} = \mathbf{DF}+\mathbf{EF}$ (right-hand distributive law).
- $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ (associative law).

- For scalars, the number 1 is the identity element for multiplication because it reproduces whatever it is multiplied by.
- For matrices, there is an identity element with similar properties.

Identity Matrix

The $n \times n$ matrix with 1's on the main diagonal and 0's elsewhere

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

is called the ***identity matrix*** of order n. For every $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A}\mathbf{I}_n = \mathbf{A} \quad \text{and} \quad \mathbf{I}_m\mathbf{A} = \mathbf{A}.$$

The subscript on $\mathbf{I}_n$ is neglected whenever the size is obvious from the context.

Analogous to scalar algebra, we define the 0^{th} power of a square matrix to be the identity matrix of corresponding size. That is, if $\mathbf{A}$ is $n \times n$, then

$$\mathbf{A}^0 = \mathbf{I}_n.$$

Positive powers of $\mathbf{A}$ are also defined in the natural way. That is,

$$\mathbf{A}^n = \underbrace{\mathbf{A}\mathbf{A}\cdots\mathbf{A}}_{n \text{ times}}.$$

The associative law guarantees that it makes no difference how matrices are grouped for powering. For example, $\mathbf{A}\mathbf{A}^2$ is the same as $\mathbf{A}^2\mathbf{A}$, so that

$$\mathbf{A}^3 = \mathbf{A}\mathbf{A}\mathbf{A} = \mathbf{A}\mathbf{A}^2 = \mathbf{A}^2\mathbf{A}.$$

Also, the usual laws of exponents hold. For nonnegative integers r and s,

$$\mathbf{A}^r\mathbf{A}^s = \mathbf{A}^{r+s} \quad \text{and} \quad (\mathbf{A}^r)^s = \mathbf{A}^{rs}.$$

We are not yet in a position to define negative or fractional powers, and due to the lack of conformability, powers of nonsquare matrices are never defined.

- The operation of transposition has an interesting effect upon a matrix product—a reversal of order occurs.

Reverse Order Law for Transposition

For conformable matrices $\mathbf{A}$ and $\mathbf{B}$,

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T.$$

The case of conjugate transposition is similar. That is,

$$(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*.$$

- For every matrix $\mathbf{A}_{m\times n}$, the products $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are symmetric matrices.
- For matrices $\mathbf{A}_{m\times n}$ and $\mathbf{B}_{n\times m}$, $trace(\mathbf{AB}) = trace(\mathbf{BA})$.
- Any product of conformable matrices can be permuted cyclically without altering the trace of the product.

$$trace(\mathbf{ABC}) = trace(\mathbf{BCA}) = trace(\mathbf{CAB}).$$

- However, a noncyclical permutation may not preserve the trace.

$$trace(\mathbf{ABC}) \neq trace(\mathbf{BAC}).$$

- Executing multiplication between two matrices by partitioning one or both factors into **submatrices**—a matrix contained within another matrix—can be a useful technique.

Block Matrix Multiplication

Suppose that $\mathbf{A}$ and $\mathbf{B}$ are partitioned into submatrices—often referred to as ***blocks***—as indicated below.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1r} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{s1} & \mathbf{A}_{s2} & \cdots & \mathbf{A}_{sr} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \cdots & \mathbf{B}_{1t} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \cdots & \mathbf{B}_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{r1} & \mathbf{B}_{r2} & \cdots & \mathbf{B}_{rt} \end{pmatrix}.$$

If the pairs $(\mathbf{A}_{ik}, \mathbf{B}_{kj})$ are conformable, then $\mathbf{A}$ and $\mathbf{B}$ are said to be ***conformably partitioned***. For such matrices, the product $\mathbf{AB}$ is formed by combining the blocks exactly the same way as the scalars are combined in ordinary matrix multiplication. That is, the (i, j)-block in $\mathbf{AB}$ is

$$\mathbf{A}_{i1}\mathbf{B}_{1j} + \mathbf{A}_{i2}\mathbf{B}_{2j} + \cdots + \mathbf{A}_{ir}\mathbf{B}_{rj}.$$

Block multiplication is particularly useful when there are patterns in the matrices to be multiplied. Consider the partitioned matrices

$$\mathbf{A} = \left(\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right) = \begin{pmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix}, \quad \mathbf{B} = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 2 & 1 & 2 \\ 3 & 4 & 3 & 4 \end{array}\right) = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{C} \end{pmatrix},$$

where

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Using block multiplication, the product $\mathbf{AB}$ is easily computed to be

$$\mathbf{AB} = \begin{pmatrix} \mathbf{C} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{C} \end{pmatrix} = \begin{pmatrix} 2\mathbf{C} & \mathbf{C} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} = \left(\begin{array}{cc|cc} 2 & 4 & 1 & 2 \\ 6 & 8 & 3 & 4 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{array}\right).$$

Matrix Inversion

- If α ia a nonzero scalar, then for each number β the equation $\alpha x = \beta$ has a unique solution given by $x = \alpha^{-1}\beta$.
- The properties $\alpha\alpha^{-1} = 1$ and $\alpha^{-1}\alpha = 1$ are the key ingredients.
- We want to solve matrix equations in the same fashion as we solve scalar equations.

Matrix Inversion

For a given square matrix $\mathbf{A}_{n\times n}$, the matrix $\mathbf{B}_{n\times n}$ that satisfies the conditions

$$\mathbf{AB} = \mathbf{I}_n \quad \text{and} \quad \mathbf{BA} = \mathbf{I}_n$$

is called the ***inverse*** of $\mathbf{A}$ and is denoted by $\mathbf{B} = \mathbf{A}^{-1}$. Not all square matrices are invertible—the zero matrix is a trivial example, but there are also many nonzero matrices that are not invertible. An invertible matrix is said to be ***nonsingular***, and a square matrix with no inverse is called a ***singular matrix***.

- Notice that matrix inversion is defined for square matrices only.

If

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \text{where} \quad \delta = ad - bc \neq 0,$$

then

$$\mathbf{A}^{-1} = \frac{1}{\delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

because it can be verified that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_2$.

- Although not all matrices are invertible, **when an inverse exists, it is unique.**
- Since matrix inversion was defined analogously to scalar inversion, so we have
  1. If $\mathbf{A}$ is a nonsingular matrix, then there is a unique solution for $\mathbf{X}$ in the matrix equation $\mathbf{A}_{n\times n}\mathbf{X}_{n\times p} = \mathbf{B}_{n\times p}$ and the solution is

     $$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}.$$

  2. A system of n linear equations in n unknowns can be written as a single matrix equation $\mathbf{A}_{n\times n}\mathbf{x}_{n\times 1} = \mathbf{b}_{n\times 1}$, so when $\mathbf{A}$ is nonsingular, the system has a unique solution given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

- It must be stressed that the representation of the solution as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ is mostly a notation or theoretical convenience.
- Since not all square matrices are invertible, methods are needed to distinguish between nonsingular and singular matrices.
- There is a variety of ways to describe be class of nonsingular matrices.

Existence of an Inverse

For an $n \times n$ matrix $\mathbf{A}$, the following statements are equivalent.

- $\mathbf{A}^{-1}$ exists ($\mathbf{A}$ is nonsingular).
- $rank(\mathbf{A}) = n$.
- $\mathbf{A} \xrightarrow{\text{Gauss–Jordan}} \mathbf{I}$.
- $\mathbf{Ax} = \mathbf{0}$ implies that $\mathbf{x} = \mathbf{0}$.

- The definition of matrix inversion says that in order to compute $\mathbf{A}^{-1}$, it is necessary to solve both of the matrix equations $\mathbf{AX} = \mathbf{I}$ and $\mathbf{XA} = \mathbf{I}$.

- These two equations are necessary to rule out the possibility of nonsquare inverses.
- But when only square matrices are involved, then any one of the two equations will suffice.
- If $\mathbf{A}$ and $\mathbf{X}$ are square matrices,

$$\mathbf{AX} = \mathbf{I} \Longrightarrow \mathbf{XA} = \mathbf{I}.$$

  In other words, if $\mathbf{A}$ and $\mathbf{X}$ are square and $\mathbf{AX} = \mathbf{I}$, then $\mathbf{X} = \mathbf{A}^{-1}$.
- Although we usually try to avoid computing the inverse of a matrix, there are times when an inverse must be found.

Computing an Inverse

Gauss–Jordan elimination can be used to invert $\mathbf{A}$ by the reduction

$$[\mathbf{A} \,|\, \mathbf{I}] \xrightarrow{\text{Gauss–Jordan}} [\mathbf{I} \,|\, \mathbf{A}^{-1}].$$

The only way for this reduction to fail is for a row of zeros to emerge in the left-hand side of the augmented array, and this occurs if and only if $\mathbf{A}$ is a singular matrix.

Problem: If possible, find the inverse of $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$.

Solution:

$$[\mathbf{A} \,|\, \mathbf{I}] = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|rrr} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{ccc|rrr} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|rrr} 1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 1 & 0 & -1 & 2 & -1 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{array}\right)$$

Therefore, the matrix is nonsingular, and $\mathbf{A}^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$. If we wish to check this answer, we need only check that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

- Earlier in this section it was stated that one almost never solves a nonsingular linear system $\mathbf{Ax} = \mathbf{b}$ by first computing $\mathbf{A}^{-1}$ and then the product $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
- To appreciate why this is true, pay attention to how much effort is required to perform one matrix inversion.

Operation Counts for Inversion

Computing $\mathbf{A}^{-1}_{n\times n}$ by reducing $[\mathbf{A}|\mathbf{I}]$ with Gauss–Jordan requires

- n^3 multiplications/divisions,
- $n^3 - 2n^2 + n$ additions/subtractions.

- Remarkably, it takes almost exactly as much effort to perform one matrix multiplication as to perform one matrix inversion.
- This fact always seems to be counter to a novice's intuition-it feels like matrix inversion should be a more difficult task than multiplication, but this is not the case.

Properties of Matrix Inversion

For nonsingular matrices $\mathbf{A}$ and $\mathbf{B}$, the following properties hold.

- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
- The product $\mathbf{AB}$ is also nonsingular.
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ (the reverse order law for inversion).
- $(\mathbf{A}^{-1})^{T} = (\mathbf{A}^{T})^{-1}$ and $(\mathbf{A}^{-1})^{*} = (\mathbf{A}^{*})^{-1}$.

- In general the product of two rank-r matrices does not necessarily have to produce another matrix of rank r.
- However, the product of two invertible matrices is again invertible. This generalizes to any number of matrices.

Products of Nonsingular Matrices Are Nonsingular

If $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are each $n \times n$ nonsingular matrices, then the product $\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k$ is also nonsingular, and its inverse is given by the reverse order law. That is,

$$(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_k)^{-1} = \mathbf{A}_k^{-1}\cdots\mathbf{A}_2^{-1}\mathbf{A}_1^{-1}.$$

Inverses of Sums and Sensitivity

- The reverse order law for inversion makes the inverse of a product easy to deal with, but the inverse of a sum is much more difficult.
- To begin with $(\mathbf{A}+\mathbf{B})^{-1}$ may not exist even if $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ each exist.
- Moreover, if $(\mathbf{A}+\mathbf{B})^{-1}$ exists, then, with rare exceptions, $(\mathbf{A}+\mathbf{B})^{-1} \neq \mathbf{A}^{-1}+\mathbf{B}^{-1}$.
- There is no useful general formula for $(\mathbf{A}+\mathbf{B})^{-1}$, but there are some special sums for which something can be said.
- One of the most easily inverted sums is $\mathbf{I}+\mathbf{c}\mathbf{d}^T$ in which $\mathbf{c}$ and $\mathbf{d}$ are $n \times 1$ nonzero columns such that $1+\mathbf{d}^T\mathbf{c} \neq 0$.
- It is straightforward to verify by direct multiplication that

$$(\mathbf{I}+\mathbf{c}\mathbf{d}^T)^{-1} = \mathbf{I} - \frac{\mathbf{c}\mathbf{d}^T}{1+\mathbf{d}^T\mathbf{c}}.$$

Sherman–Morrison Formula

- If $\mathbf{A}_{n\times n}$ is nonsingular and if $\mathbf{c}$ and $\mathbf{d}$ are $n \times 1$ columns such that $1 + \mathbf{d}^T\mathbf{A}^{-1}\mathbf{c} \neq 0$, then the sum $\mathbf{A} + \mathbf{c}\mathbf{d}^T$ is nonsingular, and

$$\left(\mathbf{A} + \mathbf{c}\mathbf{d}^T\right)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{c}\mathbf{d}^T\mathbf{A}^{-1}}{1 + \mathbf{d}^T\mathbf{A}^{-1}\mathbf{c}}.$$

- The *Sherman–Morrison–Woodbury formula* is a generalization. If $\mathbf{C}$ and $\mathbf{D}$ are $n \times k$ such that $(\mathbf{I} + \mathbf{D}^T\mathbf{A}^{-1}\mathbf{C})^{-1}$ exists, then

$$(\mathbf{A} + \mathbf{C}\mathbf{D}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{I} + \mathbf{D}^T\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{D}^T\mathbf{A}^{-1}.$$

- The sherman-Morrison-Woodbury formula can be verified with direct multiplication.
- Suppose $\mathbf{A}^{-1}$ is known from a previous calculation, but new one entry in $\mathbf{A}$ needs to be changed or updated—we need to add α to a_{ij}.
- Sherman—Morrison shows how the previously computed information in $\mathbf{A}^{-1}$ can be updated to produce the new inverse.

- Let $\mathbf{c} = \mathbf{e}_i$ and $\mathbf{d} = \alpha\mathbf{e}_j$, where $\mathbf{e}_i$ and $\mathbf{e}_j$ are the i^{th} and j^{th} unit columns, respectively.
- the matrix $\mathbf{c}\mathbf{d}^T$ has α in the (i,j)-position and zeros elsewhere so that

$$\mathbf{B} = \mathbf{A} + \mathbf{c}\mathbf{d}^T = \mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^T$$

  is the updaed matrix.
- According to the Sherman-Morrison formula,

$$\begin{aligned}\mathbf{B}^{-1} &= (\mathbf{A} + \alpha\mathbf{e}_i\mathbf{e}_j^T)^{-1} = \mathbf{A}^{-1} - \alpha\frac{\mathbf{A}^{-1}\mathbf{e}_i\mathbf{e}_j^T\mathbf{A}^{-1}}{1 + \alpha\mathbf{e}_j^T\mathbf{A}^{-1}\mathbf{e}_i} \\ &= \mathbf{A}^{-1} - \alpha\frac{[\mathbf{A}^{-1}]_{*i}[\mathbf{A}^{-1}]_{j*}}{1 + \alpha[\mathbf{A}^{-1}]_{ji}}.\end{aligned}$$

- This show how $\mathbf{A}^T$ changes when a_{ij} is perturbed, and it provides a useful algorithm for updating $\mathbf{A}^{-1}$.

Problem: Start with $\mathbf{A}$ and $\mathbf{A}^{-1}$ given below. Update $\mathbf{A}$ by adding 1 to a_{21}, and then use the Sherman–Morrison formula to update $\mathbf{A}^{-1}$:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}.$$

Solution: The updated matrix is

$$\mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} = \mathbf{A} + \mathbf{e}_2 \mathbf{e}_1^T.$$

Applying the Sherman–Morrison formula yields the updated inverse

$$\mathbf{B}^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{e}_2\mathbf{e}_1^T\mathbf{A}^{-1}}{1 + \mathbf{e}_1^T\mathbf{A}^{-1}\mathbf{e}_2} = \mathbf{A}^{-1} - \frac{[\mathbf{A}^{-1}]_{*2}[\mathbf{A}^{-1}]_{1*}}{1 + [\mathbf{A}^{-1}]_{12}}$$

$$= \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} - \frac{\begin{pmatrix} -2 \\ 1 \end{pmatrix} \begin{pmatrix} 3 & -2 \end{pmatrix}}{1 - 2} = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}.$$

- Another sum that often require inversion is $\mathbf{I} - \mathbf{A}$, but we have to careful because $(\mathbf{I} - \mathbf{A})^{-1}$ need not always exist.
- However, we are safe when the entries in $\mathbf{A}$ are sufficiently small.

Neumann Series

If $\lim_{n\to\infty} \mathbf{A}^n = \mathbf{0}$, then $\mathbf{I} - \mathbf{A}$ is nonsingular and

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \cdots = \sum_{k=0}^{\infty} \mathbf{A}^k.$$

This is the ***Neumann series***. It provides approximations of $(\mathbf{I} - \mathbf{A})^{-1}$ when $\mathbf{A}$ has entries of small magnitude. For example, a first-order approximation is $(\mathbf{I} - \mathbf{A})^{-1} \approx \mathbf{I} + \mathbf{A}$.

- While there is no useful formula for $(\mathbf{A} + \mathbf{B})^{-1}$ in general, the Neumann series allows us to say something when $\mathbf{B}$ has small entries relative to $\mathbf{A}$, or vice versa.
- For example, if $\mathbf{A}^{-1}$ exists, and if the entries in $\mathbf{B}$ are small enough in magnitude to insure $\lim_{n\to\infty}(\mathbf{A}^{-1}\mathbf{B})^n = 0$.

$$
\begin{aligned}
(\mathbf{A}+\mathbf{B})^{-1} &= \left(\mathbf{A}(\mathbf{I}-[-\mathbf{A}^{-1}\mathbf{B}])\right)^{-1} = \left(\mathbf{I}-[-\mathbf{A}^{-1}\mathbf{B}]\right)^{-1}\mathbf{A}^{-1} \\
&= \left(\sum_{k=0}^{\infty}[-\mathbf{A}^{-1}\mathbf{B}]^{k}\right)\mathbf{A}^{-1}
\end{aligned}
$$

and a first-order approximation is

$$
(\mathbf{A}+\mathbf{B})^{-1} \approx \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}.
$$

- Consequently, if $\mathbf{A}$ is perturbed by a small matrix $\mathbf{B}$, possibly resulting from errors due to inexact measurements or perhaps form roundoff error, then the resulting change in $\mathbf{A}^{-1}$ is about $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$.
- In other words, the effect of a small perturbation (or error) $\mathbf{B}$ is magnified by multiplication (on both sides) with $\mathbf{A}^{-1}$.

- So if $\mathbf{A}^{-1}$ has large entries, small perturbations (or errors) in $\mathbf{A}$ can produce large perturbations (or errors) in the resulting inverse.

Sensitivity and Conditioning

- A nonsingular matrix $\mathbf{A}$ is said to be ***ill conditioned*** if a small relative change in $\mathbf{A}$ can cause a large relative change in $\mathbf{A}^{-1}$. The degree of ill-conditioning is gauged by a ***condition number*** $\kappa = \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\|$, where $\|\star\|$ is a matrix norm.
- The sensitivity of the solution of $\mathbf{Ax} = \mathbf{b}$ to perturbations (or errors) in $\mathbf{A}$ is measured by the extent to which $\mathbf{A}$ is an ill-conditioned matrix.

- If a nonsingular system $\mathbf{Ax} = \mathbf{b}$ are slightly perturbed to produced the system $(\mathbf{A}+\mathbf{B})\widetilde{\mathbf{x}} = \mathbf{b}$, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ and $\widetilde{\mathbf{x}} = (\mathbf{A}+\mathbf{B})^{-1}\mathbf{b}$ so that

$$\mathbf{x} - \widetilde{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{b} - (\mathbf{A}+\mathbf{B})^{-1}\mathbf{b} \approx \mathbf{A}^{-1}\mathbf{Bx}.$$

- $\|\mathbf{x} - \widetilde{\mathbf{x}}\| \lesssim \|\mathbf{A}^{-1}\|\|\mathbf{B}\|\|\mathbf{x}\|$,

$$\frac{\|\mathbf{x} - \widetilde{\mathbf{x}}\|}{\|\mathbf{x}\|} \lesssim \|\mathbf{A}^{-1}\|\|\mathbf{B}\| = \kappa \left\{ \frac{\|\mathbf{B}\|}{\|\mathbf{A}\|} \right\}.$$

- The condition number κ is small, a small relative change in $\mathbf{A}$ can not produce a large relative change in $\mathbf{x}$, but for large values of κ, a small relative change in $\mathbf{A}$ can possibly result in a large relative change in $\mathbf{x}$.
- Consider the system

$$.835x + .667y = .168,$$

$$.333x + .266y = .067,$$

- $\mathbf{A} = \begin{pmatrix} .835 & .667 \\ .333 & .266 \end{pmatrix}$ and $\mathbf{A}^{-1} = \begin{pmatrix} -266000 & 667000 \\ 333000 & -835000 \end{pmatrix}$.
- The condition number for $\mathbf{A}$ is $\kappa = 1.7 \times 10^6$.
- **A Rule of Thumb** If Gaussian elimination with partial pivoting is used to solve a nonsingular system $\mathbf{Ax} = \mathbf{b}$ using t-digit floating-point arithmetic, then, assuming no other source of error exists, it can be argued that when κ is of order 10^p, the computed solution is expected to be accurate to at least $t - p$ significant digits, more or less.

Elementary Matrices and Equivalence

- A common theme in mathematics is to break complicated objects into more elementary components.
- Factor large polynomials into products of smaller polynomials.
- This section is to lay the groundwork for similar ideas in matrix algebra by considering how a general matrix might be factored into a product of more elementary matrices.
- Matrices of the form $\mathbf{I} - \mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ columns such that $\mathbf{v}^T\mathbf{u} \neq 1$ are called **elementary matrices.**
- All such matrices are nonsingular and

$$(\mathbf{I} - \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{I} - \frac{\mathbf{u}\mathbf{v}^T}{\mathbf{v}^T\mathbf{u} - 1}.$$

- Notice that inverses of elementary matrices are elementary matrices.

- We are primarily interested in the elementary matrices associated with the three elementary row (or column) operations hereafter referred to as follows.
  - Type I is the interchanging rows (columns) i and j;
  - Type II is multiplying row (column) i by $\alpha \neq 0$;
  - Type III is adding a multiple of row (column) i to row (column) j.
- An elementary matrix of Type I, II, or III is created by performing an elementary operation of Type I, II, or III to an identity matrix.
- For example

$$\mathbf{E}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix},$$

- $\mathbf{E}_1 = \mathbf{I} - \mathbf{u}\mathbf{u}^T$, where $\mathbf{u} = \mathbf{e}_1 - \mathbf{e}_2$,
- $\mathbf{E}_2 = \mathbf{I} - (1-\alpha)\mathbf{e}_2\mathbf{e}_2^T$ and $\mathbf{E}_3 = \mathbf{I} + \alpha\mathbf{e}_3\mathbf{e}_1^T$.

Properties of Elementary Matrices

- When used as a *left-hand* multiplier, an elementary matrix of Type I, II, or III executes the corresponding *row* operation.
- When used as a *right-hand* multiplier, an elementary matrix of Type I, II, or III executes the corresponding *column* operation.

The sequence of row operations used to reduce $\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 13 \end{pmatrix}$ to $\mathbf{E_A}$ is indicated below.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \\ 3 & 6 & 13 \end{pmatrix} \begin{matrix} R_2 - 2R_1 \\ R_3 - 3R_1 \end{matrix} \longrightarrow \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{\textit{Interchange } R_2 \textit{ and } R_3} \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \; R_1 - 4R_2 \longrightarrow \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \mathbf{E_A}.$$

The reduction can be accomplished by a sequence of left-hand multiplications with the corresponding elementary matrices as shown below.

$$\begin{pmatrix} 1 & -4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{A} = \mathbf{E_A}.$$

- $\mathbf{A}$ is a nonsingular matrix if and only if $\mathbf{A}$ is the product of elementary matrices of Type I, II, or III.

Proof. If $\mathbf{A}$ is nonsingular, then the Gauss–Jordan technique reduces $\mathbf{A}$ to $\mathbf{I}$ by row operations. If $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k$ is the sequence of elementary matrices that corresponds to the elementary row operations used, then

$$\mathbf{G}_k \cdots \mathbf{G}_2\mathbf{G}_1\mathbf{A} = \mathbf{I} \text{ or, equivalently, } \mathbf{A} = \mathbf{G}_1^{-1}\mathbf{G}_2^{-1} \cdots \mathbf{G}_k^{-1}.$$

Since the inverse of an elementary matrix is again an elementary matrix of the same type, this proves that $\mathbf{A}$ is the product of elementary matrices of Type I, II, or III. Conversely, if $\mathbf{A} = \mathbf{E}_1\mathbf{E}_2 \cdots \mathbf{E}_k$ is a product of elementary matrices, then $\mathbf{A}$ must be nonsingular because the $\mathbf{E}_i$'s are nonsingular, and a product of nonsingular matrices is also nonsingular. ∎

Equivalence

- Whenever $\mathbf{B}$ can be derived from $\mathbf{A}$ by a combination of elementary row and column operations, we write $\mathbf{A} \sim \mathbf{B}$, and we say that $\mathbf{A}$ and $\mathbf{B}$ are ***equivalent matrices***. Since elementary row and column operations are left-hand and right-hand multiplication by elementary matrices, respectively, we can say that

$$\mathbf{A} \sim \mathbf{B} \iff \mathbf{PAQ} = \mathbf{B} \quad \text{for nonsingular } \mathbf{P} \text{ and } \mathbf{Q}.$$

- Whenever $\mathbf{B}$ can be obtained from $\mathbf{A}$ by performing a sequence of elementary *row* operations only, we write $\mathbf{A} \overset{\text{row}}{\sim} \mathbf{B}$, and we say that $\mathbf{A}$ and $\mathbf{B}$ are ***row equivalent***. In other words,

$$\mathbf{A} \overset{\text{row}}{\sim} \mathbf{B} \Longleftrightarrow \mathbf{PA} = \mathbf{B} \quad \text{for a nonsingular } \mathbf{P}.$$

- Whenever $\mathbf{B}$ can be obtained from $\mathbf{A}$ by performing a sequence of *column* operations only, we write $\mathbf{A} \overset{\text{col}}{\sim} \mathbf{B}$, and we say that $\mathbf{A}$ and $\mathbf{B}$ are ***column equivalent***. In other words,

$$\mathbf{A} \overset{\text{col}}{\sim} \mathbf{B} \Longleftrightarrow \mathbf{AQ} = \mathbf{B} \quad \text{for a nonsingular } \mathbf{Q}.$$

- If it's possible to go from $\mathbf{A}$ to $\mathbf{B}$ by elementary row and column operations, then clearly it's possible to start with $\mathbf{B}$ and get back to $\mathbf{A}$ because elementary operations are reversible.
- It therefore makes sense to talk about the equivalence of a pair of matrices without regard to order. $\mathbf{A} \sim \mathbf{B} \Longleftrightarrow \mathbf{B} \sim \mathbf{A}$.
- Each type of equivalence is transitive in the sense that

$$\mathbf{A} \sim \mathbf{B} \quad \textit{and} \quad \mathbf{B} \sim \mathbf{C} \Longrightarrow \mathbf{A} \sim \mathbf{C}.$$

Rank Normal Form

If $\mathbf{A}$ is an $m \times n$ matrix such that $rank(\mathbf{A}) = r$, then

$$\mathbf{A} \sim \mathbf{N}_r = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

$\mathbf{N}_r$ is called the ***rank normal form*** for $\mathbf{A}$, and it is the end product of a complete reduction of $\mathbf{A}$ by using both row and column operations.

- Given matrices $\mathbf{A}$ and $\mathbf{B}$, how do we decide whether or not $\mathbf{A} \sim \mathbf{B}$?
  - We can use a trial-and-error approach by attempting to reduce $\mathbf{A}$ to $\mathbf{B}$ by elementary operations.
  - $\mathbf{A} \sim \mathbf{B}$ if and only if $rank(\mathbf{A}) = rank(\mathbf{B})$.
  - **Corollary:** Multiplication by nonsingular matrices cannot change rank.
- Transposition does not change the rank—i.e. for all $m \times n$ matrices,

$$rank(\mathbf{A}) = rank(\mathbf{A}^T) \quad rank(\mathbf{A}) = rank(\mathbf{A}^*).$$

LU Factorization

- Now, we are back solving a nonsingular system of linear equations using Gaussian elimination with back substitution.
- This time, however, the goal is to describe and understand the process in the context of matrices.
- If $\mathbf{A}\mathbf{x} = \mathbf{b}$ is a nonsingular system, then the object of Gaussian elimination is to reduce $\mathbf{A}$ to an upper-triangular matrix using elementary row operations.
- If no zero pivots are encountered, then row interchanges are not necessary.
- The reduction can be accomplished by using only elementary row operations of Type III.
- For example, consider the reducing the matrix

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 7 & 7 \\ 6 & 18 & 22 \end{pmatrix}$$

to upper-triangular form as shown:

$$\begin{pmatrix} 2 & 2 & 2 \\ 4 & 7 & 7 \\ 6 & 18 & 22 \end{pmatrix} \begin{matrix} R_2 - 2R_1 \\ R_3 - 3R_1 \end{matrix} \longrightarrow \begin{pmatrix} 2 & 2 & 2 \\ 0 & 3 & 3 \\ 0 & 12 & 16 \end{pmatrix} R_3 - 4R_2$$

$$\longrightarrow \begin{pmatrix} 2 & 2 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 4 \end{pmatrix} = \mathbf{U}.$$

We learned in the previous section that each of these Type III operations can be executed by means of a left-hand multiplication with the corresponding elementary matrix $\mathbf{G}_i$, and the product of all of these $\mathbf{G}_i$'s is

$$\mathbf{G}_3\mathbf{G}_2\mathbf{G}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & -4 & 1 \end{pmatrix}.$$

In other words, $\mathbf{G}_3\mathbf{G}_2\mathbf{G}_1\mathbf{A} = \mathbf{U}$, so that $\mathbf{A} = \mathbf{G}_1^{-1}\mathbf{G}_2^{-1}\mathbf{G}_3^{-1}\mathbf{U} = \mathbf{L}\mathbf{U}$, where $\mathbf{L}$ is the lower-triangular matrix

$$\mathbf{L} = \mathbf{G}_1^{-1}\mathbf{G}_2^{-1}\mathbf{G}_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{pmatrix}.$$

Thus $\mathbf{A} = \mathbf{L}\mathbf{U}$ is a product of a lower-triangular matrix $\mathbf{L}$ and an upper-triangular matrix $\mathbf{U}$. Naturally, this is called an $\mathbf{LU}$ ***factorization*** of $\mathbf{A}$.

- Observe the $\mathbf{U}$ is the end product of Gaussian elimination and has the pivots on its diagonal on its diagonal, while $\mathbf{L}$ has 1's on its diagonal.
- $\mathbf{L}$ has the remarkable property that below its diagonal, each entry l_{ij} is precisely the multiplier used in the elimination to annihilate the (i,j)-position.

LU Factorization

If $\mathbf{A}$ is an $n \times n$ matrix such that a zero pivot is never encountered when applying Gaussian elimination with Type III operations, then $\mathbf{A}$ can be factored as the product $\mathbf{A} = \mathbf{LU}$, where the following hold.

- $\mathbf{L}$ is lower triangular and $\mathbf{U}$ is upper triangular.
- $\ell_{ii} = 1$ and $u_{ii} \neq 0$ for each $i = 1, 2, \ldots, n$.
- Below the diagonal of $\mathbf{L}$, the entry ℓ_{ij} is the multiple of row j that is subtracted from row i in order to annihilate the (i, j)-position during Gaussian elimination.
- $\mathbf{U}$ is the final result of Gaussian elimination applied to $\mathbf{A}$.

- The decomposition of $\mathbf{A}$ into $\mathbf{A} = \mathbf{LU}$ is called the $\mathbf{LU}$ **factorization of** $\mathbf{A}$, and the matrices $\mathbf{L}$ and $\mathbf{U}$ are called the $\mathbf{LU}$ factors of $\mathbf{A}$.

Once the LU factors for a nonsingular matrix $\mathbf{A}_{n\times n}$ have been obtained, it's relatively easy to solve a linear system $\mathbf{Ax} = \mathbf{b}$. By rewriting $\mathbf{Ax} = \mathbf{b}$ as

$$\mathbf{L}(\mathbf{Ux}) = \mathbf{b} \quad \text{and setting} \quad \mathbf{y} = \mathbf{Ux},$$

we see that $\mathbf{Ax} = \mathbf{b}$ is equivalent to the two triangular systems

$$\mathbf{Ly} = \mathbf{b} \quad \text{and} \quad \mathbf{Ux} = \mathbf{y}.$$

First, the lower-triangular system $\mathbf{Ly} = \mathbf{b}$ is solved for $\mathbf{y}$ by ***forward substitution.*** That is, if

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \ell_{21} & 1 & 0 & \cdots & 0 \\ \ell_{31} & \ell_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{n1} & \ell_{n2} & \ell_{n3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix},$$

set

$$y_1 = b_1, \quad y_2 = b_2 - \ell_{21}y_1, \quad y_3 = b_3 - \ell_{31}y_1 - \ell_{32}y_2, \quad \text{etc.}$$

The forward substitution algorithm can be written more concisely as

$$y_1 = b_1 \quad \text{and} \quad y_i = b_i - \sum_{k=1}^{i-1} \ell_{ik}y_k \quad \text{for} \quad i = 2, 3, \ldots, n.$$

After $\mathbf{y}$ is known, the upper-triangular system $\mathbf{Ux} = \mathbf{y}$ is solved using the standard back substitution procedure by starting with $x_n = y_n/u_{nn}$, and setting

$$x_i = \frac{1}{u_{ii}}\left(y_i - \sum_{k=i+1}^{n} u_{ik}x_k\right) \quad \text{for} \quad i = n-1,\ n-2,\ \ldots, 1.$$

- Problem: Use the LU factorization of $\mathbf{A}$ to solve $\mathbf{Ax} = \mathbf{b}$ where

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & 2 \\ 4 & 7 & 7 \\ 6 & 18 & 22 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 12 \\ 24 \\ 12 \end{pmatrix}.$$

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{U} = \begin{pmatrix} 2 & 2 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 4 \end{pmatrix}.$$

The strategy is to set $\mathbf{Ux} = \mathbf{y}$ and solve $\mathbf{Ax} = \mathbf{L}(\mathbf{Ux}) = \mathbf{b}$ by solving the two triangular systems

$$\mathbf{Ly} = \mathbf{b} \quad \text{and} \quad \mathbf{Ux} = \mathbf{y}.$$

First solve the lower-triangular system $\mathbf{Ly} = \mathbf{b}$ by using forward substitution:

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 24 \\ 12 \end{pmatrix} \implies \begin{aligned} y_1 &= 12, \\ y_2 &= 24 - 2y_1 = 0, \\ y_3 &= 12 - 3y_1 - 4y_2 = -24. \end{aligned}$$

Now use back substitution to solve the upper-triangular system $\mathbf{Ux} = \mathbf{y}$:

$$\begin{pmatrix} 2 & 2 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 0 \\ -24 \end{pmatrix} \implies \begin{aligned} x_1 &= (12 - 2x_2 - 2x_3)/2 = 6, \\ x_2 &= (0 - 3x_3)/3 = 6, \\ x_3 &= -24/4 = -6. \end{aligned}$$

- If only one system $\mathbf{Ax} = \mathbf{b}$ is to be solved, then there is no significant difference between the technique of reducing the augmented matrix $[\mathbf{A}|\mathbf{b}]$to a row echelon form and the LU factorization method.
- However, suppose it becomes necessary to later solve other systems $\mathbf{Ax} = \widetilde{\mathbf{b}}$ with the same coefficient matrix but with different right-hand sides, which is frequently the case in applied work.
- If the LU factors of A were computed and saved when the original system was solved, then they need not be recomputed, and the solutions to all subsequent systems $\mathbf{Ax} = \widetilde{\mathbf{b}}$ are therefore relatively cheap to obtain.
- **Summary**
  - To solve a nonsingular system $\mathbf{Ax} = \mathbf{b}$ using the LU factorization $\mathbf{A} = \mathbf{LU}$, first solve $\mathbf{Ly} = \mathbf{b}$ for $\mathbf{y}$ with the forward substitution algorithm and solve $\mathbf{Ux} = \mathbf{y}$ for $\mathbf{x}$ with the back substitution procedure.
  - The advantage of this approach is that once the LU factors for $\mathbf{A}$ have been computed, any other linear system $\mathbf{Ax} = \widetilde{\mathbf{b}}$ can be solved with only n^2 multiplications/divisions and $n^2 - n$ additions/ subtractions.

- Not all nonsingular matrices posses an LU factorization.
- For example, there is clearly no nonzero value of u_{11} that will satisfy

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{pmatrix}$$

  The problem is the zero pivot in the (1,1)-position.
- A nonsingular matrix $\mathbf{A}$ has an LU factorization if and only if a zero pivot does not emerge during row reduction to upper-triangular form with Type III operations.
- There is another interesting way to characterize the existence of LU factors.This characterization is given in terms of the leading principal submatrices of $\mathbf{A}$ that are defined to be those submatrices taken from the upper-left-hand corner of $\mathbf{A}$. That is,

$$\mathbf{A}_1 = \begin{pmatrix} a_{11} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \ldots, \mathbf{A}_k = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}, \ldots.$$

Existence of LU Factors

Each of the following statements is equivalent to saying that a nonsingular matrix $\mathbf{A}_{n\times n}$ possesses an LU factorization.

- A zero pivot does not emerge during row reduction to upper-triangular form with Type III operations.
- Each leading principal submatrix $\mathbf{A}_k$ is nonsingular.

- Up to this point we have avoided dealing with row interchanges because if a row interchange is needed to remove a zero pivot, then no LU factorization is possible.
- However, we know that practical computation necessitates row interchanges in the form of partial pivoting.
- So even if no zero pivots emerge, it is usually the case that we must still somehow account for row interchanges.
- When row interchanges are allowed, zero pivots can always be avoided when the original matrix $\mathbf{A}$ is nonsingular.

- Consequently, we may conclude that for every nonsingular matrix $\mathbf{A}$, there exists a permutation matrix $\mathbf{P}$ (a product of elementary interchange matrices) such that $\mathbf{PA}$ has an LU factorization.
- This means that we can proceed just as in the case when no interchanges are used and successively overwrite the array originally containing A with each multiplier replacing the position it annihilates.
- Whenever a row interchange occurs, the corresponding multipliers will be correctly interchanged as well.
- The permutation matrix $\mathbf{P}$ is simply the cumulative record of the various interchanges used, and the information in $\mathbf{P}$ is easily accounted for by a simple technique that is illustrated in the following example.

Problem: Use partial pivoting on the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -3 & 4 \\ 4 & 8 & 12 & -8 \\ 2 & 3 & 2 & 1 \\ -3 & -1 & 1 & -4 \end{pmatrix}$$

and determine the LU decomposition $\mathbf{PA} = \mathbf{LU}$, where $\mathbf{P}$ is the associated permutation matrix.

$$[\mathbf{A}|\mathbf{p}] = \left(\begin{array}{cccc|c} 1 & 2 & -3 & 4 & 1 \\ 4 & 8 & 12 & -8 & 2 \\ 2 & 3 & 2 & 1 & 3 \\ -3 & -1 & 1 & -4 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 4 & 8 & 12 & -8 & 2 \\ 1 & 2 & -3 & 4 & 1 \\ 2 & 3 & 2 & 1 & 3 \\ -3 & -1 & 1 & -4 & 4 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|c} 4 & 8 & 12 & -8 & 2 \\ \mathbf{1/4} & 0 & -6 & 6 & 1 \\ \mathbf{1/2} & -1 & -4 & 5 & 3 \\ \mathbf{-3/4} & 5 & 10 & -10 & 4 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 4 & 8 & 12 & -8 & 2 \\ \mathbf{-3/4} & 5 & 10 & -10 & 4 \\ \mathbf{1/2} & -1 & -4 & 5 & 3 \\ \mathbf{1/4} & 0 & -6 & 6 & 1 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|c} 4 & 8 & 12 & -8 & 2 \\ \mathbf{-3/4} & 5 & 10 & -10 & 4 \\ \mathbf{1/2} & \mathbf{-1/5} & -2 & 3 & 3 \\ \mathbf{1/4} & \mathbf{0} & -6 & 6 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 4 & 8 & 12 & -8 & 2 \\ \mathbf{-3/4} & 5 & 10 & -10 & 4 \\ \mathbf{1/4} & \mathbf{0} & -6 & 6 & 1 \\ \mathbf{1/2} & \mathbf{-1/5} & -2 & 3 & 3 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|c} 4 & 8 & 12 & -8 & 2 \\ \mathbf{-3/4} & 5 & 10 & -10 & 4 \\ \mathbf{1/4} & \mathbf{0} & -6 & 6 & 1 \\ \mathbf{1/2} & \mathbf{-1/5} & \mathbf{1/3} & 1 & 3 \end{array}\right).$$

Therefore,

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -3/4 & 1 & 0 & 0 \\ 1/4 & 0 & 1 & 0 \\ 1/2 & -1/5 & 1/3 & 1 \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} 4 & 8 & 12 & -8 \\ 0 & 5 & 10 & -10 \\ 0 & 0 & -6 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

LU Factorization with Row Interchanges

- For each nonsingular matrix $\mathbf{A}$, there exists a permutation matrix $\mathbf{P}$ such that $\mathbf{PA}$ possesses an LU factorization $\mathbf{PA} = \mathbf{LU}$.

- To compute $\mathbf{L}$, $\mathbf{U}$, and $\mathbf{P}$, successively overwrite the array originally containing $\mathbf{A}$. Replace each entry being annihilated with the multiplier used to execute the annihilation. Whenever row interchanges such as those used in partial pivoting are implemented, the multipliers in the array will automatically be interchanged in the correct manner.

- Although the entire permutation matrix $\mathbf{P}$ is rarely called for, it can be constructed by permuting the rows of the identity matrix $\mathbf{I}$ according to the various interchanges used. These interchanges can be accumulated in a "permutation counter column" $\mathbf{p}$ that is initially in natural order $(1, 2, \ldots, n)$

- To solve a nonsingular linear system $\mathbf{Ax} = \mathbf{b}$ using the LU decomposition with partial pivoting, permute the components in $\mathbf{b}$ to construct $\tilde{\mathbf{b}}$ according to the sequence of interchanges used—i.e., according to $\mathbf{p}$—and then solve $\mathbf{Ly} = \tilde{\mathbf{b}}$ by forward substitution

- It is easy to combine the advantages of partial pivoting with the LU decomposition in order to solve a nonsingular system $\mathbf{Ax} = \mathbf{b}$.
- Because permutation matrices are nonsingular, the system $\mathbf{Ax} = \mathbf{b}$ is equivalent to

$$\mathbf{PAx} = \mathbf{Pb}.$$

- Hence we can employ the LU solution techniques discussed earlier to solve this permuted system.
- That is, if we have already performed the factorization $\mathbf{PA} = \mathbf{LU}$, then we can solve $\mathbf{Ly} = \mathbf{Pb}$ for $\mathbf{y}$ by forward substitution, and then solve $\mathbf{Ux} = \mathbf{y}$ by back substitution.

Problem: Use the LU decomposition obtained with partial pivoting to solve the system $\mathbf{Ax} = \mathbf{b}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & -3 & 4 \\ 4 & 8 & 12 & -8 \\ 2 & 3 & 2 & 1 \\ -3 & -1 & 1 & -4 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 60 \\ 1 \\ 5 \end{pmatrix}.$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -3/4 & 1 & 0 & 0 \\ 1/4 & 0 & 1 & 0 \\ 1/2 & -1/5 & 1/3 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 60 \\ 5 \\ 3 \\ 1 \end{pmatrix} \implies \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 60 \\ 50 \\ -12 \\ -15 \end{pmatrix}.$$

Then solve $\mathbf{Ux} = \mathbf{y}$ by applying back substitution:

$$\begin{pmatrix} 4 & 8 & 12 & -8 \\ 0 & 5 & 10 & -10 \\ 0 & 0 & -6 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 60 \\ 50 \\ -12 \\ -15 \end{pmatrix} \implies \mathbf{x} = \begin{pmatrix} 12 \\ 6 \\ -13 \\ -15 \end{pmatrix}.$$

The LDU factorization

- There’ s some asymmetry in an LU factorization because the lower factor has 1’ s on its diagonal while the upper factor has a nonunit diagonal.
- This is easily remedied by factoring the diagonal entries out of the upper factor as shown below:

$$\begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix} = \begin{pmatrix} u_{11} & 0 & \cdots & 0 \\ 0 & u_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{pmatrix} \begin{pmatrix} 1 & u_{12}/u_{11} & \cdots & u_{1n}/u_{11} \\ 0 & 1 & \cdots & u_{2n}/u_{22} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

- Setting $\mathbf{D} = diag(u_{11}, u_{22}, \cdots, u_{nn})$ (the diagonal matrix of pivots) and redefining $\mathbf{U}$ to be the rightmost upper-triangular matrix shown above allows any $\mathbf{LU}$ factorization to be written as

$$\mathbf{A} = \mathbf{LDU}.$$

- Where $\mathbf{L}$ and $\mathbf{U}$ are lower- and uppertriangular matrices with 1' s on both of their diagonals.
- This is called the **LDU factorization** of $\mathbf{A}$.
- It is uniquely determined, and when $\mathbf{A}$ is symmetric, the LDU factorization is $\mathbf{A} = \mathbf{LDL}^T$.

Exercises

1. Construct an example of a 3×3 matrix $\mathbf{A}$ that satisfies the following conditions.
   (a) $\mathbf{A}$ is both symmetric and skew symmetric.
   (b) $\mathbf{A}$ is both hermitian and symmetric.
   (c) $\mathbf{A}$ is skew hermitian.
2. If $\mathbf{A}$ and $\mathbf{B}$ are two matrices of the same shape, prove that each of the following statements is true.
   (a) $(\mathbf{A}+\mathbf{B})^* = \mathbf{A}^* + \mathbf{B}^*$.
   (b) $(\alpha\mathbf{A})^* = \overline{\alpha}\mathbf{A}^*$.
3. Prove that each of the following statements is true.
   (a) If $\mathbf{A} = [a_{ij}]$ is skew symmetric, then $a_{jj} = 0$ for each j.
   (b) If $\mathbf{A}$ is real and symmetric, then $\mathbf{B} = i\mathbf{A}$ is skew hermitian.
4. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices. If $\mathbf{A}x = \mathbf{B}x$ holds for all $n \times 1$ columns x, prove that $\mathbf{A} = \mathbf{B}$.
5. For every matrix $\mathbf{A}_{m\times n}$, demonstrate that the products $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ are hermitian matrices.

6. Prove $trace(\mathbf{ABC}) = trace(\mathbf{BCA})$.
7. If $\mathbf{A}$ is nonsingular and symmetric, prove that $\mathbf{A}^{-1}$ is symmetric.
8. For a square matrix $\mathbf{A}$, explain why each of the following statements must be true
   (a) If $\mathbf{A}$ contains a zero row or a zero column, then $\mathbf{A}$ is singular.
   (b) If $\mathbf{A}$ contains two identical rows or two identical columns, then $\mathbf{A}$ is singular.
9. Suppose $\mathbf{A} = \begin{pmatrix} 2 & 0 & -1 \\ -1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix}$ and $\mathbf{A}^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 0 & 2 \end{pmatrix}$.
   (a) Use the Sherman-Morrison formula to determine the inverse of the matrix $\mathbf{B}$ that is obtained by changing the (3,2)-entry in $\mathbf{A}$ from 0 to 2.
   (b) Let $\mathbf{C}$ be the matrix that agrees with $\mathbf{A}$ except that $c_{32} = 2$ and $c_{33} = 2$. Use the Sherman-Morrison formula to find $\mathbf{C}^{-1}$.

10. Let $\mathbf{A} = \begin{pmatrix} 1 & 4 & 5 \\ 4 & 18 & 26 \\ 3 & 16 & 30 \end{pmatrix}$.
    (a) Determine the LU factors of $\mathbf{A}$.
    (b) Use the LU factors to solve $\mathbf{A}\mathbf{x}_1 = \mathbf{b}_1$ as well as $\mathbf{A}\mathbf{x}_2 = \mathbf{b}_2$, where $\mathbf{b}_1 = (6, 0, -6)^T$ and $\mathbf{b}_2 = (6, 6, 12)^T$.
11. Let $\mathbf{A}$ and $\mathbf{b}$ be the matrices

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 & 17 \\ 3 & 6 & -12 & 3 \\ 2 & 3 & -3 & 2 \\ 0 & 2 & -2 & 6 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 17 \\ 3 \\ 3 \\ 4 \end{pmatrix}.$$

    (a) Use partial pivoting and find the permuation matrix $\mathbf{P}$ such that $\mathbf{PA} = \mathbf{LU}$.
    (b) Use the information in $\mathbf{P}$, $\mathbf{L}$, $\mathbf{U}$ to soleve $\mathbf{A}\mathbf{x} = \mathbf{b}$.