Shu Lu *and* Quoc Tran-Dinh

# Introduction to Optimization

## From Linear Programming to Nonlinear Programming

August 22, 2019

STOR-UNC-Chapel Hill

## Chapter 2

# Overview of Linear Algebra

In this chapter, we review some concepts and tools from linear algebra that will be used in this course. We focus on the definitions and properties of vectors and matrices, linear systems, matrix inverses, elementary row operations, and the Gauss-Jordan method.

## 2.1 Vectors and vector operations

### *2.1.1 Definitions and examples*

An $n$-dimensional row vector is an array of numbers arranged in one row as

$$x = (x_1, x_2, \cdots, x_n).$$

For each $i = 1, \cdots, n$, $x_i$ is the $i$th entry (or element, component, coordinate) of the row vector $x$. Alternatively, an $n$-dimensional column vector is an array of numbers arranged in one column as

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

The notation $\mathbb{R}^n$ stands for the space of all $n$-dimensional column vectors (or row vectors, depending on the context). For example, $x = (2, 3, -1)$ is a row vector in $\mathbb{R}^3$, and $x = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ is a column vector in $\mathbb{R}^2$.

The transpose operator, denoted by $(\cdot)^T$, is used to transform a row vector into a column vector, and vice versa. In other words, if $x$ is an $n$-dimensional row (or column) vector, then $x^T$ is the $n$-dimensional column (or row) vector with the same elements placed in the same order.

Vectors we use in this course will mostly be column vectors. We often write an $n$-dimensional column vector as $x = (x_1, \cdots, x_n)^T$, the transpose of a row vector.

The vector of all zeros is called the **zero vector**. That is

$$0 = (0, 0, \cdots, 0)^T.$$

The vector of all ones is denoted by

$$1 = (1, 1, \cdots, 1)^T.$$

The vector

$$\mathbf{e}_i = (0, 0, \cdots, 0, \underbrace{1}_{\text{the } i\text{-th entry}}, 0, \cdots, 0)^T$$

is called the $i$-th unit vector of $\mathbb{R}^n$. The space $\mathbb{R}^n$ has $n$ unit vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$. For example, $\mathbb{R}^3$ has three unit vectors $\mathbf{e}_1 = (1, 0, 0)^T$, $\mathbf{e}_2 = (0, 1, 0)^T$ and $\mathbf{e}_3 = (0, 0, 1)^T$.

### *2.1.2 Basic operations*

Given two $n$-dimensional column vectors $x, y \in \mathbb{R}^n$, and a number $c \in \mathbb{R}$, the following operations can be defined.

**Addition.** The sum of $x$ and $y$, denoted by $x + y$, is a vector $z = (z_1, \cdots, z_n)^T$ such that $z_i = x_i + y_i$ for $i = 1, \cdots, n$.

**Subtraction.** The difference of $x$ and $y$, denoted by $x - y$, is a vector $z = (z_1, \cdots, z_n)^T$ such that $z_i = x_i - y_i$ for $i = 1, \cdots, n$.

**Scalar multiplication.** The product of the scalar $c$ and the vector $x$, denoted by $cx$, is a vector $z = (z_1, \cdots, z_n)^T$ such that $z_i = cx_i$ for $i = 1, \cdots, n$.

**Inner product.** The inner product of $x$ and $y$, denoted by $x^T y$, is a number computed by

$$x^T y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

**Norm.** The norm (also called the Euclidean norm) of a vector $x$, denoted by $\|x\|_2$, is a number computed as

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

**Euclidean distance.** The Euclidean distance between $x$ and $y$ is given by

$$\|x - y\|_2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

The following properties are trivial:

$$
\begin{aligned}
& x-y = x+(-1)y, \\
& x+y = y+x, \\
& c(x+y) = cx+cy, \text{ and} \\
& \|x\|_2 \geq 0, \text{ and } \|x\|_2 = 0 \text{ iff } x = 0. \\
& \|x\|_2 = \sqrt{x^T x}, \text{ or } \|x\|_2^2 = x^T x.
\end{aligned}
$$

The norm of $x$ is also the Euclidean distance between $x$ and the origin. For example, given $x \in \mathbb{R}^2$, and the origin $0 = (0,0)^T$, then $\|x\| = \sqrt{x_1^2 + x_2^2} = \sqrt{(x_1-0)^2 + (x_2-0)^2}$, which is exactly the Euclidean distance between $x$ and 0. Note that we do not have the division operator (or quotients) between two vectors.

*Example 2.1.* Here are some examples. Given $x = (1,2)^T \in \mathbb{R}^2$, $y = (-3,2)^T \in \mathbb{R}^2$ and $c = 3$. We have

$$
x+y = \begin{pmatrix} -2 \\ 4 \end{pmatrix}, \ x-y = \begin{pmatrix} 4 \\ 0 \end{pmatrix}, \ 3y = \begin{pmatrix} -9 \\ 6 \end{pmatrix}, \ x^T y = -3+4 = 1, \text{ and } \|x\|_2 = \sqrt{1^2+2^2} = \sqrt{5}.
$$

## 2.2 Matrices and matrix operations

### *2.2.1 Definitions and special cases*

An $m \times n$ **matrix** is a rectangular array of numbers arranged in $m$ rows and $n$ columns as follows:

$$
A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}.
$$

The entry that appears at the intersection of the $i$-th row and $j$-th column of $A$ is called the $(i,j)$ entry of $A$. It is written as $A_{ij}$ or sometimes as $a_{ij}$. The notation $A \in \mathbb{R}^{m \times n}$ or $(a_{ij})_{m \times n}$ means that $A$ is an $m \times n$ matrix. We conventionally use square brackets to present matrices in an explicit form as above.

*Example 2.2.* The following matrices are of the dimensions $2 \times 2$, $2 \times 3$, $2 \times 1$ and $1 \times 3$, respectively, with entries such as $A_{11} = 1, A_{12} = 2, A_{21} = 3, A_{22} = 4$ and $B_{21} = 4$, etc.

$$
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \ B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \ C = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \ D = \begin{bmatrix} 3 & 0 & 4 \end{bmatrix}.
$$

Clearly, if a matrix has one row (or one column), it becomes a row vector (or a column vector). In other words, an $m \times 1$ matrix is an $m$-dimensional column vector, while a $1 \times n$ matrix is an $n$-dimensional row vector. In the above example, $C$ is a 2-dimensional column vector, and $D$ is a 3-dimensional row vector.

Hence, an $m \times n$ matrix $A$ is formed from $n$ column vectors $A_j$ of size $m$, or from $m$ row vectors $a_i^T$ of size $n$. Here, for each $i = 1, \cdots, m$ and each $j = 1, \cdots, n$,

$$A_j = \begin{pmatrix} A_{1j} \\ \vdots \\ A_{mj} \end{pmatrix}, \ a_i = \begin{pmatrix} A_{i1} \\ \vdots \\ A_{in} \end{pmatrix}, \ \text{and } a_i^T = (A_{i1}, \cdots, A_{in}).$$

*Example 2.3.* $A = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 1 & 4 \\ 4 & -2 & 3 \end{bmatrix}$ is formed from 3 column vectors $A_1 = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}$, $A_2 = \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}$, and $A_3 = \begin{pmatrix} -3 \\ 4 \\ 3 \end{pmatrix}$, or from 3 row vectors $a_1^T = (1, 2, -3)$, $a_2^T = (2, 1, 4)$ and $a_3^T = (4, -2, 3)$.

A **zero** matrix is a matrix where all elements are zeros. For example, $O = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is the $2 \times 2$ zero matrix.

If the number of rows in a matrix equals the number of its columns, then this matrix is called a **square matrix**. For example, $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is a $2 \times 2$ square matrix. An $n \times n$ square matrix is called a **diagonal matrix** if all of its off-diagonal entries are 0's. We often use $\mathrm{diag}(d_1, \cdots, d_n)$ to denote the diagonal matrix

$$\begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}.$$

A square matrix is called an **identity matrix** if its diagonal entries are all 1's, and its off-diagonal entries are all 0's. That is, the identity matrix is a special diagonal matrix with unit diagonal entries. We use $\mathbb{I}_m$ to denote the $m \times m$ identity matrix. For instance,

$$\mathbb{I}_1 = \begin{bmatrix} 1 \end{bmatrix}, \ \ \mathbb{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \ \mathbb{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

### 2.2.2 Basic matrix operations

There are a number of basic operations that can be applied to matrices.

**Scalar-matrix multiplication:** Given a matrix $A$ and a scalar $c$, the product $cA$ is computed by multiplying every entry of $A$ by $c$. That is

$$cA = \begin{bmatrix} cA_{11} & cA_{12} & \cdots & cA_{1n} \\ cA_{21} & cA_{22} & \cdots & cA_{2n} \\ \vdots & \vdots & & \vdots \\ cA_{m1} & cA_{m2} & \cdots & cA_{mn} \end{bmatrix}.$$

*Example 2.4.* If $A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}$, then $3A = \begin{bmatrix} 3 & 6 \\ -3 & 0 \end{bmatrix}$, and $-2A = \begin{bmatrix} -2 & -4 \\ 2 & 0 \end{bmatrix}$.

**Addition:** Two matrices of the same dimension can be added. Given two $m \times n$ matrices $A$ and $B$, their sum $A+B$ is an $m \times n$ matrix computed by summing up the corresponding entries of $A$ and $B$. That is

$$A+B = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \cdots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{22}+B_{22} & \cdots & A_{2n}+B_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \cdots & A_{mn}+B_{mn} \end{bmatrix}.$$

*Example 2.5.* If $A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} -1 & -2 \\ 2 & 1 \end{bmatrix}$, then $A+B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$.

**Subtraction:** Two matrices of the same dimension can be subtracted. Given two $m \times n$ matrices $A$ and $B$, their difference $A-B$ is an $m \times n$ matrix computed by subtracting the entries of $B$ from the corresponding entries of $A$. That is

$$A-B = \begin{bmatrix} A_{11}-B_{11} & A_{12}-B_{12} & \cdots & A_{1n}-B_{1n} \\ A_{21}-B_{21} & A_{22}-B_{22} & \cdots & A_{2n}-B_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1}-B_{m1} & A_{m2}-B_{m2} & \cdots & A_{mn}-B_{mn} \end{bmatrix}.$$

*Example 2.6.* If $A = \begin{bmatrix} 1 & 2 \\ -1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} -1 & -2 \\ 2 & 1 \end{bmatrix}$, then $A-B = \begin{bmatrix} 2 & 4 \\ -3 & -1 \end{bmatrix}$.

Only matrices of the same dimension can be added or subtracted. It is obvious that $A-B = A+(-1)B$.

**Transpose:** Given an $m \times n$ matrix $A$, its transpose $A^T$ is an $n \times m$ matrix obtained by turning rows of $A$ into columns (and columns of $A$ into rows). That is

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}^T = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{m1} \\ A_{12} & A_{22} & \cdots & A_{m2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{mn} \end{bmatrix}.$$

*Example 2.7.* If $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{bmatrix}$.

The transpose of an $n$-dimensional column vector is an $n$-dimensional row vector, and vice versa, as we have seen before.

*Example 2.8.* If $B = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, then $B^T = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

**Matrix multiplication:** Given two matrices $A = (a_{ij})_{m\times n} \in \mathbb{R}^{m\times n}$ and $B = (b_{ij})_{p\times q} \in \mathbb{R}^{p\times q}$, the product $AB$ is well defined if $n = p$. When it is well defined, we denote the product $AB$ by $C = (c_{ij})_{m\times q}$, which is an $m \times q$ matrix such that its entries are computed by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} \cdot b_{kj} \quad \forall i = 1, \cdots m, \quad j = 1, \cdots q.$$

**Condition of matrix multiplication:** Two matrices $A = (a_{ij})_{m\times n} \in \mathbb{R}^{m\times n}$ and $B = (b_{ij})_{p\times q} \in \mathbb{R}^{p\times q}$ can be multiplied as $AB$ if the number of columns of $A$ is equal to the number of rows of $B$, i.e., $n = p$.

We note that the fact $AB$ is well defined does not mean that the product $BA$ is also well defined.

*Example 2.9.* In the following calculations, the products $AB$ are well defined.

If $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 1 & 2 \end{bmatrix}$, then

$$AB = \begin{bmatrix} 1(1)+2(2)+3(1) & 1(3)+2(0)+3(2) \\ (-1)(1)+0(2)+2(1) & (-1)(3)+0(0)+2(2) \end{bmatrix} = \begin{bmatrix} 8 & 9 \\ 1 & 1 \end{bmatrix}.$$

If $A = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 0 \end{bmatrix}$, then

$$AB = \begin{bmatrix} 3(1) & 3(3) & 3(0) \\ 2(1) & 2(3) & 2(0) \\ 1(1) & 1(3) & 1(0) \end{bmatrix} = \begin{bmatrix} 3 & 9 & 0 \\ 2 & 6 & 0 \\ 1 & 3 & 0 \end{bmatrix}.$$

This product is also called the outer product of two vectors $A$ and $B$.

If $A = \begin{bmatrix} 3 & 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$, then

$$AB = \begin{bmatrix} 3(1) + 2(3) + 1(0) \end{bmatrix} = \begin{bmatrix} 9 \end{bmatrix}.$$

This becomes the inner product of two column vectors $A^T$ and $B$.

### 2.2.3 Properties of matrix operations

Similar to operations on real numbers, matrix operations enjoy properties such as associativity and distributivity. The following lists basic properties satisfied by matrix operations. In this list, we use $c$ to denote an arbitrary real number, and use $A$, $B$, $C$ to denote matrices.

The dimensions of $A$, $B$ and $C$ need to be consistent for the involved operations to be well defined. For example, the first part of the third item below says that when the dimensions of $A$, $B$ $C$ are such that $A+B$, $AC$ and $BC$ are well defined, then computing $(A+B)C$ can be equivalently done by multiplying $A$ and $B$ by $C$ separately and then summing the products up.

- *Associativity of scalar-matrix multiplication:*

$$(cA)(B) = A(cB) = c(AB).$$

- *Associativity of matrix multiplication:*

$$(AB)C = A(BC).$$

- *Distributivity of matrix multiplication:*

$$(A+B)C = AC + BC \quad \text{and} \quad C(A+B) = CA + CB.$$

- *Transposes of matrix product/sum:*

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

- *Multiplication with identity matrices:* for any $m \times n$ matrix $A$, we have

$$\mathbb{I}_m A = A \mathbb{I}_n = A,$$

  where $\mathbb{I}_m$ and $\mathbb{I}_n$ are two identity matrices of sizes $m$ and $n$ respectively.

It is important to keep in mind that matrix multiplication is NOT commutative in general. Namely, in general,

$$AB \neq BA.$$

For example, if $A \in \mathbb{R}^{3\times 4}$ and $B \in \mathbb{R}^{4\times 5}$, then $AB$ is a $3 \times 5$ matrix, but $BA$ is not even well defined!

*Example 2.10.* If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $C = \begin{bmatrix} 2 & 0 & 3 & 0 \end{bmatrix}$, then $AB = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ and $BC = \begin{bmatrix} 2 & 0 & 3 & 0 \\ -2 & 0 & -3 & 0 \end{bmatrix}$. Consequently,

$$(AB)C = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 3 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 0 & -3 & 0 \\ -2 & 0 & -3 & 0 \end{bmatrix}$$

and

$$A(BC) = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 2 & 0 & 3 & 0 \\ -2 & 0 & -3 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 0 & -3 & 0 \\ -2 & 0 & -3 & 0 \end{bmatrix}.$$

### 2.2.4 Block matrices

A **block matrix** or a **partitioned matrix** is a partition of a matrix into rectangular smaller matrices called blocks. The matrix is split up by horizontal and vertical lines that go all the way across.

*Example 2.11.* The matrix

$$P = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \end{bmatrix}$$

can be partitioned into 4 blocks

$$P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

with

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 6 & 7 & 8 \end{bmatrix}, B = \begin{bmatrix} 4 & 5 \\ 9 & 10 \end{bmatrix}, C = \begin{bmatrix} 11 & 12 & 13 \end{bmatrix}, \text{ and } D = \begin{bmatrix} 14 & 15 \end{bmatrix}.$$

If $A$ and $B$ are two matrices of the same dimension, and they are partitioned in the same way, then $A+B$ can be computed block by block. For example, if

$$A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \text{ and } B = \begin{bmatrix} B_1 & B_2 \end{bmatrix},$$

and the dimensions of $A_1$ (respectively, $A_2$) are the same as those of $B_1$ (respectively, $B_2$), then

$$A + B = \begin{bmatrix} A_1 + B_1 & A_2 + B_2 \end{bmatrix}.$$

If $M$ and $N$ are two matrices of dimensions $m \times n$ and $n \times k$ respectively, and the columns of $M$ are partitioned in the same way as rows of $N$, then we can compute $MN$ using block-by-block multiplication. More precisely, if we have

$$M = \begin{bmatrix} A_1 & B_1 \\ C_1 & D_1 \end{bmatrix} \text{ and } N = \begin{bmatrix} A_2 & B_2 \\ C_2 & D_2 \end{bmatrix},$$

then the equality

$$MN = \begin{bmatrix} A_1A_2 + B_1C_2 & A_1B_2 + B_1D_2 \\ C_1A_2 + D_1C_2 & C_1B_2 + D_1D_2 \end{bmatrix}$$

holds as long as all the product matrices, such as $A_1A_2$, are well defined. In other words, the number of columns in $A_1$ needs to equal the number of rows in $A_2$, and the number of columns in $B_1$ needs to equal the number of rows in $C_2$.

*Example 2.12.* Given two matrices

$$A = \left[\begin{array}{cc|cc} 1 & 3 & 2 & -1 \\ 3 & 2 & 1 & 0 \\ -2 & 4 & 0 & 5 \\ \hline -1 & -2 & 3 & 1 \\ 5 & 4 & 2 & -1 \end{array}\right]_{5\times 4} \quad \text{and} \quad B = \left[\begin{array}{cc|cc} 2 & 1 & 3 & 4 \\ -1 & 0 & 2 & 1 \\ \hline 1 & -1 & 0 & 3 \\ 4 & 5 & -2 & -3 \end{array}\right]_{4\times 4},$$

then, the product of $A$ and $B$ can be computed as

$$AB = \begin{bmatrix} \begin{bmatrix} 1 & 3 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 4 & 5 \end{bmatrix} & \begin{bmatrix} 1 & 3 \\ 3 & 2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ 0 & 5 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ -2 & -3 \end{bmatrix} \\ \begin{bmatrix} -1 & -2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 4 & 5 \end{bmatrix} & \begin{bmatrix} -1 & -2 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 0 & 3 \\ -2 & -3 \end{bmatrix} \end{bmatrix}$$

$$= \left[\begin{array}{cc|cc} -3 & -6 & 11 & 16 \\ 5 & 2 & 13 & 17 \\ 12 & 23 & -8 & -19 \\ \hline 7 & 1 & -9 & 0 \\ 4 & -2 & 25 & 33 \end{array}\right].$$

For an $m \times n$ matrix $A$, recall that $A_j$ denotes its $j$-th column, and $a_i$ denotes the transpose of the $i$-th row of $A$. The matrix $A$ can be partitioned by columns or by rows:

$$A = \begin{bmatrix} A_1 \ A_2 \ \cdots \ A_n \end{bmatrix} = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_m^T \end{bmatrix}.$$

The product $Ax$ of a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $x \in \mathbb{R}^n$ can be written as

$$Ax = \begin{bmatrix} a_1^T x \\ a_2^T x \\ \vdots \\ a_m^T x \end{bmatrix}$$

by partitioning $A$ by rows, or as

$$Ax = \sum_{i=1}^{n} x_i A_i$$

by partitioning $A$ by columns.

### *2.2.5 Invertible matrices, dependence and independence*

For each **square** matrix $A \in \mathbb{R}^{m \times m}$, one can compute its **determinant**, denoted by $\det(A)$, which is a real number. For a matrix $A \in \mathbb{R}^{1 \times 1}$, $\det(A)$ is simply defined as $\det A = a_{11}$. For a matrix $A \in \mathbb{R}^{2 \times 2}$, the determinant is given by the formula

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

For a matrix $A \in \mathbb{R}^{3\times 3}$, we can compute it by

$$\begin{aligned}\det(A) &= a_{11}\det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{bmatrix}\right) - a_{12}\det\left(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33}\end{bmatrix}\right) + a_{13}\det\left(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{332}\end{bmatrix}\right) \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{13}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{33}a_{21} - a_{11}a_{23}a_{32}.\end{aligned}$$

For higher dimensions, the computation of $\det A$ is much more complicated, and we shall not discuss it in this course. Students can read, e.g., [**?**] for further details.

- If a matrix $A$ satisfies $\det A = 0$, then it is called a *singular matrix*.
- If a matrix $A$ satisfies $\det A \neq 0$, then it is called a *nonsingular matrix* or *an invertible matrix*.

*Example 2.13.* For example, given $A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 3 \\ 3 & 2 & 1\end{bmatrix}$, then

$$\det(A) = 1\times1\times1+2\times3\times3+2\times2\times(-1)-3\times1\times(-1)-2\times2\times1-2\times3\times1 = 1+18-4+3-4-6 = 8.$$

Hence, $A$ is a nonsingular matrix, or $A$ is invertible.

> Given a nonsingular matrix $A \in \mathbb{R}^{n\times n}$, there exists a unique $n\times n$ matrix, denoted by $A^{-1}$, that satisfies
>
> $$A(A^{-1}) = \mathbb{I}_n.$$
>
> Here $\mathbb{I}_n$ is the $n\times n$ identity matrix. The matrix $A^{-1}$ is called the inverse of $A$, which also satisfies $(A^{-1})A = \mathbb{I}_n$.

For instance, the inverse of $A = \begin{bmatrix} 2 & 1 \\ 1 & -1\end{bmatrix}$ is $A^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3}\end{bmatrix}$. To check this, first calculate $\det(A) = -3$ to find $A$ to be nonsingular, and then compute $AA^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & -1\end{bmatrix}\begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3}\end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1\end{bmatrix}$. This suffices to verify that the given matrix is the inverse of $A$. One can further check that $A^{-1}A = \mathbb{I}_2$ as well.

While it is possible for two non-square matrices $A$ and $B$ to satisfy $AB = \mathbb{I}$ or $BA = \mathbb{I}$, we do not call them inverses of each other since they are non-square.

#### 2.2.5.1 Facts about invertible matrices

- If $A$ is invertible, then $A^T$ is also invertible, with

$$(A^T)^{-1} = (A^{-1})^T.$$

- If both $A$ and $B$ are invertible, and they have the same dimension, then $AB$ is invertible, with

$$(AB)^{-1} = B^{-1}A^{-1}.$$

#### 2.2.5.2 Linear combinations

Let $v^{(1)}, v^{(2)}, \cdots, v^{(k)}$ be $k$ column vectors of the same dimension. That is, each $v^{(i)} \in \mathbb{R}^m$ for $i = 1, \cdots, k$.

- A **linear combination** of these $k$ vectors is any vector $u \in \mathbb{R}^m$ that can be written as

$$u = c_1 v^{(1)} + c_2 v^{(2)} + \cdots c_k v^{(k)}$$

  for some real numbers $c_1, c_2, \cdots, c_k$. The numbers $c_1, \cdots, c_k$ are called coefficients of the linear combination.
- The vectors $v^{(1)}, v^{(2)}, \cdots, v^{(k)}$ are **linearly dependent** if there exist $k$ coefficients $c_1, c_2, \cdots, c_k$, not all being zeros, with

$$c_1 v^{(1)} + c_2 v^{(2)} + \cdots c_k v^{(k)} = 0.$$

  Here $0 = (0, 0, \cdots, 0)^T$ is the zero vector in $\mathbb{R}^m$.
- The vectors $v^{(1)}, v^{(2)}, \cdots, v^{(k)}$ are **linearly independent** if they are not linearly dependent. In other words, it is linearly independent if the following equality

$$c_1 v^{(1)} + c_2 v^{(2)} + \cdots c_k v^{(k)} = 0$$

  holds if and only if $c_1 = c_2 = \cdots = c_k = 0$.

The above definitions apply to row vectors as well.

*Example 2.14.* Consider the following 4 vectors in $\mathbb{R}^3$:

$$v^{(1)} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \ v^{(2)} = \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \ v^{(3)} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \text{ and } v^{(4)} = \begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix}.$$

- The vectors $v^{(1)}, v^{(2)}, v^{(3)}$ are linearly independent. To see this, consider a linear combination $c_1 v^{(1)} + c_2 v^{(2)} + c_3 v^{(3)} = 0$ and write this explicitly as

$$\begin{cases} c_1 & -2c_2 & +2c_3 = 0 \\ 2c_1 & +c_2 & = 0 \\ c_1 & & = 0. \end{cases}$$

The last equation shows that $c_1 = 0$. Substitute $c_1 = 0$ into the second equation, we have $c_2 = 0$. Substitute both $c_1 = 0$ and $c_2 = 0$ into the first equation, we get $c_3 = 0$. This shows that $(c_1, c_2, c_3) = 0$ is the only coefficient vector that satisfies $c_1 v^{(1)} + c_2 v^{(2)} + c_3 v^{(3)} = 0$. Hence, $v^{(1)}, v^{(2)}, v^{(3)}$ is linearly independent.

- The vectors $v^{(1)}, v^{(2)}, v^{(4)}$ are linearly dependent. Indeed, we can easily observe that $v^{(1)} + v^{(2)} = v^{(4)}$. This is equivalent to $v^{(1)} + v^{(1)} - v^{(4)} = 0$. Hence, the coefficients $c_1 = 1$, $c_2 = 1$ and $c_4 = -1$ satisfy $c_1 v^{(1)} + c_2 v^{(2)} + c_4 v^{(4)} = 0$.

The following theorem gives the relation between independence of vectors and invertibility of matrices. The proof is omitted. Two statements are said to be equivalent if they imply each other.

**Theorem 2.1.** *Let A be a square matrix. The following statements are equivalent.*

1. *A is invertible.*
2. *The rows of A are linearly independent.*
3. *The columns of A are linearly independent.*

When $A$ is not a square matrix, we have the following result. A **submatrix** of a matrix is obtained by deleting any collection of rows and/or columns.

**Theorem 2.2.** *Let $A \in \mathbb{R}^{m \times n}$ where $m \leq n$. The following statements are equivalents.*

1. *The rows of A are linearly independent.*
2. *There exist m columns of A that are linearly independent.*
3. *A has a nonsingular $m \times m$ submatrix.*

Proofs of these two theorems can be found in standard linear algebra books, e.g., [**?**].

*Example 2.15.* Consider the following $4 \times 7$ matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 & 0 & 1 & 2 \\ 0 & 0 & 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

One can readily check that the rows of $A$ are linearly independent. It follows from the above theorem that one must be able to find 4 columns of $A$ that are linearly independent. Indeed, the first 4 columns are linearly independent. This does not mean any 4 columns of $A$ are linearly independent; for example you cannot pick columns 1, 5, 6, and 7.

## 2.3 Systems of linear equations

### *2.3.1 Definitions and examples*

Consider a system of linear equations

$$Ax = b,$$

where $A \in \mathbb{R}^{m \times n}$ is a given left-hand-side coefficient matrix and $b \in \mathbb{R}^m$ is a given right-hand-side vector. We can write this system in the explicit form as

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

This system can also be represented in the augmented matrix form $[A \mid b]$. The *solution set* of this system is the set of all of its solutions, that is, the set of all vectors $x \in \mathbb{R}^n$ that satisfy $Ax = b$.

Depending on its solution set, a linear system may belong to one of the following three types:

1. *A linear system with infinitely many solutions:* This happens when $b$ is a linear combination of columns of $A$, and columns of $A$ are linearly dependent.
2. *A linear system with a unique solution:* This happens when $b$ is a linear combination of columns of $A$, and columns of $A$ are linearly independent.
3. *A linear system with no solution:* This happens when $b$ is not a linear combination of columns of $A$.

In general, if $\bar{x}$ is a known solution to the system $Ax = b$, then the solution set of this system is

$$\{\bar{x} + d \mid Ad = 0\}.$$

If $A$ is an invertible square matrix, then the system $Ax = b$ has a unique solution $x$ for any $b \in \mathbb{R}^n$, and this solution is $x = A^{-1}b$.

> We say that a system of linear equations $Ax = b$ is equivalent to a system of linear equations $Cx = d$ if any solution of $Ax = b$ is also a solution of $Cx = d$, and vice versa.

*Example 2.16.* Consider two systems of linear equations:

$$\begin{cases} x_1 + x_2 + x_3 & = 3 \\ x_1 + 2x_2 + 3x_3 & = 6, \end{cases} \quad \text{and} \quad \begin{cases} x_1 + 3x_2 + 5x_3 & = 9 \\ 2x_1 + 3x_2 + 4x_3 & = 9. \end{cases}$$

Let $(x_1, x_2, x_3)$ be a solution of the first system. Then, $x_1 + x_2 + x_3 = 3$ and $x_1 + 2x_2 + 3x_3 = 6$. Summing up these two equations, we get $2x_1 + 3x_2 + 4x_3 = 9$, which is the second equation of the second system. Moreover, if we compute $2 \times (x_1 + 2x_2 + 3x_3 = 6) - (x_1 + x_2 + x_3 = 3)$, we obtain the first equation of the second system. Similarly, we can obtain the first system from the second system by combining its equations. Hence, these two systems are equivalent.

### *2.3.2 Elementary row operations*

When we apply the following elementary row operations on a system of linear equations, they do NOT change the solution set of this system.

**Type 1:** Multiplying an equation by a nonzero number.

**Type 2:** Adding a multiple of one equation to another equation.

**Type 3:** Switching two equations.

More generally, consider a system of linear equations represented by

$$Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. We can pre-multiply both sides of it by an invertible matrix $M \in \mathbb{R}^{m\times m}$, to obtain a new system $MAx = Mb$. Note that

$$Ax = b \quad \Leftrightarrow \quad MAx = Mb$$

so the two systems $Ax = b$ and $MAx = Mb$ have exactly the same solution set.

*Pre-multiplying both sides of a system of linear equations by an invertible matrix does not change the solution set of this system.*

In fact, conducting an elementary row operation on a system of equations is equivalent to pre-multiplying the system by a special matrix.

*Example 2.17.* Consider the following system of linear equations $Ax = b$:

$$\begin{cases} x_1 + x_2 + x_3 &= 3 \\ x_1 + 2x_2 + 3x_3 &= 6, \end{cases} \quad \text{or in the matrix form} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

Now, let $M = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ be an invertible matrix $\det(M) = -3 \neq 0$). Then, the system $MAx = Mb$ becomes

$$\begin{cases} 3x_1 + 5x_2 + 7x_3 = 15 \\ 3x_1 + 4x_2 + 5x_3 = 12. \end{cases}$$

In fact, the first equation of the latter system is the sum of the first equation and two times the second equation of the former system, and the second equation of the latter system is the sum of two times the first equation and the second equation of the former system.

### *2.3.3 Gauss-Jordan's method for solving systems of linear equations*

Consider a system of $m$ linear equations in the augmented-matrix form $[A \mid b]$. The Gauss-Jordan (GJ) method proceeds as follows.

1. Initialization: set the counter $k = 1$.
2. Find the leftmost column in $A$ that is not all zeros from rows $k$ to $m$. If no such a column exists, stop the procedure right away.

   The following illustrates this step for the case in which $k = 1$.

$$\left[\begin{array}{cccccc|c} 0 & \cdots & 0 & * & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & * & \cdots & * & * \end{array}\right]$$
$$\Uparrow$$

3. Use elementary row operations, we transform this column to the unit vector $\mathbf{e}_k$ (i.e., a column with 1 as the $k$th entry, and 0 elsewhere). In order to do so, we carry out the following steps:

   a. If the $k$th entry of this column is zero, switch the $k$th row with another row so that the $k$th entry of this column becomes $a \neq 0$.
   b. Multiply the $k$th row by $1/a$.
   c. Add multiples of the $k$th row to other rows so that all other entries of this column become zeros.

   The following illustrates these steps for the case in which $k = 1$.

$$\left[\begin{array}{cccccc|c} 0 & \cdots & 0 & 0 & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & * & \cdots & * & * \end{array}\right] \xrightarrow{\text{After step (a)}} \left[\begin{array}{cccccc|c} 0 & \cdots & 0 & a & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & * & \cdots & * & * \end{array}\right] \xrightarrow{\text{After step (c)}} \left[\begin{array}{cccccc|c} 0 & \cdots & 0 & 1 & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & * & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & * & * \end{array}\right].$$

4. If $k = m$, stop the procedure. Otherwise, increase $k$ by 1, and go back to Step 2.

At the end of the Gauss-Jordan method, we obtain a system of linear equations in the so-called "reduced row-echelon form", from which we can read off its solution set.

*Example 2.18.* The linear system

$$\begin{cases} 2x_1 + 2x_2 + x_3 = 9 \\ 2x_1 - x_2 + 2x_3 = 6 \\ x_1 - x_2 + 2x_3 = 5 \end{cases}$$

can be written into the matrix form $Ax = b$ as:

$$\begin{bmatrix} 2 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & -1 & 2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 6 \\ 5 \end{pmatrix}.$$

Hence, its augmented-matrix form becomes

$$[A \mid b] = \left[\begin{array}{ccc|c} 2 & 2 & 1 & 9 \\ 2 & -1 & 2 & 6 \\ 1 & -1 & 2 & 5 \end{array}\right]$$

**The Gauss-Jordan method:** We apply Gauss-Jordan's method to solve this system, which consists of the following steps:

*Step 1:* Multiply row 1 by $1/2$ to get

$$\left[\begin{array}{ccc|c} 1 & 1 & 1/2 & 9/2 \\ 2 & -1 & 2 & 6 \\ 1 & -1 & 2 & 5 \end{array}\right].$$

*Step 2:* Add -2(row 1) to row 2 to get

$$\left[\begin{array}{ccc|c} 1 & 1 & 1/2 & 9/2 \\ 0 & -3 & 1 & -3 \\ 1 & -1 & 2 & 5 \end{array}\right].$$

*Step 3:* Add -1(row 1) to row 3 to get

$$\left[\begin{array}{ccc|c} 1 & 1 & 1/2 & 9/2 \\ 0 & -3 & 1 & -3 \\ 0 & -2 & 3/2 & 1/2 \end{array}\right].$$

*Step 4:* Multiply row 2 by $-1/3$ to get

$$\left[\begin{array}{ccc|c} 1 & 1 & 1/2 & 9/2 \\ 0 & 1 & -1/3 & 1 \\ 0 & -2 & 3/2 & 1/2 \end{array}\right].$$

*Step 5:* Add -1(row 2) to row 1 to get

$$\left[\begin{array}{ccc|c} 1 & 0 & 5/6 & 7/2 \\ 0 & 1 & -1/3 & 1 \\ 0 & -2 & 3/2 & 1/2 \end{array}\right].$$

*Step 6:* Add 2(row 2) to row 3 to get

$$\left[\begin{array}{ccc|c} 1 & 0 & 5/6 & 7/2 \\ 0 & 1 & -1/3 & 1 \\ 0 & 0 & 5/6 & 5/2 \end{array}\right].$$

*Step 7:* Multiply row 3 by $6/5$ to get

$$\left[\begin{array}{ccc|c} 1 & 0 & 5/6 & 7/2 \\ 0 & 1 & -1/3 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

*Step 8:* Add -5/6(row 3) to row 1 to get

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & -1/3 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

*Step 9:* Add 1/3(row 3) to row 2 to get

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

Since each operation does not change the solution set of the system, the last system has the same solution set as the original system. Here, the last system has a unique solution $x = (1,2,3)^T$, which is also the unique solution of the original system.

*Example 2.19.* Consider the following simple system:

$$\begin{cases} x_1 + 2x_2 & = 3 \\ 2x_1 + 4x_2 & = 4. \end{cases}$$

If we apply Gauss-Jordan's method to this system, then we have

$$\left[\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & 4 \end{array}\right] \quad \Rightarrow \quad \left[\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 0 & -2 \end{array}\right].$$

The second row in the last system says $0x_1 + 0x_2 = -2$, which can never be true. Therefore, this system has no solution.

*Example 2.20.* Consider the linear system:

$$\begin{cases} x_1 + x_2 \phantom{{}+x_3} = 1 \\ \phantom{x_1 +{}} x_2 + x_3 = 3 \\ x_1 + 2x_2 + x_3 = 4. \end{cases}$$

If we apply Gauss-Jordan's method to this system, then we have

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 3 \\ 1 & 2 & 1 & 4 \end{array}\right] \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 3 \\ 0 & 1 & 1 & 3 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 1 & 1 & 3 \end{array}\right] \quad \Rightarrow \quad \left[\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

We can rewrite the last system as

$$\begin{cases} x_1 - x_3 = -2 \\ x_2 + x_3 = 3. \end{cases}$$

This system has infinitely many solutions: for any arbitrary value of $x_3 \in \mathbb{R}$, $x_1 = -2 + x_3$ and $x_2 = 3 - x_3$. Its solution set can be represented as

$$\{x = (-2+t, 3-t, t) \mid t \in \mathbb{R}\}.$$

*Example 2.21.* Solve the following system of equations with three variables and three equations:

$$\begin{cases} x_1 + x_2 + 2x_3 = 9 \\ x_1 + x_2 + x_3 = 6 \\ \phantom{x_1} - x_2 + 2x_3 = 5. \end{cases}$$

First, we write this system in the matrix form $Ax = b$ with

$$[A \mid b] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 1 & 1 & 1 & 6 \\ 0 & -1 & 2 & 5 \end{array}\right].$$

Then, we apply the GJ method to transform this matrix into the row-echelon form as:

$$[A \mid b] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 1 & 1 & 1 & 6 \\ 0 & -1 & 2 & 5 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 0 & -1 & -3 \\ 0 & -1 & 2 & 5 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & -1 & 2 & 5 \\ 0 & 0 & -1 & -3 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -2 & -5 \\ 0 & 0 & -1 & -3 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 4 & 14 \\ 0 & 1 & -2 & -5 \\ 0 & 0 & -1 & -3 \end{array}\right] \Rightarrow \cdots \Rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

We finally obtain the solution of this system as $x = (2,1,3)^T$.

#### 2.3.3.1 Back substitution

There is an alternative way to apply Gauss-Jordan's method. Instead of transforming $[A \mid b]$ until $A$ contains an identity submatrix, we can apply Gauss-Jordan's method to transform $[A \mid b]$ until $A$ becomes an upper trapezoidal matrix (a matrix is an upper trapezoidal matrix if nonzero elements are found only in the upper triangle of the matrix, including the main diagonal). Following that, we apply the **back substitution method** to solve the resulting system.

Below we briefly describe this method for the special case in which the number of variables equals the number of equations in the system. This method can readily be applied to general systems.

Given a system of linear equations $Ax = b$ with $A \in \mathbb{R}^{n\times n}$, we perform two steps:

1. Apply the GJ method to $[A \mid b]$ to transform it into a row echelon form $[\tilde{A} \mid \tilde{b}]$ (or upper triangular):

$$\left[\begin{array}{cccc|c} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} & \tilde{b}_1 \\ 0 & \tilde{a}_{22} & \cdots & \tilde{a}_{2n} & \tilde{b}_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \tilde{a}_{nn} & \tilde{b}_n \end{array}\right]$$

2. Then apply the back substitution method to solve the linear system $\tilde{A}x = \tilde{b}$. If all diagonal elements of $\tilde{A}$ are nonzero, this method performs backward steps from the bottom to the top as follows:

   - Compute $x_n$ from $\tilde{a}_{nn}x_n = \tilde{b}_n$ as $x_n = \frac{\tilde{b}_n}{\tilde{a}_{nn}}$.
   - Once $x_{i+1}, \cdots, x_n$ are computed, we can compute $x_i$ recursively as $x_i = \frac{\tilde{b}_i - \sum_{j=i+1}^n \tilde{a}_{ij}x_j}{\tilde{a}_{ii}}$ for $i = n-1$ down to $i = 1$.

   When some diagonal elements are zero, the system either has no solution or infinitely many solutions. In the latter case, the solution set can be written by expressing some variables in terms of other variables.

*Example 2.22.* To illustrate this method, we reconsider Example 2.21 by solving

$$\begin{cases} x_1 + x_2 + 2x_3 = 9 \\ x_1 + x_2 + x_3 = 6 \\ \quad - x_2 + 2x_3 = 5. \end{cases}$$

We have $[A \mid b] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 1 & 1 & 1 & 6 \\ 0 & -1 & 2 & 5 \end{array}\right]$.

*Step 1:* After applying the GJ method we can transform this matrix into an upper triangular matrix:

$$[\tilde{A} \mid \tilde{b}] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 0 & 1 & -2 & -5 \\ 0 & 0 & -1 & -3 \end{array}\right].$$

Hence, we can write this system as

$$\begin{cases} x_1 + x_2 + 2x_3 = 9 \\ \quad x_2 - 2x_3 = -5 \\ \quad\quad -x_3 = -3. \end{cases}$$

*Step 2:* Now, we apply the back substitution step as

– From the last equation, we can compute $x_3 = \frac{-3}{-1} = 3$.

– From the second equation, we have $x_2 = -5 + 2x_3 = -5 + 2 \times 3 = 1$.

– From the first equation, we have $x_1 = 9 - x_2 - 2x_3 = 9 - 1 - 2 \times 3 = 2$.

Hence, the solution of the original linear system is $x = (2, 1, 3)^T$.

### 2.3.4 Gauss-Jordan's method for computing the inverse of a nonsingular matrix

We can use the Gauss-Jordan method to compute the inverse of an $n \times n$ nonsingular matrix $A$. This method works as follows:

- First, we write the matrix $A$ and the identity matrix $\mathbb{I}$ side by side, as $[A \mid \mathbb{I}]$.
- Second, we use elementary row operations to transform the matrix $A$ into the row-echelon form (or upper triangular), so that its left-bottom corner below the diagonal is all zero.
- Third, we further use elementary row operations to transform the matrix $A$ into a diagonal matrix.
- Fourth, we normalize the diagonal elements of the diagonal matrix, to obtain a matrix of the form $[\mathbb{I} \mid B]$, where $B$ is exactly the inverse of $A$.

*Example 2.23 (Inverse matrix).* Using the Gauss-Jordan method to compute the inverse of $A = \begin{bmatrix} 1 & 2 & -4 \\ 2 & 1 & -1 \\ 3 & 0 & 1 \end{bmatrix}$.

Our method performs as follows:

- First, we write
$$[A \mid \mathbb{I}] = \left[\begin{array}{ccc|ccc} 1 & 2 & -4 & 1 & 0 & 0 \\ 2 & 1 & -1 & 0 & 1 & 0 \\ 3 & 0 & 1 & 0 & 0 & 1 \end{array}\right].$$
- Next, we transform this matrix into the "row-echelon form". The following steps are performed:
  1. Keep row 1 unchanged.
  2. Replace row 2 by row 2 subtracting two times of row 1 from row 2. That is, $R_2 \leftarrow R_2 - 2 \times R_1$.
  3. Replace row 3 by row 3 subtracting three times of row 1 from row 3. That is, $R_3 \leftarrow R_3 - 3 \times R_1$.
  4. Then, replace row 3 by row 3 subtracting two times of row 2 from row 3. That is, $R_3 \leftarrow R_3 - 2 \times R_2$.
$$\left[\begin{array}{ccc|ccc} 1 & 2 & -4 & 1 & 0 & 0 \\ 0 & -3 & 7 & -2 & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array}\right].$$
- The next step is to transform this matrix to a "row-diagonal form".
  1. Keep the last row unchanged.
  2. Replace row 2 by row 2 adding to row 3 after multiplying it by 7. That is, $R_2 \leftarrow R_2 + 7 \times R_3$.
  3. Replace row 1 by row 1 subtracting four times of row 3 from row 1. That is, $R_1 \leftarrow R_1 - 4 \times R_3$.
  4. Then, replace row 1 by row 1 adding to row 2 after multiplying it by $\frac{2}{3}$. That is, $R_1 \leftarrow R_1 + \frac{2}{3} \times R_2$.
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{3} & \frac{-2}{3} & \frac{2}{3} \\ 0 & -3 & 0 & 5 & -13 & 7 \\ 0 & 0 & -1 & 1 & -2 & 1 \end{array}\right].$$
- Finally, we normalize this matrix to get the identity matrix on the left by dividing the second row by $-3$, and the last row by $-1$.
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{3} & \frac{-2}{3} & \frac{2}{3} \\ 0 & 1 & 0 & \frac{-5}{3} & \frac{13}{3} & \frac{-7}{3} \\ 0 & 0 & 1 & -1 & 2 & -1 \end{array}\right] = [\mathbb{I} \mid B].$$

- This matrix is of the form $[\mathbb{I} \mid B]$, we conclude that $B = A^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{-2}{3} & \frac{2}{3} \\ \frac{-5}{3} & \frac{13}{3} & \frac{-7}{3} \\ -1 & 2 & -1 \end{bmatrix}$.

### 2.3.5 *Checking linearly independence of vectors*

Given $n$ column vectors $\{A_1, A_2, \cdots, A_n\}$ in $\mathbb{R}^m$, we can use Gauss-Jordan method to check if they are linearly independent. If we let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

be the $m \times n$ matrix formed by these vectors, then the equation

$$x_1 A_1 + x_2 A_2 + \cdots x_n A_n = 0$$

can be equivalently written as

$$Ax = 0,$$

where $x \in \mathbb{R}^n$ is the column vector with entries given by the coefficients $x_1, \cdots, x_n$. If the above equation has nonzero solutions, then the vectors $A_1, \cdots, A_n$ are linearly dependent. Otherwise, if the only solution to the above system is zero, then these vectors are linearly independent.

## 2.4 Exercises

**Exercise 2.1.** Given two matrices $A = \begin{bmatrix} 4 & 3 & 2 \\ 2 & 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & -2 \end{bmatrix}$. Answer the following questions:

1. Compute the product $AB$ using the definition of matrix multiplication.
2. Partition $A$ into 2 blocks by a vertical line between its 2nd and 3rd columns, and partition $B$ into 2 blocks by a horizontal line between its 2nd and 3rd rows. Compute $AB$ using block operations.

**Exercise 2.2.** Use the Gauss-Jordan method to solve the following linear systems. Show each step of your calculations.

1.
$$\begin{cases} \phantom{x_1 +} x_2 + 2x_3 = 3 \\ x_1 + 2x_2 + x_3 = 4 \\ x_1 + x_2 - 2x_3 = 0. \end{cases}$$

2.
$$\begin{cases} x_1 - 4x_2 + 2x_3 = -4 \\ 2x_2 - x_3 = 1 \\ -x_1 + 2x_2 - x_3 = 3 \\ -2x_1 + 6x_2 - 3x_3 = 7. \end{cases}$$

3.
$$\begin{cases} x_1 + x_4 = 5 \\ x_2 + 2x_4 = 5 \\ x_3 + 0.5x_4 = 1 \\ 2x_3 + x_4 = 3. \end{cases}$$

**Exercise 2.3.** Given the following four linear equation systems of the form $Ax = b$, where their augmented matrix $\bar{A} = [A \mid b]$ are respectively given by

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 2 & 4 \\ 0 & 1 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 2 \end{array}\right], \quad \left[\begin{array}{cccc|c} 1 & 0 & 1 & 9 & 3 \\ 0 & 1 & -1 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{array}\right], \quad \left[\begin{array}{cccc|c} 1 & 0 & 1 & 9 & 0 \\ 0 & -1 & 1 & -3 & 1 \\ 1 & -1 & 2 & 6 & -2 \end{array}\right].$$

For each system, determine if it has a unique solution, multiple solutions, or no solution. Write down the solution set of each system when it exists.

**Exercise 2.4.** Use the Gauss-Jordan method to compute the inverse matrix of the following matrices:

$$(a)\ A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \qquad (b)\ A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 0 & 1 \\ 2 & 3 & -2 \end{bmatrix}.$$

Show the steps of your Gauss-Jordan transformation. Verify the result by computing $AA^{-1}$.

To show a square matrix $A$ is not invertible (that is $A^{-1}$ does not exist), we can also use the Gauss-Jordan method. First, we apply the Gauss-Jordan method to transform $[A \mid \mathbb{I}]$ into a trapezoidal form $[\tilde{A} \mid \tilde{B}]$, where $\tilde{A}$ is a lower triangular matrix. Then, if there exists a zero entry on the diagonal of $\tilde{A}$, then we can conclude that $A$ is not invertible.

Now, use this theory and the Gauss-Jordan method to show that $A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & -2 \\ 5 & 2 & -3 \end{bmatrix}$ is not invertible.

**Exercise 2.5.** Determine if each of the following statements is true or false. Justify your answer by a proof or a counter example. Proofs should be based on the material covered in this course.

1. The equality $x^T y = y^T x$ holds for any two $n$-dimensional column vectors $x$ and $y$.

2. If $A$, $B$ and $C$ are invertible matrices of the same dimension, then the product of their transposes, $A^TB^TC^T$, is invertible.
3. If $A$, $B$ and $C$ are invertible matrices of the same dimension, then their sum $A+B+C$ is invertible.
4. Suppose that $A$ is an invertible matrix. Switch its top two rows to obtain a new matrix $B$. The matrix $B$ must be invertible as well.
5. Let $v^1, v^2, \cdots, v^k$ be column vectors of the same dimension, and let $v = v^1 + v^2 + \cdots + v^k$. The vectors $v^1, v^2, \cdots, v^k, v$ are linearly dependent.
6. The system $Ax = b$ has a solution, if $b$ is a linear combination of columns of $A$.
7. For a system of linear equations that contains 2 equations and 3 variables (i.e., $A \in \mathbb{R}^{2\times 3}$), there are always infinitely many solutions.

**Exercise 2.6.** 1. Given $k$ column vectors $v^i \in \mathbb{R}^3$ for $i = 1, \cdots k$ as follows:

$$v^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad v^2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad v^3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v^4 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \quad v^5 = \begin{pmatrix} 1 \\ 1 \\ -3 \end{pmatrix}.$$

Find all the possible groups of 3 vectors such that each group is linearly independent. Is the group $\{v^1, v^2, v^3, v^4\}$ linearly independent? If it is not, please explain.

2. Given a matrix $A$

$$A = \begin{bmatrix} -1 & 2 & 3 & 4 \\ 5 & 0 & -1 & -1 \\ 8 & -6 & -10 & -13 \end{bmatrix}.$$

   a. Show that the last row of $A$ is a linear combination of the first two rows.
   b. Let $B$ be a $3\times 3$ matrix formed by three columns of $A$. Is $B$ invertible? Does your answer depend on which three columns of $A$ are included in $B$?

**Exercise 2.7.** Given two vector $x, y \in \mathbb{R}^n$, we say that $x$ and $y$ are orthogonal if $x^Ty = 0$. If, in addition, $\|x\|_2 = \|y\|_2 = 1$, then we say that $x$ and $y$ are orthonormal. Find an orthonormal vector $b$ for each of the following vectors $a$:

$$a = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T, \quad \text{and } a = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{6}}\right)^T.$$

Find a vector $c$ such that it is simultaneously orthonormal with two vectors $a = \frac{1}{\sqrt{14}}(1,2,3)^T$ and $b = \frac{1}{\sqrt{5}}(2,-1,0)^T$. Is $c$ unique?

**Exercise 2.8.** An $m \times n$ matrix $A$ is called column orthogonal if each column of $A$ is orthogonal to other columns in $A$. If, in addition, the length of each column $A_j$ is unit (i.e., $\|A_j\|_2 = 1$), then $A$ is called a column orthonormal matrix. Given an orthogonal matrix $A$, we can always orthonormalize it by dividing each column by its norm. Check if the following matrices are column orthogonal/orthonormal.

$$A = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad A = \frac{1}{4}\begin{bmatrix} 1 & 1 & -1 \\ 2 & -2 & 0 \\ -2 & 1 & 0 \end{bmatrix} \text{ and } A = \begin{bmatrix} -\frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{2} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

If $A$ is orthonormal, compute $A^{-1}$ and compare it to $A^T$.