Toddler Linear Algebra

Jes Modian

May 18, 2024

Abstract

Linear Algebra is the study of linear maps on finite dimensional vector spaces. It involves stuff like matrices and vectors, and helps us solve systems of linear equations, find equilibrium probability in a Markov chain, and much more. Linear Algebra is said to be the most important subject in undergraduate mathematics, as it appears everywhere. However, when I am learning it, things like $\det(AB) = \det(A)\det(B)$ often appear out of nowhere and I am just told to accept it. So I am going to figure them out in this note.

Contents

0 Preliminaries

0.1 Properties of equality

Properties of equality (used for solving systems of equation):

1. $a = a$ for any object a (Reflexive property)
2. If $a = b$ and $c = d$, then $a + c = b + d$. (Addition of two equations / Additive property)
3. If $a = b$ and $c = d$, then $a - c = b - d$. (Subtraction of two equations / Subtractive property)
4. If $a = b$ and $c = d$, then $ac = bd$. (Multiplication of two equations / Multiplicative property)
5. If $a = b$ and $c = d \neq 0$, then $\frac{a}{c} = \frac{b}{d}$. (Division of two equations / Divisive property)
6. If $a = b$, then $f(a) = f(b)$ (Taking function of both sides of an equation)

From (4) we can multiply an equation by any constant, say multiply $a = b$ by c to get $ac = bc$.
(4) becomes if $a = b$ and $c = c$, then $ac = bc$. $a = b$ is true by assumption and $c = c$ by reflexive property, so $ac = bc$ must be true.

(5) can be inferred from (4) since $\frac{a}{c}$ is just $a(\frac{1}{c})$. Thus we can also divide an equation by a non-zero constant.

(6) We can take function of both sides of an equation since by definition, function can only have one output for the same input.

Some more properties:

7. If $ca = cb$ and $c \neq 0$, then $a = b$

   Proof. Divide both sides by c to get $\frac{ca}{c} = \frac{cb}{c}$, thus $a = b$. □

8. If $a \neq b$ and $c \neq 0$, then $ca \neq cb$.

   Proof. The contrapositive is if $ca = cb$ and $c \neq 0$, then $a = b$, which is property (7). □

0.2 Summation notation

The summation notation (sigma notation) is used to write sums concisely.

Definition

For a sequence $a_m, a_{m+1}, \dots, a_n$, the summation of all terms is defined as:

$$\sum_{i=m}^{n} a_i = a_m + a_{m+1} + \ldots + a_{n-1} + a_n$$

Here, i (called index) is a dummy variable and can be replaced by any other variable that is not m or n (to avoid conflict).
m and n must be integers because they are term-numbers. If $m > n$, then there are zero terms to sum up so the summation is zero.

Example

$$\sum_{i=3}^{6} i^2 = 3^2 + 4^2 + 5^2 + 6^2 = 86$$

$$\sum_{i=3}^{6} 5 = 5 + 5 + 5 + 5 = 20$$

$$\sum_{i=5}^{5} i(i+1) = 5(5+1) = 30$$

$$\sum_{i=8}^{7} i = 0$$

Double summation

Let $a_{i,j}$ be a "2D sequence" term that depends on two variables: i and j. The summation of all terms for $m \leq i \leq n$ and $p_i \leq j \leq q_i$ is defined as:

$$\sum_{i=m}^{n}\sum_{j=p_i}^{q_i} a_{i,j} = \sum_{j=p_m}^{q_m} a_{m,j} + \sum_{j=p_{m+1}}^{q_{m+1}} a_{m+1,j} + \ldots + \sum_{j=p_n}^{q_n} a_{n,j}$$

where p_i and q_i may be constants or expressions in terms of i.

Example:

$$\sum_{i=1}^{3}\sum_{j=1}^{2} ij = 1(1) + 1(2) + 2(1) + 2(2) + 3(1) + 3(2) = 18$$

$$\sum_{i=1}^{2}\sum_{j=3i}^{3i+2} \frac{i}{i+j} = (\frac{1}{1+3} + \frac{1}{1+4} + \frac{1}{1+5}) + (\frac{2}{2+6} + \frac{2}{2+7} + \frac{2}{2+8}) = \frac{37}{60} + \frac{121}{180} = \frac{58}{45}$$

Properties

Let c be a constant.

1. $\displaystyle\sum_{i=1}^{n} c = cn$ (constant summation)

Proof. $\displaystyle\sum_{i=1}^{n} c = \underbrace{c + c + \ldots + c}_{n \text{ times}} = cn$ □

2. $\displaystyle\sum_{i=m}^{n} ca_i = c\sum_{i=m}^{n} a_i$ (constant rule)

Proof. $\displaystyle\sum_{i=m}^{n} ca_i = ca_m + ca_{m+1} + \ldots + ca_{n-1} + ca_n = c(a_m + a_{m+1} + \ldots + a_{n-1} + a_n) = c\sum_{i=m}^{n} a_i$ □

3. $\displaystyle\sum_{i=m}^{n} a_i \pm \sum_{i=m}^{n} b_i = \sum_{i=m}^{n}(a_i \pm b_i)$ (distributive property)

Proof.

$$\begin{aligned}\sum_{i=m}^{n} a_i + \sum_{i=m}^{n} b_i &= (a_m + a_{m+1} \ldots + a_n) + (b_m + b_{m+1} + \ldots + b_n)\\ &= (a_m + b_m) + (a_{m+1} + b_{m+1}) \ldots + (a_n + b_n)\\ &= \sum_{i=m}^{n}(a_i + b_i)\end{aligned}$$

Difference of two summations is done similarly:

$$\begin{aligned}\sum_{i=m}^{n} a_i - \sum_{i=m}^{n} b_i &= (a_m + a_{m+1} \ldots + a_n) - (b_m + b_{m+1} + \ldots + b_n)\\ &= (a_m - b_m) + (a_{m+1} - b_{m+1}) \ldots + (a_n - b_n)\\ &= \sum_{i=m}^{n}(a_i - b_i)\end{aligned}$$

□

4. $\displaystyle\sum_{i=m}^{n} a_i = \sum_{i=m}^{k} a_i + \sum_{i=k+1}^{n} a_i$ (splitting summation)

Proof. $\sum_{i=m}^{n} a_i = a_m + \ldots + a_k + a_{k+1} + \ldots + a_n = (a_m + \ldots + a_k) + (a_{k+1} + \ldots + a_n) = \sum_{i=m}^{k} a_i + \sum_{i=k+1}^{n} a_i$ $\square$

5. $\sum_{i=m}^{n} a_i = \sum_{i=k}^{n} a_i - \sum_{i=k}^{m-1} a_i$ (variant of splitting summation)

Proof. By property of splitting summation, $\sum_{i=k}^{m-1} a_i + \sum_{i=m}^{n} a_i = \sum_{i=k}^{n} a_i \implies \sum_{i=m}^{n} a_i = \sum_{i=k}^{n} a_i - \sum_{i=k}^{m-1} a_i$ $\square$

6. $\sum_{i=m}^{n} a_i = \sum_{i=m+k}^{n+k} a_{i-k}$ (index shift)

Proof. $\sum_{i=m+k}^{n+k} a_{i-k} = a_{(m+k)-k} + \ldots + a_{(n+k)-k} = a_m + \ldots + a_n = \sum_{i=m}^{n} a_i$ $\square$

7. $\sum_{i=m}^{n} a_i = \sum_{i=0}^{n-m} a_{n-i}$ (reverse order of summation)

Proof.

$$\begin{aligned}
\sum_{i=0}^{n-m} a_{n-i} &= a_{n-0} + a_{n-1} + \ldots + a_{n-(n-m-1)} + a_{n-(n-m)} \\
&= a_n + a_{n-1} + \ldots + a_{m+1} + a_m \\
&= a_m + a_{m+1} + \ldots + a_{n-1} + a_n
\end{aligned}$$

$\square$

8. $\sum_{i=0}^{n} a_i = \sum_{i=0}^{n} a_{n-i}$ (A particular case of above by putting $m = 0$)

9. $\left(\sum_{i=m}^{n} a_i\right)\left(\sum_{j=p}^{q} b_j\right) = \sum_{i=m}^{n}\sum_{j=p}^{q} a_i b_j$ (product of summation is double summation of products)

Proof.

$$\begin{aligned}
\left(\sum_{i=m}^{n} a_i\right)\left(\sum_{j=p}^{q} b_j\right) &= \left(\sum_{i=m}^{n} a_i\right)(b_p + b_{p+1} + \ldots + p_q) \\
&= \sum_{i=m}^{n} a_i(b_p + b_{p+1} + \ldots + p_q) \qquad \text{(constant rule)} \\
&= \sum_{i=m}^{n} (a_i b_p + a_i b_{p+1} + \ldots + a_i p_q) \\
&= \sum_{i=m}^{n} (\sum_{j=p}^{q} a_i b_j) \\
&= \sum_{i=m}^{n}\sum_{j=p}^{q} a_i b_j \qquad \text{(brackets can be dropped for double summation)}
\end{aligned}$$

$\square$

10. $\sum_{i=m}^{n}\sum_{j=p}^{q} a_i b_j = \sum_{j=p}^{q}\sum_{i=m}^{n} a_i b_j$ (interchange of independent double summation)

Proof.

$$\begin{aligned}
\sum_{i=m}^{n}\sum_{j=p}^{q} a_i b_j &= \left(\sum_{i=m}^{n} a_i\right)\left(\sum_{j=p}^{q} b_j\right) && \text{(product of double summation)}\\
&= (a_m + a_{m+1} + \ldots + a_n)\left(\sum_{j=p}^{q} b_j\right)\\
&= \sum_{j=p}^{q} b_j (a_m + a_{m+1} + \ldots + a_n) && \text{(constant rule)}\\
&= \sum_{j=p}^{q} (a_m b_j + a_{m+1} b_j + \ldots + a_n b_j)\\
&= \sum_{j=p}^{q}\sum_{i=m}^{n} a_i b_j
\end{aligned}$$

□

11. $\displaystyle\sum_{i=m}^{n}\sum_{j=p}^{q} a_{i,j} = \sum_{j=p}^{q}\sum_{i=m}^{n} a_{i,j}$ (interchange of double summation with independent bounds)

Proof.

$$\begin{aligned}
\sum_{i=m}^{n}\sum_{j=p}^{q} a_{i,j} &= \sum_{j=p}^{q} a_{m,j} + \sum_{j=p}^{q} a_{m+1,j} + \ldots + \sum_{j=p}^{q} a_{n,j}\\
&= a_{m,p} + a_{m,p+1} + \ldots + a_{m,q}\\
&\quad + a_{m+1,p} + a_{m+1,p+1} + \ldots + a_{m+1,q}\\
&\quad + \ldots\\
&\quad + a_{n,p} + a_{n,p+1} + \ldots + a_{n,q}\\
&= \sum_{i=m}^{n} a_{i,p} + \sum_{i=m}^{n} a_{i,p+1} + \ldots + \sum_{i=m}^{n} a_{i,q}\\
&= \sum_{j=p}^{q}\sum_{i=m}^{n} a_{i,j}
\end{aligned}$$

□

This is a stronger version of the previous property.

Conditional summation notation

In conditional summation notatation, all the terms a_i where the term number i satisfies a certain condition $p(i)$ are summed up. The upper bound is no longer written above, and the lower bound in the bottom in normal summation is replaced by the condition $p(i)$.

Example:

$$\sum_{1\leq k\leq 100} k^2 = \sum_{k=1}^{100} k^2 = 1^2 + 2^2 + \ldots + 100^2$$

Let $S = \{2, 3, 5, 7\}$.

$$\sum_{i\in S} i = 2 + 3 + 5 + 7 = 17$$

$$\sum_{i\in\mathbb{Z}^+} \frac{1}{2^i} = \sum_{i=1}^{\infty} \frac{1}{2^i} = 2$$

1 Introduction

Note: I am going to copy stuff from the books *Linear Algebra Done Right* [1], *Linear Algebra and Its Applications* [2], *Linear Algebra With Applications* [3], and my university lecture slides.

Warning: This is a note where exposition is removed to save space and definitions come out of nowhere, unlike the textbooks that actually explain the motivation behind the concepts.

1.1 Complex numbers

Definition 1.1. A **complex number** is an ordered pair (a, b), where $a, b \in \mathbb{R}$, written as $a + bi$. a is called the **real part** and b is called the **imaginary part** .

- The set of all complex numbers is denoted by $\mathbb{C}$:

$$\mathbb{C} = \{a + bi \ : \ a, b \in \mathbb{R}\}$$

- Addition and multiplication on $\mathbb{C}$ are defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

for $a, b, c, d \in \mathbb{R}$.

The definition is motivated by treating i as a variable with the special property that $i^2 = -1$, and applying the usual rules of arithmetic.

Example:

$$\begin{aligned}(2 + 3i)(4 + 5i) &= (2)(4) + (2)(5i) + (3i)(4) + (3i)(5i) \\ &= 8 + 15(-1) + 10i + 12i \\ &= -7 + 22i\end{aligned}$$

Two complex numbers are equal if and only if the real part is equal to the imaginary part, namely, if $a + bi = c + di$, then $a = c$ and $b = d$.

Note that $\mathbb{R}$ is a subset of $\mathbb{C}$, meaning every real number is a complex number. But the vice versa is not true. $1 + i$ is complex but not real.

Theorem 1.1 (Properties of complex arithmetic)**.** For all α, β, λ in $\mathbb{C}$,

(i) **commutativity**
$\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$

(ii) **associativity**
$(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ and $(\alpha\beta)\lambda = \alpha(\beta\lambda)$

(iii) **identities**
$\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$

(iv) **additive inverse**
For every $\alpha \in \mathbb{C}$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha + \beta = 0$.

(v) **multiplicative inverse**
For every $\alpha \in \mathbb{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha\beta = 1$.

(vi) **distributive property**
$\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$

Note: These are also the properties of real numbers if we replace $\mathbb{C}$ with $\mathbb{R}$.

Proof. Let $\alpha = a + bi$, $\beta = c + di$, $\lambda = e + fi$.

(i) $\alpha + \beta = (a + bi) + (c + di) = (a + c) + (b + d)i$
$\beta + \alpha = (c + di) + (a + bi) = (c + a) + (d + b)i = (a + c) + (b + d)i = \alpha + \beta$

$\alpha\beta = (a + bi)(c + di) = (ac - bd) + (ad + bc)i$
$\beta\alpha = (c + di)(a + bi) = (ca - db) + (cb + da)i = \alpha\beta$

(ii) $(\alpha + \beta) + \lambda = (a + bi + c + di) + e + fi = ((a + c) + (b + d)i) + e + fi = (a + c + e) + (b + d + f)i$
$\alpha + (\beta + \lambda) = a + bi + (c + di + e + fi) = a + bi + ((c + e) + (d + f)i) = (a + c + e) + (b + d + f)i$

$(\alpha\beta)\lambda = ((ac - bd) + (ad + bc)i)(e + fi) = (ac - bd)e - (ad + bc)f + ((ac - bd)f + (ad + bc)e)$

$$\begin{aligned}
\alpha(\beta\lambda) &= (a + bi)((ce - df) + (cf + de)i) \\
&= a(ce - df) - b(cf + de) + (a(cf + de) + b(ce - df))i \\
&= ace - adf - bcf - bde + (acf + ade + bce - bdf)i \\
&= (ac - bd)e - (ad + bc)f + ((ac - bd)f + (ad + bc)e)i
\end{aligned}$$

(iii) Here, 0 means $0 + 0i$ and 1 means $1 + 0i$.
$\lambda + 0 = e + fi + 0 + 0i = e + fi$
$(e + fi)(1 + 0i) = (1e - 0f) + (0e + 1f)i = e + fi$

(iv) For $\alpha = a + bi$, there is only one $\beta = -a - bi$ that will make $\alpha + \beta = 0 + 0i$.

(v) For $\alpha = a + bi \neq 0$, let $\beta = c + di$ such that $\alpha\beta = (ac - bd) + (ad + bc)i = 1 + 0i$. Then we have a system of equations in which we want to solve for c and d:

$$\begin{cases} ac - bd = 1 & \dots (1) \\ ad + bc = 0 & \dots (2) \end{cases}$$

$(1) \times -b$:

$$-abc + b^2 d = -b \qquad \dots (3)$$

$(2) \times a$:

$$a^2 d + abc = 0 \qquad \dots (4)$$

$(3) + (4)$:

$$\begin{aligned}
a^2 d + b^2 d &= -b \\
d(a^2 + b^2) &= -b \\
d &= -\frac{b}{a^2 + b^2}
\end{aligned}$$

Note that we are allowed to make $a^2 + b^2$ to be denominator since a and b are not both zero, so $a^2 + b^2$ must be greater than 0.

Similarly, to solve for c:
$(1) \times a$:

$$a^2 c - abd = a \qquad \dots (5)$$

$(2) \times b$:

$$abd + b^2 c = 0 \qquad \dots (6)$$

$(5) + (6)$:

$$\begin{aligned}
a^2 c + b^2 c &= a \\
c(a^2 + b^2) &= a \\
c &= \frac{a}{a^2 + b^2}
\end{aligned}$$

Thus, for $\alpha = a + bi \neq 0$, there is a unique multiplicative inverse $\beta = \dfrac{a}{a^2 + b^2} - \dfrac{b}{a^2 + b^2}i$ such that $\alpha\beta = 1$. We can verify that $\alpha\beta$ indeed gives 1:

$$\begin{aligned}
\alpha\beta &= (a + bi)(\frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i) \\
&= \frac{a^2}{a^2 + b^2} - \frac{b^2}{a^2 + b^2} + (\frac{-ab}{a^2 + b^2} + \frac{ab}{a^2 + b^2})\, i \\
&= 1 + 0i = 1
\end{aligned}$$

(vi) $\lambda(\alpha+\beta) = (e+fi)(a+bi+c+di) = (e+fi)((a+c)+(b+d)i) = e(a+c)-f(b+d)+(e(b+d)+f(a+c))i$

$$\begin{aligned}\lambda\alpha + \lambda\beta &= (e+fi)(a+bi) + (e+fi)(c+di)\\ &= ae - bf + (be+af)i + ec - df + (de+cf)i\\ &= ae - bf + ec - df + (be+af+de+cf)i\\ &= e(a+c) - f(b+d) + (e(b+d)+f(a+c))i\\ &= \lambda(\alpha+\beta)\end{aligned}$$

□

A set equipped with addition and multiplication satisfying the above six properties is called a **field**. Thus $\mathbb{C}$ and $\mathbb{R}$ are both fields, and share the properties that a field has. In this note, we use $\mathbb{F}$ to denote $\mathbb{C}$ or $\mathbb{R}$, whichever suit our taste (since linear algebra can be more insightful when complex numbers are also considered). When we see a statement involving $\mathbb{F}$, we can replace $\mathbb{F}$ with $\mathbb{C}$ or $\mathbb{R}$, and the statement will be true.

1.2 Scalars and vectors

Definition 1.2. A **scalar** is an element of $\mathbb{F}$.

The term scalar is used to emphasize that it is a number, as opposed to a vector. It can be a complex number or real number, but more often refers to a real number. A scalar can be used to scale a vector.

Definition 1.3. A **list** of **length** n is an ordered collection of n elements (where n is non-negative integer).
• Two lists are equal if and only if they have the same length and the same elements in the same order.

A list is typically written using round brackets or square brackets, with each element separated by commas (which is sometimes omitted). eg. $(1,2,3)$ or $[1,2,3]$. Note that a list must have a finite length.
Example: $(2,5,6) = (2,5,6)$, $(1,2,3) \neq (1,3,2)$, $(0,0,0) \neq (0,0)$
If $(a,b,c) = (d,e,f)$, then $a = d$, $b = e$ and $c = f$.

Definition 1.4. $\mathbb{F}^n$ is the set of all lists of length n of elements of $\mathbb{F}$:

$$\mathbb{F}^n = \{(x_1,\dots,x_n) \;:\; x_k \in \mathbb{F} \text{ for } k = 1,\dots,n\}$$

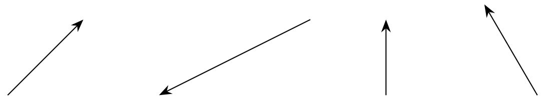
Geometrically, the 2D plane is represented by $\mathbb{R}^2$ and the 3D space is represented by $\mathbb{R}^3$. An element in $\mathbb{R}^2$ (or $\mathbb{R}^3$) is called a **point** , and the list written out in the form (x_1, x_2) is called the **coordinate** or **position** of that point.

Definition 1.5. A **vector** of **dimension** n is a list of length n with the following property: it can be added to by another vector and can be multiplied by a scalar.

A vector of dimension n is called n-dimensional vector.

The difference between lists and vectors is that addition and scalar multiplication is well-defined for vectors but need not be defined for a general list.
Right now, the vectors we are interested in are elements of $\mathbb{F}^n$, but when we get to the topic of vector spaces, vector need not be an element of $\mathbb{F}^n$, and can refer to elements of other stuff.

Geometrically, a 2D or 3D vector is represented by an arrow with a **direction** and **magnitude** (physical length of vector):



(*Some 2D vectors of different directions and magnitudes*)

A vector is written using square brackets instead of round bracket to distinguish it from a position in space. eg. $[1,2,3]$. Each element of the vector is called an **entry**. Geometrically, a vector represents how your position changes when you go from the tail to the tip of the vector.

Vectors can be written in two ways: horizontally and vertically.

A vector written horizontally is called a **row vector**. Usually, commas are written between entries when the vector is seen geometrically as an arrow, and commas are omitted when the vector is seen as a $1 \times n$ matrix. eg. $[1, 2, 3]$ and $[1 \quad 2 \quad 3]$ are the same thing.

A vector written vertically is called a **column vector**. No commas are used for column vectors. eg. $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

An n-dimensional column vector is a $n \times 1$ matrix.
Note that row vectors and column vectors with the same entries are not equal.

To denote a vector as a variable, we typically use boldface: $\mathbf{a}$, or add an overline arrow: $\vec{a}$, in order to distinguish it from scalars.
A vector with all entries zero, eg. [0,0,0], is denoted $\mathbf{0}$ or $\vec{0}$.

A general vector $\mathbf{x}$ with n entries is written as $[x_1, \ldots, x_n]$. Since I'm lazy, I'll shorten this to $[x_j]$ (lazy notation). $[0_j]$ denotes a zero vector with an unspecified number of entries.

1.3 Vector operations

Definition 1.6 (Vector addition)**.** Addition of vectors in $\mathbb{F}^n$ is defined by adding their corresponding entries:

$$[x_1, \ldots, x_n] + [y_1, \ldots, y_n] = [x_1 + y_1, \ldots, x_n + y_n]$$

Addition of vectors of different dimensions are undefined.

Example: $[1, 2] + [3, 4] = [1 + 2, 3 + 4] = [3, 7]$

Definition 1.7 (Additive inverse)**.** For $\mathbf{x} \in \mathbb{F}^n$, the additive inverse of $\mathbf{x}$, denoted by $-\mathbf{x}$, is the vector $-\mathbf{x} \in \mathbb{F}^n$ such that

$$\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$$

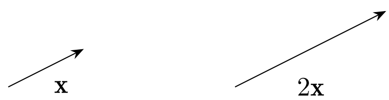
Thus if $\mathbf{x} = [x_1, \ldots, x_n]$, then $-\mathbf{x} = [-x_1, \ldots, -x_n]$

Then we can use $\mathbf{x} - \mathbf{y}$ to denote $\mathbf{x} + (-\mathbf{y})$.
Geometrically, the additive inverse of a vector is a vector with reverse direction:

Definition 1.8 (Scalar multiplication)**.** The product of a number $\lambda \in \mathbb{F}$ and a vector in $\mathbb{F}^n$ is computed by multiplying each entry of the vector by λ:

$$\lambda[x_1, \ldots x_n] = [\lambda x_1, \ldots, \lambda x_n]$$

Geometrically, multiplying a vector by λ scales (stretch or squish) the vector by a factor of λ. In order words, the magnitude of vector is multiplied by λ.



Theorem 1.2 (Algebraic properties of vectors and scalars)**.** For all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{F}^n$ and all scalars c and d:

(i) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$

(ii) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$

(iii) $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$

(iv) $\mathbf{x} + (-\mathbf{x}) = -\mathbf{x} + \mathbf{x} = \mathbf{0}$

(v) $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$

(vi) $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$

(vii) $c(d\mathbf{x}) = (cd)\mathbf{x}$

(viii) $1\mathbf{x} = \mathbf{x}$

Proof. Let $\mathbf{x} = [x_j]$, $\mathbf{y} = [y_j]$, $\mathbf{z} = [z_j]$.

(i) $\mathbf{x} + \mathbf{y} = [x_j] + [y_j] = [x_j + y_j] = [y_j + x_j] = [y_j] + [x_j] = \mathbf{y} + \mathbf{x}$

(ii) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = ([x_j] + [y_j]) + [z_j] = [x_j + y_j] + [z_j] = [x_j + y_j + z_j]$
$\mathbf{x} + (\mathbf{y} + \mathbf{z}) = [x_j] + ([y_j] + [z_j]) = [x_j] + [y_j + z_j] = [x_j + y_j + z_j]$

(iii) $\mathbf{x} + \mathbf{0} = [x_j] + [0_j] = [x_j + 0] = [x_j] = \mathbf{x}$
$\mathbf{0} + \mathbf{x} = [0_j] + [x_j] = [0 + x_j] = [x_j] = \mathbf{x}$

(iv) $\mathbf{x} + (-\mathbf{x}) = [x_j] + [-x_j] = [x_j - x_j] = [0_j] = \mathbf{0}$
$-\mathbf{x} + \mathbf{x} = [-x_j] + [x_j] = [-x_j + x_j] = [0_j] = \mathbf{0}$

(v) $c(\mathbf{x} + \mathbf{y}) = c([x_j] + [y_j]) = c[x_j + y_j] = [c(x_j + y_j)] = [cx_j + cy_j]$
$c\mathbf{x} + c\mathbf{y} = c[x_j] + c[y_j] = [cx_j] + [cy_j] = [cx_j + cy_j] = c(\mathbf{x} + \mathbf{y})$

(vi) $(c + d)\mathbf{x} = (c + d)[x_j] = [(c + d)x_j] = [cx_j + dx_j]$
$c\mathbf{x} + d\mathbf{x} = c[x_j] + d[x_j] = [cx_j] + [dx_j] = [cx_j + dx_j] = (c + d)\mathbf{x}$

(vii) $c(d\mathbf{x}) = c[dx_j] = [cdx_j] = (cd)[x_j] = (cd)\mathbf{x}$

(viii) $1\mathbf{x} = 1[x_j] = [1x_j] = [x_j] = \mathbf{x}$

$\square$

Definition 1.9 (Magnitude)**.** The **magnitude** (/**norm**/length) of a vector $\mathbf{x} = [x_1, \ldots x_n]$ is

$$|\mathbf{x}| = \sqrt{x_1^2 + \ldots + x_n^2}$$

This definition comes from the Pythagoras Theorem (applied $n - 1$ times).

Definition 1.10 (Dot product)**.** The **dot product** of $\mathbf{x}$ and $\mathbf{y}$, denoted $\mathbf{x} \cdot \mathbf{y}$, is defined by

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + \cdots + x_n y_n$$

where $\mathbf{x} = [x_1, \ldots x_n]$ and $\mathbf{y} = [y_1, \ldots y_n]$.

1.4 Matrices

1.4.1 Definition

Definition 1.11. Let m and n be positive integers. An $m \times n$ matrix A is a rectangular array of elements of $\mathbb{F}$ with m rows and n columns:

$$A = \begin{bmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{bmatrix}$$

The notation $a_{i,j}$ denotes the entry in row i, column j of A.

When i and j are both single digit numbers or both variables, the comma between i, j can be omitted to make it more concise, so it becomes

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix}$$

A matrix is typically denoted by a capital letter A, M or others. We can also denote a matrix by its entries, eg. $A = [a_{ij}]$.

To refer to a single entry instead of an entire matrix, we can denote the entry in row m, column n of matrix A by A_{mn} or a_{mn}. (The latter is more often used to emphasize that it is an element, not a matrix.)

To indicate a matrix $A = [a_{ij}]$ has size $m \times n$, we can write $A_{m\times n}$ or $[a]_{m\times n}$ or $[a]_{mn}$.

Definition 1.12. The **size** of a matrix with m rows and n columns is $m \times n$.
Two matrices are equal if and only if they have the same size and the corresponding entries are equal.

Note: The $\times$ is a symbolic times, so the size of a 3×5 matrix is not 15.

Example: If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$, then $a_{11} = 3$, $a_{12} = 4$, $a_{21} = 5$, $a_{22} = 6$.

Example: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ is a 3×2 matrix. The number of rows goes first and column after.

$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \neq \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$ since they have different sizes (first is 3×2 but second is 2×3).

Definition 1.13. An $n \times n$ matrix is called a **square matrix** of size n.

Example: Square matrices of different sizes:

Size 1: $\begin{bmatrix} 69 \end{bmatrix}$, Size 2: $\begin{bmatrix} 2 & 0 \\ 2 & 4 \end{bmatrix}$, Size 3: $\begin{bmatrix} 3 & 9 & 4 \\ 0 & 5 & 2 \\ 7 & 8 & 6 \end{bmatrix}$

Definition 1.14 (Matrix addition)**.** The sum of two matrices of the same size is the matrix obtained by adding corresponding entries in the matrices:

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \dots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

Addition for matrices of different sizes is undefined.

Example: $\begin{bmatrix} 8 & 9 \\ 6 & 4 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 1 & 9 \end{bmatrix} = \begin{bmatrix} 8+2 & 9+0 \\ 6+1 & 4+9 \end{bmatrix} = \begin{bmatrix} 10 & 9 \\ 7 & 13 \end{bmatrix}$

Definition 1.15. The product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar:

$$\lambda \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \lambda a_{11} & \dots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \dots & \lambda a_{mn} \end{bmatrix}$$

The set of all $m \times n$ matrices with entries in $\mathbb{F}$ is denoted by $\mathbb{F}^{m\times n}$.
A matrix with all entries zero is called **zero matrix**, denoted by $\mathbf{0}$.

Theorem 1.3 (Matrix addition and scalar multiplication)**.** For all A, B, C in $\mathbb{F}^{m\times n}$ and all scalars c and d:

(i) $A + B = B + A$

(ii) $(A + B) + C = A + (B + C)$

(iii) $A + \mathbf{0} = \mathbf{0} + A = A$

(iv) $A + (-A) = -A + A = \mathbf{0}$

(v) $c(A + B) = cA + cB$

(vi) $(c + d)A = cA + dA$

(vii) $c(dA) = (cd)A$

(viii) $1A = A$

The proof is similar to Theorem 1.2 (algebraic properties of vectors and scalars).

1.4.2 Matrix multiplication

Definition 1.16 (Matrix multiplication)**.** Suppose $A = [a_{ij}]$ is an $m \times n$ matrix and $B = [b_{ij}]$ is an $r \times p$ matrix. Then AB is defined, if and only if $n = r$, to be the $m \times p$ matrix whose entry in row i, column j, is given by the equation

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Example: Multiplication of two 3×3 matrices:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{bmatrix} = \begin{bmatrix} 1(11)+2(21)+3(31) & 1(12)+2(22)+3(32) & 1(13)+2(23)+3(33) \\ 4(11)+5(21)+6(31) & 4(12)+5(22)+6(32) & 4(13)+5(23)+6(23) \\ 7(11)+8(21)+9(31) & 7(12)+8(22)+9(32) & 7(13)+8(23)+9(33) \end{bmatrix}$$
$$= \begin{bmatrix} 146 & 152 & 158 \\ 335 & 350 & 365 \\ 524 & 548 & 572 \end{bmatrix}$$

Example: A 3×3 matrix multiplied by a 3×1 matrix:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 10 \\ 11 \\ 12 \end{bmatrix} = \begin{bmatrix} 1(10)+2(11)+3(12) \\ 4(10)+5(11)+6(12) \\ 7(10)+8(11)+9(12) \end{bmatrix} = \begin{bmatrix} 68 \\ 167 \\ 266 \end{bmatrix}$$

Example: A 3×2 matrix multiplied by a 2×4 matrix is a 3×4 matrix:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 6 & 5 & 4 & 3 \\ 2 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 10 & 7 & 4 & 1 \\ 26 & 19 & 12 & 5 \\ 42 & 31 & 20 & 9 \end{bmatrix}$$

Note that matrix multiplication is non-commutative, meaning $AB \neq BA$ in general (but $AB = BA$ is sometimes true).

Example:

$$\begin{bmatrix} 1 & 1 \\ 100 & 100 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}1 & 1\\ 100 & 100\end{bmatrix} = \begin{bmatrix}99 & 99\\ -99 & 99\end{bmatrix}$$

Definition 1.17. The **main diagonal** of a matrix A is the list of entries a_{ij} where $i = j$. The entries in the main diagonal are called **diagonal entries**.

Example: Main diagonal of $\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{bmatrix}$ is $(1, 5, 9)$.

Definition 1.18. A **diagonal matrix** is a square matrix in which all entries outside of the main diagonal is zero.
i.e. $A_{n\times n}$ is diagonal if $a_{ij} = 0$ for all $i \neq j$

The entries in the main diagonal can be zero or non-zero.
Examples of diagonal matrix:

$$\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3\end{bmatrix} \qquad \begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

Definition 1.19. An **identity matrix** is a square diagonal matrix with all entries in the main diagonal equal to 1.
i.e. A is identity matrix if $a_{ij} = 0$ for all $i \neq j$ and $a_{ij} = 1$ for all $i = j$.

An identity matrix of size n is denoted I_n or simply I.

Example:

$$I_1 = \begin{bmatrix}1\end{bmatrix}, \quad I_2 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \quad I_3 = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

Theorem 1.4 (Matrix multiplication properties)**.** Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

(i). $(AB)C = A(BC)$ (associative law of multiplication)

(ii). $A(B + C) = AB + AC$ (left distributive law)

(iii). $(B + C)A = BA + CA$ (right distributive law)

(iv). $\lambda(AB) = (\lambda A)B = A(\lambda B)$ for any scalar λ (associative law of scalar multiplication)

(v). $I_m A = AI_n = A$ (identity for matrix multiplication)

Proof. **(i)** Associative law: [4] Let $A = [a]_{mn}$, $B = [b]_{np}$, $C = [c]_{pq}$.

Consider $(AB)C$.

Let $AB = [r]_{mp}$, and $(AB)C = [s]_{mq}$

By definition, $r_{ik} = (AB)_{ik} = \sum_{l=1}^{n} a_{il}b_{lk}$.

Also by definition, $s_{ij} = ((AB)C)_{ij} = \sum_{k=1}^{p} r_{ik}c_{kj}$.

Thus, putting first equation into second equation,

$$\begin{aligned}
s_{ij} &= \sum_{k=1}^{p}\left(\sum_{l=1}^{n} a_{il}b_{lk}\right)c_{kj} \\
&= \sum_{k=1}^{p}\left(a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}\right)c_{kj} \qquad \text{(Expand for better visualization)} \\
&= \sum_{k=1}^{p}\left(a_{i1}b_{1k}c_{kj} + a_{i2}b_{2k}c_{kj} + \ldots + a_{in}b_{nk}c_{kj}\right) \\
&= \sum_{k=1}^{p}\left(\sum_{l=1}^{n} a_{il}b_{lk}c_{kj}\right)
\end{aligned}$$

Now consider $A(BC)$.

Re-let $BC = [r]_{nq}$, $A(BC) = [s]_{mq}$.

By definition, $r_{ij} = (BC)_{ij} = \sum_{k=1}^{p} b_{ik}c_{kj}$.

Also by definition, $s_{ij} = (A(BC))_{ij} = \sum_{l=1}^{n} a_{il}r_{lj}$.

Thus, putting first equation into second equation,

$$\begin{aligned}
s_{ij} &= \sum_{l=1}^{n} a_{il}\left(\sum_{k=1}^{p} b_{lk}c_{kj}\right) \\
&= \sum_{l=1}^{n} a_{il}\left(b_{l1}c_{1j} + b_{l2}c_{2j} + \ldots + b_{lp}c_{pj}\right) \\
&= \sum_{l=1}^{n}(a_{il}b_{l1}c_{1j} + a_{il}b_{l2}c_{2j} + \ldots + a_{il}b_{lp}c_{pj}) \\
&= \sum_{l=1}^{n}\left(\sum_{k=1}^{p} a_{il}b_{lk}c_{kj}\right) \\
&= \sum_{k=1}^{p}\left(\sum_{l=1}^{n} a_{il}b_{lk}b_{kj}\right) \qquad \text{(interchange of double summation with independent bounds)}
\end{aligned}$$

Since the entries of $(AB)C$ and $A(BC)$ are equal, we conclude that $(AB)C = A(BC)$.

(ii) Left distributive law:

$$\begin{aligned}
(A(B+C))_{ij} &= \sum_{k=1}^{n} A_{ik}(B+C)_{kj} \\
&= \sum_{k=1}^{n} A_{ik}(B_{kj} + C_{kj}) \\
&= \sum_{k=1}^{n} A_{ik}B_{kj} + \sum_{k=1}^{n} A_{ik}C_{kj} \\
&= (AB)_{ij} + (AC)_{ij} \\
&= (AB + AC)_{ij}
\end{aligned}$$

(iii) Right distributive law:

$$\begin{aligned}
((B+C)A)_{ij} &= \sum_{k=1}^{n}(B+C)_{ik}A_{kj} \\
&= \sum_{k=1}^{n}(B_{ik} + C_{ik})A_{kj} \\
&= \sum_{k=1}^{n} B_{ik}A_{kj} + \sum_{k=1}^{n} C_{ik}A_{kj} \\
&= (BA)_{ij} + (CA)_{ij} \\
&= (BA + CA)_{ij}
\end{aligned}$$

(iv) Associative law of scalar multiplication:

$$(\lambda(AB))_{ij} = \lambda(AB)_{ij} = \lambda\sum_{k=1}^{n} A_{ik}B_{kj} = \sum_{k=1}^{n} \lambda A_{ik}B_{kj}$$
$$((\lambda A)B)_{ij} = \sum_{k=1}^{n}(\lambda A)_{ik}B_{kj} = \sum_{k=1}^{n} \lambda A_{ik}B_{kj}$$
$$(A(\lambda B))_{ij} = \sum_{k=1}^{n} A_{ik}(\lambda B)_{kj} = \sum_{k=1}^{n} \lambda A_{ik}B_{kj}$$

(v) Identity for matrix multiplication:

Let $I_m = [u]_m$, $A = [a]_{mn}$. Note that $u_{ij} = 1$ if $i = j$, else $u_{ij} = 0$.

$$(I_m A)_{ij} = \sum_{k=1}^{m} u_{ik}a_{kj} = 0 + \ldots + u_{ii}a_{ij} + \ldots + 0 = a_{ij}$$
$$(AI_n)_{ij} = \sum_{k=1}^{n} a_{ik}u_{kj} = 0 + \ldots + a_{ij}u_{jj} + \ldots + 0 = a_{ij}$$

Note that $I_m A$ and AI_n have the same size $m \times n$ and have the same entries as A. Thus $I_m A = AI_n = A$.

□

Notation:

Let's introduce the column vector notation for matrix. A matrix A with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ can be denoted by

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix}$$

Example: $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}$, $\mathbf{a}_3 = \begin{bmatrix} 5 \\ 9 \\ 7 \end{bmatrix}$. Then $\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 6 & 9 \\ 3 & 8 & 7 \end{bmatrix}$

Similarly, a matrix A with rows $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ can be denoted by

$$A = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}$$

Instead of bold vectors, we can also use $\text{row}_i(A)$ to denote the i th row of A, and $\text{col}_j(A)$ to denote the j th column of A.

For identity matrix I_m, we commonly use $\mathbf{e}_i$ to denote the i th column or i th row. (Bold i is reserved for other meanings.)

Theorem 1.5 (Columns of matrix product)**.** Let B be a matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$, and A be a matrix of appropriate size. Then the product AB has the columns $A\mathbf{b}_1, A\mathbf{b}_2, \ldots, A\mathbf{b}_p$. That is,

$$AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_p \end{bmatrix}$$

Alternatively, we have

$$\text{col}_j(AB) = A\ \text{col}_j(B)$$

Proof. Let $A = [a_{ij}]$ be $m \times n$ and $B = [b_{ij}] = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_p \end{bmatrix}$ be $n \times p$.

Note that each $A\mathbf{b}_j$ is a $m \times 1$ column vector (because $m \times n$ times $n \times 1$). Also note that the i-th element of $\mathbf{b}_j$, denoted $(\mathbf{b}_j)_i$ or $(\mathbf{b}_j)_{i1}$, is simply b_{ij}.

By definition, $(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$ and $(A\mathbf{b}_j)_{i1} = \sum_{k=1}^{n} a_{ik}(\mathbf{b}_j)_{k1} = \sum_{k=1}^{n} a_{ik}b_{kj} = (AB)_{ij}$.

In other words, the i-th element of $A\mathbf{b}_j$ is exactly the (i, j)-th entry of AB. Hence, each $A\mathbf{b}_j$ is the j-th column of AB, and we have

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_p \end{bmatrix}$$

□

Example: Let $A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}$, $B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix}$.

Then $A\mathbf{b}_1 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 11 \\ -1 \end{bmatrix}$, $A\mathbf{b}_2 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 13 \end{bmatrix}$, $A\mathbf{b}_3 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 21 \\ -9 \end{bmatrix}$

So $AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & A\mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}$

Theorem 1.6 (Rows of matrix product)**.** Let A be a matrix with rows $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$, and B be a matrix of appropriate size. Then the product of AB has the rows $\mathbf{a}_1 B, \mathbf{a}_2 B, \ldots, \mathbf{a}_m$. That is,

$$AB = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} B = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}$$

Alternatively, we have

$$\text{row}_i(AB) = \text{row}_i(A)B$$

Proof. Note that each $\mathbf{a}_i B$ is a $1 \times p$ row vector. The j th element of $\mathbf{a}_i$, denoted $(\mathbf{a}_i)_j$ or $(\mathbf{a}_i)_{1j}$, is simply a_{ij}.

By definition, $(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$ and $(\mathbf{a}_i B)_{1j} = \sum_{k=1}^{n} (\mathbf{a}_i)_{1k} b_{kj} = \sum_{k=1}^{n} a_{ik} b_{kj}$. The j th entry of $\mathbf{a}_i B$ is exactly the (i, j) th entry of AB. Hence, each $\mathbf{a}_i B$ is the i th row of AB, and we have

$$AB = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}$$

□

Theorem 1.7 (Multiplying matrix by row/column of identity matrix)**.** Let A be an $m \times n$ matrix, and let $\mathbf{e}_j$ be the j th column of I_n, and $\mathbf{r}_i$ be the i th row of I_m. Then

$$A\mathbf{e}_j = \text{col}_j(A) \qquad \text{and} \qquad \mathbf{r}_i A = \text{row}_i(A)$$

Proof. Note that $AI_n = A$. Also by the above theorems, $AI_n = \begin{bmatrix} A\mathbf{e}_1 & A\mathbf{e}_2 \ldots & A\mathbf{e}_n \end{bmatrix}$.

Thus $A = \begin{bmatrix} A\mathbf{e}_1 & A\mathbf{e}_2 & \ldots & A\mathbf{e}_n \end{bmatrix}$, which means each column j of A is equal to $A\mathbf{e}_j$.

Similarly, note that $I_m A = A$. Also by the above theorems, $I_m A = \begin{bmatrix} \mathbf{r}_1 A \\ \mathbf{r}_2 A \\ \vdots \\ \mathbf{r}_m A \end{bmatrix}$.

Thus $A = \begin{bmatrix} \mathbf{r}_1 A \\ \mathbf{r}_2 A \\ \vdots \\ \mathbf{r}_m A \end{bmatrix}$, which means each row i of A is equal to $\mathbf{r}_i A$ □

Theorem 1.8 (Linear combinations of columns)**.** Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{bmatrix}$ be a $m \times n$ matrix and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ be an $n \times 1$ matrix. Then $A\mathbf{b}$ is a linear combination of columns of A with the scalars from the entries of $\mathbf{b}$:

$$A\mathbf{b} = b_1 \mathbf{a}_1 + b_2 \mathbf{a}_2 + \ldots + b_n \mathbf{a}_n$$

Proof. By definition of matrix multiplication, the i-th entry of $A\mathbf{b}$ is $(A\mathbf{b})_i = \sum_{k=1}^{n} a_{ik}b_k$.

Also, note that the i-th entry of each $\mathbf{a}_j$ is a_{ij}.

Thus, the i-th entry of the sum $b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \ldots + b_n\mathbf{a}_n$ is:

$$b_1 a_{i1} + b_1 a_{i2} + \ldots b_1 a_{in} = \sum_{k=1}^{n} a_{ik}b_k$$

Since the corresponding entries are equal, $A\mathbf{b} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \ldots + b_n\mathbf{a}_n$

□

Example: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + 1 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 19 \\ 31 \end{bmatrix}$

Theorem 1.9 (Multiplying by zero)**.** Let A be a matrix, $\mathbf{0}$ be the zero matrix and λ be a scalar. Then

(i). $\mathbf{0}A = A\,\mathbf{0} = 0$

(ii). $0A = \mathbf{0}$

(iii). If $\lambda A = \mathbf{0}$ and $A \neq \mathbf{0}$, then $\lambda = 0$.

Proof. Let A be $m \times n$.

(i) $(\mathbf{0}A)_{ij} = \sum_{i=1}^{m} \mathbf{0}_{ik}a_{kj} = \sum_{i=1}^{m} 0a_{kj} = 0$. Similarly, $(A\,\mathbf{0})_{ij} = \sum_{i=0}^{n} a_{ik}\mathbf{0}_{kj} = \sum_{i=0}^{n} a_{ik}0 = 0$

(ii) $(0A)_{ij} = 0a_{ij} = 0$

(iii) Since $A \neq \mathbf{0}$, there is at least one non-zero entry in A. Let a_{ij} be a non-zero entry in A. Then since $(\lambda A)_{ij} = \lambda a_{ij} = \mathbf{0}_{ij} = 0$ and $a_{ij} \neq 0$, we must have $\lambda = 0$. □

Note: $AB = 0$ does not necessarily mean $A = 0$ or $B = 0$.

Definition 1.20 (Triangular matrix)**.** A square matrix is **triangular** if it is either upper triangular or lower triangular (or both).

A matrix is **upper triangular** if all the entries below main diagonal is zero.
i.e. $a_{ij} = 0$ if $i > j$.

A matrix is **lower triangular** if all the entries above main diagonal is zero.
i.e. $a_{ij} = 0$ if $i < j$.

Example:

Upper triangular:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 6 & 8 & 9 \\ 0 & 0 & 12 & 13 \\ 0 & 0 & 0 & 18 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Lower triangular:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 5 & 6 & 0 & 0 \\ 9 & 10 & 12 & 0 \\ 14 & 16 & 17 & 18 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Definition 1.21 (Powers of matrix)**.** Let k be a non-negative integer. Then the k-th power of A is the product of k copies of A:

$$A^k = \underbrace{A \dots A}_{k \text{ times}}$$

If $k = 0$, the A^0 is defined as identity matrix:

$$A^0 = I$$

Matrix algebra

Since matrices are non-commutative, normal algebraic identities like $(a+b)^2 = a^2+2ab+b^2$ are no longer true when $ab \neq ba$, but it's still true when $ab = ba$.

Let's discover some identities.

$(A+B)^2 = (A+B)(A+B) = A(A+B) + B(A+B) = A^2 + AB + BA + B^2$

$A(B+C)D = A(BD+CD) = ABD + ACD$

$(A+I)^2 = A^2 + AI + IA + I^2 = A^2 + 2A + I$

$(A-I)^2 = A^2 - AI - IA + I^2 = A^2 - 2A + I$

$(A+I)(A-I) = A(A-I) + I(A-I) = A^2 - AI + IA - I^2 = A^2 - I$

$(A+I)(A^2-A+I) = A(A^2-A+I) + I(A^2-A+I) = A^3 - A^2 + A + A^2 - A + I = A^3 - I$

$(A-I)(A^2+A+I) = A(A^2+A+I) - I(A^2+A+I) = A^3 + A^2 + A + A^2 - A - I = A^3 + I$

$(A-2I)(3A+I) = 3A^2 - 5A - 2I$

The normal identities can be used when one variable is A and another is I.
When we do matrix factorization, we use I in place of 1.

1.4.3 Transpose

Definition 1.22. Given an $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A.

i.e. If $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$, then $A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}$

Example: $\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Note: The transpose of a column vector is the corresponding row vector, and vice versa:

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, then $\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$ and $(\mathbf{x}^T)^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

Theorem 1.10 (Transpose means swapping indices)**.** $A_{ij} = (A^T)_{ji}$ and $(A^T)_{ij} = A_{ji}$.

Proof. Let's unpack what transpose means: each i-th row of A becomes the i-th column of A^T. Thus, the j-th element of i-th row of A, which is A_{ij}, becomes the j-th element of i-th column of A^T, which is $(A^T)_{ji}$. Similarly, we have $(A^T)_{ij} = A_{ji}$.

□

Thus, the transpose of a square matrix is just reflecting the entries along the main diagonal (while main diagonal entries stays fixed): $a_{12} \leftrightarrow a_{21}$, $a_{13} \leftrightarrow a_{31}$, $a_{23} \leftrightarrow a_{32}$, and so on.

Theorem 1.11 (Transpose properties)**.** Let A and B denote matrices with appropriate sizes.

(i). $(A^T)^T = A$

(ii). $(A+B)^T = A^T + B^T$

(iii). $(\lambda A)^T = \lambda A^T$ for any scalar λ

(iv). $(AB)^T = B^T A^T$

(v). $I^T = I$

Proof. Let $A = [a_{ij}]$ and $B = [b_{ij}]$.

(i) $((A^T)^T)_{ij} = (A^T)_{ji} = A_{ij}$

(ii) $(A^T + B^T)_{ij} = (A^T)_{ij} + (B^T)_{ij} = A_{ji} + B_{ji}$
$((A+B)^T)_{ij} = (A+B)_{ji} = A_{ji} + B_{ji}$

(iii) $((\lambda A)^T)_{ij} = \lambda A_{ji} = (\lambda A^T)_{ij}$

(iv) $((AB)^T)_{ij} = (AB)_{ji} = \sum_{k=1}^{n} A_{jk}B_{ki}$

$$(B^T A^T)_{ij} = \sum_{k=1}^{n}(B^T)_{ik}(A^T)_{kj} = \sum_{k=1}^{n} B_{ki}A_{jk} = \sum_{k=1}^{n} A_{jk}B_{ki}$$

(v) Note that $I_{ij} = 1$ if $i = j$, else $I_{ij} = 0$.

$(I^T)_{ij} = I_{ji}$. If $i \neq j$, then $I_{ji} = 0$. Else, $I_{ji} = I_{ii} = 1$. Since all the entries remain unchanged, $I^T = I$. $\square$

Definition 1.23. A **symmetric matrix** is a square matrix that is equal to its transpose:
i.e. A is symmetric if $A^T = A$.

A **skew-symmetric** matrix is a square matrix whose transpose is equal to the negative of itself.
i.e. A is skew-symmetric if $A^T = -A$.

1.4.4 Inverse matrix

Definition 1.24. An $n \times n$ matrix is **invertible** (/**non-singular**) if there exists an $n \times n$ matrix C such that

$$AC = CA = I$$

where $I = I_n$. C is called the **inverse** of A.

Note that only square matrix can have an inverse. If A were $m \times n$ where $m \neq n$, and C is $n \times n$, then AC is $m \times n$ which is not $n \times n$, so AC could not be identity matrix I_n.

Not all square matrices have inverse. A square matrix that has no inverse is called **singular** (/**non-invertible**) matrix.

Example: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ 8 & 9 & 12 \end{bmatrix}$ has the inverse $\begin{bmatrix} -3 & 3 & -1 \\ 8 & -12 & 5 \\ -4 & 7 & -3 \end{bmatrix}$ since

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ 8 & 9 & 12 \end{bmatrix} \begin{bmatrix} -3 & 3 & -1 \\ 8 & -12 & 5 \\ -4 & 7 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note: We only check one side since if A and C are square matrices and $AC = I$, then it must be true that $CA = I$. But we will prove this later (it seems quite difficult).

Theorem 1.12. The inverse of an invertible matrix is unique.

Proof. Suppose B and C are inverses of A. Then $AC = I$ and $BA = I$ by definition, and we have

$$B = BI = B(AC) = (BA)C = IC = C$$

. So B and C are actually equal. □

Since inverse of a matrix A is unique, we can unambiguously denote it by A^{-1}.

Theorem 1.13 (Some properties of inverse)**.** .

(i). I is invertible and $I^{-1} = I$.

(ii). If A is invertible, then so is A^{-1}, and

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

(iii). If A and B are invertible, then so is AB, and

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

(iv). If A is invertible, then so is A^T, and

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

(v). Let n be a non-negative integer. If A is invertible, then so is A^n, and

$$(A^n)^{-1} = (A^{-1})^n$$

(vi). If A is invertible, then so is λA for scalar $\lambda \neq 0$, and

$$(\lambda A)^{-1} = \frac{1}{\lambda} A^{-1}$$

Proof. **(i)** Note that $II = I$, so I is the inverse of I itself.

(ii) By definition of inverse, $AA^{-1} = I$ and $A^{-1}A = I$. Hence A is the inverse of A^{-1}.

(iii) Note that $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$

Similarly, $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$. Thus, $(AB)^{-1} = B^{-1}A^{-1}$.

(iv) Use the property $(AB)^T = B^TA^T$ for the following:

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

Similarly, $A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I$. Thus, $(A^T)^{-1} = (A^{-1})^T$.

(v) $(A^n)(A^{-1})^n = (\underbrace{A \dots A}_{n \text{ times}})(\underbrace{A^{-1} \dots A^{-1}}_{n \text{ times}}) = \underbrace{A \dots (A}_{n \text{ times}}\underbrace{A^{-1}) \dots A^{-1}}_{n \text{ times}} = A \dots I \dots A^{-1} = \dots = I$

Similarly, $(A^{-1})^n(A^n) = (\underbrace{A^{-1} \dots A^{-1}}_{n \text{ times}})(\underbrace{A \dots A}_{n \text{ times}}) = I$.

Hence, $(A^n)^{-1} = (A^{-1})^n$.

(vi) $(\lambda A)(\frac{1}{\lambda}A^{-1}) = \lambda(\frac{1}{\lambda})(AA^{-1}) = 1I = I$

Similarly, $(\frac{1}{\lambda}A^{-1})(\lambda A) = (\frac{1}{\lambda})\lambda(A^{-1}A) = 1I = I$. Thus, $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$. □

Since $(A^n)^{-1} = (A^{-1})^n$, we can unambiguously denote both by A^{-n} (but this is non-standard notation for some reason).

Theorem 1.14. A is invertible if and only if A^T is invertible.

Proof. By (iv) of previous theorem, if A is invertible, then A^T is invertible.

If A^T is invertible, then there exists B such that $A^TB = BA^T = I$. Taking transpose of both sides, we have

$$\begin{aligned} (A^TB)^T &= I^T \\ B^T(A^T)^T &= I \\ B^TA &= I \end{aligned} \qquad \text{and} \qquad \begin{aligned} (BA^T)^T &= I^T \\ (A^T)^TB^T &= I \\ AB^T &= I \end{aligned}$$

Thus, B^T is the inverse of A, meaning A is invertible. (And so $A^{-1} = ((A^T)^{-1})^T$.)

□

Theorem 1.15 (Generalized reversed product rule)**.** (i).

$$(A_1A_2\dots A_{k-1}A_k)^T = A_k^TA_{k-1}^T\dots A_1^T$$

(ii). If $A_1, \dots, A_k$ are all invertible, then so is their product $A_1A_2\dots A_{k-1}A_k$, and we also have

$$(A_1A_2\dots A_{k-1}A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\dots A_1^{-1}$$

Proof. From the rule $(AB)^T = B^TA^T$, let "A" $= A_1\dots A_{k-1}$, "B" $= A_k$. Apply this rule repeatedly.

$$\begin{aligned} (A_1A_2\dots A_{k-1}A_k)^T &= A_k^T(A_1\dots A_{k-2}A_{k-1})^T \\ &= A_k^TA_{k-1}^T(A_1\dots A_{k-2})^T \\ &= \dots \\ &= A_k^TA_{k-1}^T\dots A_2^TA_1^T \end{aligned}$$

Inverse of product is done similarly. First we show that $(A_1A_2\dots A_{k-1}A_k)^{-1}$ exists.

$$\begin{aligned} (A_1A_2\dots A_{k-1}A_k)(A_k^{-1}A_{k-1}^{-1}\dots A_1^{-1}) &= A_1A_2\dots A_{k-1}(A_kA_k^{-1})A_{k-1}^{-1}\dots A_1^{-1} \\ &= A_1A_2\dots A_{k-1}IA_{k-1}^{-1}\dots A_1^{-1} \\ &= A_1A_2\dots A_{k-2}(A_{k-1}A_{k-1}^{-1})A_{k-2}^{-1}\dots A_1^{-1} \\ &= \dots \\ &= A_1A_1^{-1} = I \end{aligned}$$

Similarly, $(A_k^{-1}A_{k-1}^{-1}\dots A_1^{-1})(A_1A_2\dots A_{k-1}A_k) = I$. Thus, $(A_k^{-1}A_{k-1}^{-1}\dots A_1^{-1})$ and $(A_1A_2\dots A_{k-1}A_k)$ are inverses of each other. □

Theorem 1.16 (Cancellation law for invertibles)**.** Let A be an invertible matrix and B, C be other matrices.

(i). If $AB = AC$, then $B = C$.

(ii). If $AB = 0$, then $B = 0$

Proof. (i) If $AB = AC$ and A^{-1} exists, multiply both sides by A^{-1} at the left to get $A^{-1}AB = A^{-1}AC$ $\Rightarrow IB = IC \ \Rightarrow B = C$

(ii) If $AB = 0$ and A^{-1} exists, then $A^{-1}AB = A^{-1}0 \ \Rightarrow \ B = 0$ □

1.4.5 Block matrix (Extra)

Block Notation: Similar to column vector notation, let $\begin{bmatrix} A & B \end{bmatrix}$ denote the matrix formed by joining A and B side by side.

eg. If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, then $\begin{bmatrix} A & B \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \end{bmatrix}$

$\begin{bmatrix} A & B \end{bmatrix}$ is called a **block matrix**, consisting of **blocks** A and B. It is also called **partitioned matrix** if we view $\begin{bmatrix} A & B \end{bmatrix}$ as a partition into A and B.

Similarly, we also use $\begin{bmatrix} A \\ B \end{bmatrix}$ to denote the matrix formed by joining A on top of B.

Similarly, we can use $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ to denote the matrix formed by joining A, B, C, D in that position. Note that the blocks A, B, C, D must fit together to form a rectangle.

Example: (The lines be are indicating the boundary of partition.)

$$\text{Let } M = \left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 4 & 5 & 6 & 0 \\ 7 & 8 & 9 & 1 \\ \hline 0 & 1 & 2 & 0 \end{array}\right] = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]$$

Then $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $C = \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$, $D = [0]$.

Sometimes, it is convenient to use block matrices, since block matrix multiplication works the same as ordinary matrix multiplication but with the ordinary entries replaced with blocks, provided that each product of the smaller matrices is defined.

Example:

$$\text{Let } A = \left[\begin{array}{ccc|cc} 2 & -3 & 1 & 0 & -4 \\ 1 & 5 & -2 & 3 & -1 \\ \hline 0 & -4 & -2 & 7 & -1 \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \left[\begin{array}{cc} 6 & 4 \\ -2 & 1 \\ -3 & 7 \\ \hline -1 & 3 \\ 5 & 2 \end{array}\right] = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}. \text{ Then}$$

$$AB = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} A_{11}B_1 + A_{12}B_2 \\ A_{21}B_1 + A_{22}B_2 \end{bmatrix} = \left[\begin{array}{cc} -5 & 4 \\ -6 & 2 \\ \hline 2 & 1 \end{array}\right]$$

Given some specific partition of A and B, we say that A and B are **block compatible** if A and B can be block-multiplied.

Theorem 1.17. If matrices A and B are partitioned compatibly into blocks, the product AB can be computed by matrix multiplication using blocks as entries.

Proof. Let $A = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & & \vdots \\ A_{m1} & \dots & A_{mn} \end{bmatrix}$ consist of mn blocks, where A_{ij} are blocks of A.

Similarly let $B = \begin{bmatrix} B_{11} & \dots & B_{1p} \\ \vdots & & \vdots \\ B_{n1} & \dots & B_{mp} \end{bmatrix}$ consist of np blocks.

Let $[AB]_{ij}$ denote the (i, j) th block of AB, while $(AB)_{st}$ denote the (s, t) th entry of AB.

Note that A_{ik} has the same number of columns for all $k = 1, \dots, m$, and A_{kj} has the same number of rows for all $k = 1, \dots, n$.

Let A be $(m_1 + m_2 + \ldots + m_m) \times (n_1 + n_2 + \ldots + n_n)$ where each m_i, n_j is the number of rows, columns of A_{ij} respectively. Let B be $(n_1 + n_2 \ldots + n_n) \times (p_1 + p_2 + \ldots + p_p)$ where each n_i, p_j is the number of rows, columns of B_{ij} respectively.

$$AB = \left[\begin{array}{ccccc} A_{11} & \dots & A_{1k} & \dots & A_{1n} \\ \vdots & & \vdots & & \vdots \\ \hline A_{i1} & \dots & A_{ik} & \dots & A_{in} \\ \hline \vdots & & \vdots & & \vdots \\ A_{m1} & \dots & A_{mk} & \dots & A_{mn} \end{array}\right] \left[\begin{array}{cc|c|cc} B_{11} & \dots & B_{1j} & \dots & B_{1p} \\ \vdots & & \vdots & & \vdots \\ B_{k1} & \dots & B_{kj} & \dots & B_{kp} \\ \vdots & & \vdots & & \vdots \\ B_{n1} & \dots & B_{nj} & \dots & B_{np} \end{array}\right]$$

By matrix multiplication using blocks as entries, the (i, j)-th block of AB is

$$[AB]_{ij} = A_{i1}B_{1j} + \ldots + A_{ik}B_{kj} + \ldots + A_{in}B_{nj}$$

By ordinary matrix multiplication, the (s, t)-th entry of $[AB]_{ij}$ is

$$\begin{aligned} ([AB]_{ij})_{st} &= (A_{i1}B_{1j})_{st} + \ldots + (A_{ik}B_{kj})_{st} + \ldots + (A_{in}B_{nj})_{st} \\ &= \sum_{l=1}^{n_1} (A_{i1})_{sl}(B_{1j})_{lt} + \ldots + \sum_{l=1}^{n_k} (A_{ik})_{sl}(B_{kj})_{lt} + \ldots + \sum_{l=1}^{n_n} (A_{in})_{sl}(B_{nj})_{lt} \end{aligned} \quad (1)$$

Note that every summation ($\sum$) of RHS takes place in the same row of A (call it row s_A) and the same column of B (call it row t_B). The entries taken from s_A th row of A run from left to right without gap, and the entries taken from t_B th column of B run from up to down without gap.

To elaborate, let's write out the entries of the $A_{ik} = [a_{st}]$ and $B_{kj} = [b_{st}]$ to help us visualize:

$$A_{ik}B_{kj} = \begin{bmatrix} a_{11} & \dots & a_{1l} & \dots & a_{1n_k} \\ \vdots & & \vdots & & \vdots \\ a_{s1} & \dots & a_{sl} & \dots & a_{sn_k} \\ \vdots & & \vdots & & \vdots \\ a_{m_i1} & \dots & a_{m_il} & \dots & a_{m_in_k} \end{bmatrix} \begin{bmatrix} b_{11} & \dots & b_{1t} & \dots & b_{1p_j} \\ \vdots & & \vdots & & \vdots \\ b_{l1} & \dots & b_{lt} & \dots & b_{lp_j} \\ \vdots & & \vdots & & \vdots \\ b_{n_k1} & \dots & b_{n_kt} & \dots & b_{n_kp_j} \end{bmatrix}$$

When we proceed to $A_{i,k+1}$, the corresponding shaded row will be aligned to the shaded row of A_{ik} in A, because A_{ik} and $A_{i,k+1}$ have the same number of rows. Also, the first entry of shaded row of $A_{i,k+1}$ is right next to the last entry of shaded row of A_{ik} in A, which means the shaded area doesn't have any gap in A. Similar reasoning can be said for B_{kj} and $B_{k+1,j}$.

Thus, the total of summation (LHS of eq.(1)) takes each entry in row s_A and row t_B exactly once, which is the same as multiplying AB ordinarily: (now a_{ij}, b_{ij} means (i,j)-entry of A, B respectively)

$$([AB]_{ij})_{st} = \sum_{k=1}^{n_1+\ldots+n_n} a_{s_Ak}b_{kt_B} = (AB)_{s_At_B}$$

Thus, block multiplication is the same as ordinary entry multiplication.

$\square$

1.5 Elementary row operations

1.5.1 Definition

Definition 1.25. Three types of **elementary row operations** can be performed to modify a matrix:

1. (Scaling) Multiply all entries in a row by a **non-zero** constant.
2. (Interchange) Interchange two rows.
3. (Replacement) Replace one row by the sum of itself and a multiple of another row.

(This is related to how we can add equations and multiply equation by constant in a system of equation. More on that later.) Let R_p denote the p th row, and $\to$ denote "becomes".

Example:

1. $\begin{bmatrix} 1 & 4 & 5 \\ 2 & -3 & 4 \\ 4 & 5 & -1 \end{bmatrix} \xrightarrow{R_3 \to 3R_3} \begin{bmatrix} 1 & 4 & 5 \\ 2 & -3 & 4 \\ 12 & 15 & -3 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 4 & 5 \\ 2 & -3 & 4 \\ 4 & 5 & -1 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 2 & -3 & 4 \\ 1 & 4 & 5 \\ 4 & 5 & -1 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 4 & 5 \\ 2 & -3 & 4 \\ 4 & 5 & -1 \end{bmatrix} \xrightarrow{R_2 \to R_2 - 2R_1} \begin{bmatrix} 1 & 4 & 5 \\ 2-2(1) & -3-2(4) & 4-2(5) \\ 4 & 5 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 5 \\ 0 & -11 & -6 \\ 4 & 5 & -1 \end{bmatrix}$

Definition 1.26. Two matrices are **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into another.

We write $A \sim B$ or $A \to B$ to denote A is row equivalent to B.

Theorem 1.18. Row equivalence satisfies the following properties:

1. $A \sim A$ (reflexive property)
2. If $A \sim B$, then $B \sim A$ (symmetric property)
3. If $A \sim B$ and $B \sim C$, then $A \sim C$ (transitive property)

Proof. (i) We define it so that doing nothing is still a sequence (of length zero) of elementary row operations.

(ii) Note that every elementary row operation has an inverse, i.e. one that undoes the elementary row operation:

Inverse of $R_p \to R_p + kR_q$ is $R_p \to R_p - kR_q$, since $(a_{pj} + ka_{qj}) - ka_{qj} = a_{pj}$

Inverse of $R_p \leftrightarrow R_q$ is $R_p \leftrightarrow R_q$, since interchanging something twice returns it to the original position.

Inverse of $R_p \to kR_p$ is $R_p \to \frac{1}{k}R_p$, since $(\frac{1}{k})(ka_{pj}) = a_{pj}$

This means a sequence of elementary row operations also has an inverse. If A is transformed to B by elementary row operations $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$, then we can use $\mathcal{E}_k^{-1}, \ldots, \mathcal{E}_2^{-1}, \mathcal{E}_1^{-1}$ to transform B back to A.

(iii) If $A \sim B$ and $B \sim C$, then we can first transform A to B and then transform B to C, which counts as transforming A to C.

□

This means row equivalence is an **equivalence relation**.

1.5.2 Elementary matrix

Definition 1.27. An **elementary matrix** is a matrix obtained by performing a single elementary row operation on an identity matrix.

It is typically denoted by E, and the three types of elementary row operations available give three types elementary matrix: type I (scaling), type II (interchange), type III (replacement).

Example:

1. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_3 \to 5R_3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_3 \to R_3 - 4R_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$

Notation: Let $\mathcal{E}(A)$ non-standardly denote the matrix obtained by performing the elementary row operation $\mathcal{E}$ on A.

Theorem 1.19 (Elementary matrix theorem)**.** If an elementary row operation is performed on an $m \times n$ matrix A, the result is EA where E is the elementary matrix obtained by performing the same row operation on I_m.

i.e. $\mathcal{E}(A) = \mathcal{E}(I_m)A$

Example: Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$, $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Then

$E_1A = \begin{bmatrix} a & b & c \\ d & e & f \\ g - 4a & h - 4b & i - 4c \end{bmatrix}$, $E_2A = \begin{bmatrix} d & e & f \\ a & b & c \\ g & h & i \end{bmatrix}$, $E_3A = \begin{bmatrix} a & b & c \\ d & e & f \\ 5g & 5h & 5i \end{bmatrix}$

Proof. Let $\mathbf{e}_i$ denote the i th row of I_m. Let $A = [a_{ij}]$.

Note that each row in A that remain unchanged after an elementary row operation is equal to the corresponding row in EA, because for i =unchanged row, $(EA)_{ij} = \sum_{k=1}^{n} (\mathbf{e}_i)_k a_{kj} = 0 + \mathbf{e}_{ii}a_{ij} + 0 = a_{ij}$, which is entry of A. So let's focus on the rows that get changed.

Note that each row of EA is equal to the corresponding row of E times A. i.e. $\text{row}_i(EA) = \text{row}_i(E)\,A$ by Theorem 1.6. Let's consider each elementary row operation.

1. Scaling operation:

Suppose row m is scaled by k. That is, $R_m \to kR_m$.

Then $\text{row}_m(E) = k\mathbf{e}_m$, and

$\text{row}_m(EA) = \text{row}_m(E)A = k\mathbf{e}_m A = k \cdot \text{row}_m(A)$ (where $\mathbf{e}_m A = \text{row}_m(A)$ by Theorem 1.7).

Thus, EA is the result of multiplying row m by k, as required.

2. Interchange operation:

Suppose row r and row m are interchanged (swapped). That is, $R_r \leftrightarrow R_m$.

Then $\text{row}_r(E) = \mathbf{e}_m$ and $\text{row}_m(E) = \mathbf{e}_r$, so

$\text{row}_r(EA) = \text{row}_r(E)A = \mathbf{e}_m A = \text{row}_m(A)$

$\text{row}_m(EA) = \text{row}_m(E)A = \mathbf{e}_r A = \text{row}_r(A)$.

Thus, EA is the result of interchanging row r and row m of A, as required.

3. Replacement operation:[5]

Suppose the operation adds k times row m to row $r \neq m$. That is, $R_r \to R_r + kR_m$.

Then $\text{row}_r(E) = \mathbf{e}_r + k\mathbf{e}_m$, and

$\text{row}_r(EA) = (\mathbf{e}_r + k\mathbf{e}_m)A = \mathbf{e}_r A + k(\mathbf{e}_m A) = \text{row}_r(A) + k \cdot \text{row}_m(A)$.

Thus, EA is the result of adding k times row m of A to row r, as required.

□

Corollary 1.20. Every elementary matrix E is invertible, and the inverse of E is an elementary matrix that corresponds to the inverse of the elementary row operation that produces E.

Proof. Let F be the matrix that corresponds to the inverse of elementary row operation that produces E. Then F is an elementary matrix since inverse of each elementary row operation is also a single elementary row operation.

Since performing the inverse elementary row operation on E produces I, by the previous theorem we have $FE = I$. Note that performing the original elementary row operation on F also produces I, so we also have $EF = I$.

Thus, F is the inverse of E. □

A corollary of the elementary matrix theorem is:

Corollary 1.21 (Elementary matrix corollary)**.** A and B are row equivalent if and only if there is a sequence of elementary matrices $E_1, E_2, \ldots, E_k$ such that $B = E_k E_{k-1} \ldots E_1 A$.

Proof. If A and B are row equivalent then there exists a sequence of elementary row operations that transforms A to B. Let $\mathcal{E}_i(A)$ be the matrix after i th elementary row operation in the sequence and $E_1, E_2, \ldots, E_k$ be the elementary matrices that corresponds to these elementary row operations.

By elementary matrix theorem we have $\mathcal{E}_1(A) = E_1 A$, $\mathcal{E}_2(A) = E_2(E_1 A)$, $\ldots$, $\mathcal{E}_k(A) = E_k E_{k-1} \ldots E_1 A = B$.

Conversely, if $B = E_k E_{k-1} \ldots E_1 A$, then by the theorem, $A \sim E_1 A \sim E_2 E_1 A \sim \ldots \sim E_n E_{n-1} \ldots E_1 A = B$ □

1.5.3 Row echelon form

A zero row refers to a row that has all entries zero. A non-zero row refers to a row that has at least one non-zero entry.

A **leading entry** of a non-zero row refers to the leftmost non-zero entry in the row.

Definition 1.28. A matrix is in **row echelon form** (and called a row-echelon matrix) if it has the following three properties:

1. All non-zero rows are above any rows of all zeros.

2. Each leading entry is to the right of the leading entry of the row above it.

A matrix is in **reduced echelon form** (/reduced row echelon form / RREF) if it has the additional properties:

3. The leading entry in each non-zero row is 1.

4. Each leading 1 is the only non-zero entry in its column.

A zero matrix is defined to be in reduced row echelon form.

Examples of row echelon matrices:

$$\begin{bmatrix} 3 & 4 & -5 & 2 \\ 0 & 5 & 3 & 4 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 3 & 5 \\ 0 & 2 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Not row echelon matrices:

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$$

Reduced row echelon matrices:

$$\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 0 & 1 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Every matrix can be transformed into row echelon form and reduced row echelon form using elementary row operations (we'll prove this later). This process is called **row reduction**. There is an algorithm we can use to transform a matrix into row echelon form reliably, called **Gaussian algorithm**.

Process 1.29. Gaussian algorithm proceeds iteratively as follows:

1. Find any leftmost non-zero entry a in the matrix, and move the row containing a to the top position (via row interchange).

2. Use row replacement to make each entry below a become zero. (e.g. To make a_{p1} zero, we can use $R_p \to R_p - \frac{a_{p1}}{a} R_1$.)

3. This completes the top row, and all further row operations are performed on the remaining rows.

4. Repeat steps 1-3. Stop when no more rows remain or the remaining rows consist entirely of zeros.

Notation and terminology: The leading entries are also called **pivots**. A pivot is a non-zero number in a pivot position (position of leading entry) that is used to create zeros via row operations. When we say A has n pivots, we mean that the row echelon form of A has n leading entries.

Sometimes I can combine two elementary operations, eg. $R_1 \to kR_1 - mR_2$ represents $R_1 \to kR_1$ followed by $R_1 \to R_1 - mR_2$) Combined elementary row operations are simply called row operations (so the word "elementary" actually matters). When elementary row operation is used, we can omit the arrow and just write $R_1 - mR_2$ instead of $R_1 \to R_1 - mR_2$ to save space.

Let's illustrate this with an example:

$$A = \begin{bmatrix} 0 & 5 & -2 & 4 \\ 6 & -3 & 4 & 12 \\ 4 & 12 & -3 & 19 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{bmatrix} 6 & -3 & 4 & 12 \\ 0 & 5 & -2 & 4 \\ 4 & 12 & -3 & 19 \end{bmatrix} \xrightarrow{R_1 \to \frac{1}{6} R_1} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{2}{3} & 2 \\ 0 & 5 & -2 & 4 \\ 4 & 12 & -3 & 19 \end{bmatrix} \xrightarrow[R_2 \to \frac{1}{5} R_2]{R_3 \to R_3 - 4R_1}$$

$$\begin{bmatrix} 1 & -\frac{1}{2} & \frac{2}{3} & 2 \\ 0 & 1 & -\frac{2}{5} & \frac{4}{5} \\ 0 & 14 & -3 & 19 \end{bmatrix} \xrightarrow{R_3 - 14R_2} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{2}{3} & 2 \\ 0 & 1 & -\frac{2}{5} & \frac{4}{5} \\ 0 & 0 & \frac{13}{5} & \frac{39}{5} \end{bmatrix} \xrightarrow{R_3 \to \frac{5}{13} R_3} \begin{bmatrix} 1 & -\frac{1}{2} & \frac{2}{3} & 2 \\ 0 & 1 & -\frac{2}{5} & \frac{4}{5} \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

For hand computation in practice, we often don't make each row have leading 1s so as to avoid messy fractions and save time:

$$
\begin{aligned}
A &= \begin{bmatrix} 2 & 5 & -2 & 6 \\ 6 & -3 & 0 & 0 \\ 4 & 12 & -3 & 19 \end{bmatrix} \\
&\sim \begin{bmatrix} 2 & 5 & -2 & 6 \\ 0 & 18 & -6 & 18 \\ 0 & 2 & 1 & 7 \end{bmatrix} && \begin{matrix} (R_2 \to 3R_1 - R_2) \\ (R_3 \to R_3 - 2R_1) \end{matrix} \\
&\sim \begin{bmatrix} 2 & 5 & -2 & 6 \\ 0 & 18 & -6 & 18 \\ 0 & 0 & 15 & 45 \end{bmatrix} && (R_3 \to 9R_3 - R_2)
\end{aligned}
$$

After making row echelon form, we can further perform row reduce to make reduced row echelon form. I'll called it extended Gaussian algorithm.

Process 1.30. To transform a row echelon matrix into reduced row echelon form, begin with the rightmost leading entry of the matrix. Make the leading entry 1 by scaling operation, and then create zeros above it. Repeat the process by working to the left on each of the leading entries.

Let's use the example again:

$$
\begin{bmatrix} 1 & -\frac{1}{2} & \frac{2}{3} & 2 \\ 0 & 1 & -\frac{2}{5} & \frac{4}{5} \\ 0 & 0 & 1 & 3 \end{bmatrix} \xrightarrow[R_1 - \frac{2}{3}R_3]{R_2 + \frac{2}{5}R_3} \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix} \xrightarrow{R_1 + \frac{1}{2}R_2} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}
$$

As we can see, the final matrix is in reduced row echelon form.

Theorem 1.22 (Existence of reduced echelon form)**.** Each matrix is row equivalent to a reduced echelon matrix (and hence, a row echelon matrix).

Proof. (Let's copy from [6] since I'm tired of making up my own words.)

Existence:

Let A be an $m \times n$ matrix.

We prove this by mathematical induction on the number of columns of A. When A has one column, either A is the zero vector which is already in reduced row echelon form (RREF), or it has a non-zero entry a. Move a to the top row by interchanging rows and multiply the top row by $1/a$. For each of the entry a_{i1} where $2 \le i \le m$, use the elementary row operation $R_i \to R_i - a_{i1}R_1$ to turn it into zero (since $a_{i1} - a_{i1}(1) = 0$). The result is the vector with 1 at top and zeros elsewhere, which is in RREF.

For the inductive step, assume A is $m \times n$ and the statement is true for all matrices with $n-1$ columns. We then know that there is a series of row operations we can do to A that result in a matrix X whose first $n-1$ columns form a RREF matrix. Suppose the matrix formed by these $n-1$ columns has k rows of zeros at the bottom.

If the final column has zeros in all of its bottom k entries, the matrix is in reduced echelon form. If not, swap a non-zero entry to the top of these k rows, multiply that row so that the entry becomes 1, and use it as a pivot to eliminate all other non-zero entries in the final column. The result is in RREF.

(To elaborate, if the pivot is in row p, then for all other rows i, use $R_i \to R_i - a_{in}R_p$ to make the entries in final column zero. This leaves the first $n-1$ columns unchanged since first $n-1$ entries of R_p are all zero.)

By mathematical induction, for A with any column number, A can be row reduced into reduced echelon form.

□

Lemma 1.23. Every non-zero $n \times 1$ column vector is row equivalent to the $n \times 1$ vector with 1 at the top and zero elsewhere, and not row equivalent to the $n \times 1$ zero vector.

A zero vector is row equivalent to itself only.

Proof. Use the process in the proof above (base case) to reduce the non-zero vector into $\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$.

A non-zero vector cannot be row equivalent to the zero vector, since if it were, then the zero vector is also row equivalent to the non-zero vector, meaning there is a sequence of elementary row operations that transforms the zero vector into the non-zero vector, which is impossible (since $k0 = 0$, $0 \leftrightarrow 0$ does nothing, $0 + k0 = 0$).

Thus, a zero vector is row equivalent to itself only. $\square$

Note that the reduced row echelon form of a matrix is also necessarily unique, but the proof involves the concept of system of linear equations, so I had put it to 1.30.

1.6 System of linear equations

Definition 1.31. A **linear equation** in n **variables** is an equation that can be written in the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$$

where $x_1, \ldots, x_n$ are variables, and b and the **coefficients** $a_1, \ldots a_n$ are scalars in $\mathbb{F}$.

For now, let the constants and coefficients be all in $\mathbb{R}$.

Definition 1.32. A **system of linear equations** (/**linear system**) is a collection of one or more linear equations involving the same variables $x_1, \ldots, x_n$. A linear system of m equations in n variables can be written as:

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

A **solution** of the system is a list $(s_1, s_2, \ldots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \ldots, s_n$ are substituted for $x_1, \ldots, x_n$ respectively.

Example: Linear system in 2 equations, 2 unknowns:

$$\begin{aligned} 3x - 5y &= 1 \\ 2x + y &= 5 \end{aligned}$$

Then $(x, y) = (2, 1)$ is a solution since substituting values make both equation true: $3(2) - 5(1) = 1$ and $2(2) + 1 = 5$.

Linear system in 3 equations, 3 unknowns:

$$\begin{aligned} 3x - 5y + 2z &= 4 \\ -2x + 6y - 5z &= 7 \\ x - 4y + 4z &= 3 \end{aligned}$$

Then $(x, y, z) = (139, 115, 81)$ is a solution since $3(139) - 5(115) + 2(81) = 4$, $-2(139) + 6(115) - 5(81) = 7$, $139 - 4(115) + 4(81) = 3$.

A linear system can have infinite solutions. The system

$$\begin{aligned} x + y &= 1 \\ 2x + 2y &= 2 \end{aligned}$$

have infinite solutions since $(x, y) = (t, 1 - t)$ makes both equations true for all $t \in \mathbb{R}$. t is called a **free variable**.

Definition 1.33. The **solution set** of a system of equation is the set of all solutions of the system. Two systems of equations are **equivalent** if they have the same solution set.

The solution set of the above system can be written as $\{(t, 1 - t) : t \in \mathbb{R}\}$.

Let me use $\cap$ between two equations to denote they belong to the same linear system.

So the system $x + y = 1 \ \cap \ 2x + 2y = 2$ is equivalent to the system $5x + 5y = 5 \ \cap \ 7x + 7y = 7$.

1.6.1 Elementary operations

Theorem 1.24. The following operations, called **elementary operations**, can be performed on system of linear equations to produce equivalent systems that have the same solution set:

1. Interchange two equations.
2. Multiply one equation by a non-zero number.
3. Add a multiple of one equation to another equation.

These are analogous to the elementary row operations mentioned before.

Proof. Note that equation interchange has no effect since it just changes the order we write the equations.

Suppose a system of linear equation is linear equations is transformed into a new system by an elementary operation. Then every solution of the original system is automatically a solution of the new system because adding equations or multiplying an equation by a non-zero number always results in a valid equation.

(To elaborate, suppose $(s_1, \ldots, s_n)$ is a solution to the original system. If an equation of the system $a_1x_1 + \ldots + a_nx_n = b$ get multiplied k to produce the new system with the corresponding equation being $ka_1x_1 + \ldots + ka_nx_n = kb$, then since $a_1s_1 + \ldots + a_ns_n = b$ is true, $a_1ks_1 + \ldots + a_nks_n = kb$ is also true by multiplicative property of equality, so $(ks_1, \ldots, ks_n)$ is also a solution of the new system.

If a multiple of equation, say $kc_1x_1 + \ldots + kc_nx_n = kd$ is added to $a_1x_1 + \ldots + a_nx_n = b$ to produce $a_1x_1 + \ldots + a_nx_n + (kc_1x_1 + \ldots + kc_nx_n) = b + kd$, then note that $a_1s_1 + \ldots + a_ns_n + (kc_1s_1 + \ldots + kc_ns_n) = b + kd$ is still true because both sides get added by equal stuff, since $kc_1s_1 + \ldots + kc_ns_n = kd$ is true because it's within the original system. So $(s_1, \ldots, s_n)$ is also a solution of the new system.)

In the same way, each solution of the new system must be a solution to the original system because the original system can be obtained from the new one by the inverse of the original elementary operation (which is also an elementary operation).

Thus, the original and new systems have the same solution set. □

Definition 1.34. A linear system is called **consistent** if it has at least one solution, and **inconsistent** if it has no solution.

Example of inconsistent system: $x + y = 1 \ \cap \ x + y = 2$. We get the false equation "$1 = 0$" if we subtract one equation from another.

Theorem 1.25. If a sequence of elementary operations on a linear system produces a false equation that does not depend on any variables, then the system is inconsistent.

Proof. Suppose the original system has a solution. Then the new system containing the false equation also has the same solution (by previous theorem), meaning there is a way of substituting values to make all the equations in the system true, including the false equation, which is a contradiction. Thus the original system cannot have a solution. □

Theorem 1.26. A linear system has either no solution, exactly one solution, or infinitely many solutions.

Proof. Suppose a linear system has two distinct solutions $(s_1, \ldots, s_n)$ and $(t_1, \ldots, t_n)$. Let $a_{i1}x_1 + \ldots + a_{in}x_n = b_i$ be the i th equation of the system. Then

$$\begin{aligned} a_{i1}s_1 + \ldots + a_{in}s_n &= b_i && \text{(by definition)} \\ (a_{i1}s_1 + \ldots + a_{in}s_n) - (a_{i1}t_1 + \ldots + a_{in}t_n) &= b_i - b_i && \text{(since } a_{i1}t_1 + \ldots + a_{in}t_n = b_i) \\ a_{i1}(s_1 - t_1) + \ldots + a_{in}(s_n - t_n) &= 0 \\ ka_{i1}(s_1 - t_1) + \ldots + ka_{in}(s_n - t_n) &= 0 && \text{(multiply both sides by } k \in \mathbb{R}) \end{aligned}$$

Then for any $k \in \mathbb{R}$, $(s_1 + k(s_1 - t_1), \ldots, s_n + k(s_n - t_n))$ is also a solution since substituting it into the equation gives

$$\begin{aligned} a_{i1}(s_1 + k(s_1 - t_1)) + \ldots + a_{in}(s_n + k(s_n - t_n)) &= b_i \\ (a_{i1}s_1 + \ldots + a_{in}s_n) + (ka_{i1}(s_1 - t_1) \ldots + ka_{in}(s_n - t_n)) &= b_i \\ a_{i1}s_1 + \ldots + a_{in}s_n + 0 &= b_i \\ b_i &= b_i \end{aligned}$$

Since $(s_1, \dots, s_n)$ and $(t_1, \dots, t_n)$ are distinct solutions, there exists j where $1 \le j \le n$ such that $s_j \neq t_j$. This means $s_j - t_j \neq 0$.

Then for scalars $k_1 \neq k_2$, we have $k_1(s_j - t_j) \neq k_2(s_j - t_j)$ and thus $s_j + k_1(s_j - t_j) \neq s_j + k_2(s_j - t_j)$, meaning the solution is different, i.e. $(s_1 + k_1(s_1 - t_1), \dots, s_n + k_1(s_n - t_n)) \neq (s_1 + k_2(s_1 - t_1), \dots, s_n + k_2(s_n - t_n))$. Thus there are infinitely many solutions corresponding to infinitely many k. □

1.6.2 Augmented matrix

Given a linear system

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 3x_3 &= -6 \\ -3x_1 + 4x_3 &= 10 \end{aligned}$$

with the variables aligned in columns, we can record the coefficients in a matrix to form a **coefficient matrix**, with the "missing variables" seen as having a coefficient of 0:

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -3 \\ -3 & 0 & 4 \end{bmatrix}$$

We can also include the constants in the right hand side of equation to form an **augmented matrix**.

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ -3 & 0 & 4 & 10 \end{array}\right]$$

Note: The vertical line is for visual purpose only. An augmented matrix has no fundamental difference from a normal matrix.

Definition 1.35. Given a linear system

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

, the **coefficient matrix** of the system is

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & b_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & b_{mn} \end{bmatrix}$$

and the **augmented matrix** of the system is

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{bmatrix}$$

1.6.3 Gaussian elimination

To solve a linear system, we can perform elementary row operations on the corresponding augmented matrix to produce equivalent linear systems. In particular, we can transform the augmented matrix into row-echelon form using Gaussian algorithm and then solve the linear system easily. This process is called **Gaussian elimination**.

Process 1.36. To solve a linear system with Gaussian elimination, proceed as follows:

1. Transform the augmented matrix of the system to row-echelon form using Gaussian algorithm.
2. If a row $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & b \end{array}\right]$ appears where $b \neq 0$, the system is inconsistent and there are no solutions.
3. Otherwise, assign the non-leading variables* (if any) as parameters, and use the equations corresponding to the reduced row-echelon matrix to solve for the leading variables in terms of the parameters.

*A non-leading variable is a variable that corresponds to a non-leading entry.

Example:

No solution: Given the linear system:

$$\begin{aligned} 3x + y - \ \ 4z &= -1 \\ x \qquad + 10z &= 5 \\ 4x + y + \ \ 6z &= 1 \end{aligned}$$

Row reduce the corresponding augmented matrix:

$$\left[\begin{array}{ccc|c} 3 & 1 & -4 & -1 \\ 1 & 0 & 10 & 5 \\ 4 & 1 & 6 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{ccc|c} 1 & 0 & 10 & 5 \\ 3 & 1 & -4 & -1 \\ 4 & 1 & 6 & 1 \end{array}\right] \xrightarrow[R_3 - 4R_1]{R_2 - 3R_1} \left[\begin{array}{ccc|c} 1 & 0 & 10 & 5 \\ 0 & 1 & -34 & -16 \\ 0 & 1 & -34 & -19 \end{array}\right] \xrightarrow{R_3 - R_2} \left[\begin{array}{ccc|c} 1 & 0 & 10 & 5 \\ 0 & 1 & -34 & -16 \\ 0 & 0 & 0 & -3 \end{array}\right]$$

The bottom row gives $0 = -3$ which is false, so the system has no solution.

Unique solution: Given the linear system:

$$\begin{aligned} x_1 - 2x_2 + \ \ x_3 &= 0 \\ 2x_2 - 3x_3 &= -6 \\ -3x_1 \qquad\quad + 4x_3 &= 10 \end{aligned}$$

Row reduce the corresponding augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ -3 & 0 & 4 & 10 \end{array}\right] \xrightarrow{R_3 + 3R_1} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ 0 & -6 & 7 & 10 \end{array}\right] \xrightarrow{R_3 + 3R_2} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ 0 & 0 & -2 & -8 \end{array}\right]$$

The last augmented matrix corresponds to

$$\begin{aligned} x_1 - 2x_2 + \ \ x_3 &= 0 \\ 2x_2 - 3x_3 &= -6 \\ -\ 2x_3 &= -8 \end{aligned}$$

And thus $x_3 = 4$. Now we do **backward substitution**. Substitute value of x_3 into second equation to get

$$\begin{aligned} 2x_3 - 3(4) &= -6 \\ x_3 &= 3 \end{aligned}$$

Now substitute value of x_2 and x_3 into first equation to get

$$\begin{aligned} x_1 - 2(3) + 4 &= 0 \\ x_1 &= 2 \end{aligned}$$

Thus, $(x_1, x_2, x_3) = (2, 3, 4)$ is the solution to the linear system.

Infinite solutions: Given the linear system:

$$\begin{aligned} x_1 - 2x_2 - \ \ x_3 + 3x_4 &= 1 \\ 2x_1 - 4x_2 + \ \ x_3 \qquad\ &= 5 \\ x_1 - 2x_2 + 2x_3 - 3x_4 &= 4 \end{aligned}$$

Row reduce the corresponding augmented matrix:

$$\left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 2 & -4 & 1 & 0 & 5 \\ 1 & -2 & 2 & -3 & 4 \end{array}\right] \xrightarrow[R_3-R_1]{R_2-2R_1} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right] \xrightarrow[\frac{1}{3}R_2]{R_3-R_2} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The last augmented matrix corresponds to

$$\begin{aligned} x_1 - 2x_2 \quad + \ x_4 &= 2 \\ x_3 - 2x_4 &= 1 \\ 0 &= 0 \end{aligned}$$

Use the non-leading variables x_2 and x_4 as free variables: $x_2 = s$, $x_4 = t$, and solve for x_3 and x_1 in terms of s and t:

$$x_3 = 1 + 2t \quad \text{and} \quad x_1 = 2 + 2s - t$$

The solution is

$$\begin{cases} x_1 &= 2 + 2s - t \\ x_2 &= s \\ x_3 &= 1 + 2t \\ x_4 &= t \end{cases}$$

or we can write the solution set: $\{(2 + 2s - t,\ s,\ 1 + 2t,\ t) : s, t \in \mathbb{R}\}$

Alternatively, we can keep row reducing the augmented matrix into reduced row echelon form, and then find the solution directly. This alternative process is called **Gauss-Jordan elimination**.

Example: Given the linear system:

$$\begin{aligned} x_1 - 2x_2 + \ x_3 &= 0 \\ 2x_2 - 3x_3 &= -6 \\ -3x_1 \qquad + 4x_3 &= 10 \end{aligned}$$

Row reduce the corresponding augmented matrix into reduced echelon form:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ -3 & 0 & 4 & 10 \end{array}\right] \xrightarrow{R_3+3R_1} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ 0 & -6 & 7 & 10 \end{array}\right] \xrightarrow{R_3+3R_2} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & -6 \\ 0 & 0 & -2 & -8 \end{array}\right] \xrightarrow[\text{then } R_2+3R_3]{R_3 \to -\frac{1}{2}R_3} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & 0 & 6 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

$$\xrightarrow[R_2 \to \frac{1}{2}R_2]{R_1-R_3} \left[\begin{array}{ccc|c} 1 & -2 & 0 & -4 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{array}\right] \xrightarrow{R_1+2R_2} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{array}\right]$$

Thus $(x, y, z) = (2, 3, 4)$.

Theorem 1.27. Suppose a system of m equations in n variables is consistent, and the row echelon form of its corresponding augmented matrix has r non-zero rows. Then

(i). $r \leq n$

(ii). If $r = n$, then the system has a unique solution.

(iii). If $r < n$, then the system has infinitely many solution.

Proof. **(i)** (First ignore the assumption that the system is consistent.) Note that the augmented matrix is $m \times (n + 1)$, where the last column entries are constants on RHS of the equations. The rowe chelon matrix has r non-zero rows, which means it has r leading entries, and each leading entry in the row echelon matrix occupies a distinct column. Thus the row echelon augmented matrix of any linear system has at most $n + 1$ leading entries.

If $r = n + 1$, the row echelon augmented matrix (where each leading entry is scaled to 1) must be in the form (* means any number)

$$\left[\begin{array}{cccccc|c}1 & * & * & \dots & * & & * \\ 0 & 1 & * & \dots & * & & * \\ 0 & 0 & 1 & \dots & * & & * \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & & * \\ 0 & 0 & 0 & \dots & 0 & & 1 \\ \vdots & \multicolumn{5}{l}{\text{(0 rows if any)}} & \end{array}\right]$$

The last non-zero row corresponds to the equation $0 = 1$, which means the system is inconsistent (i.e. has no solutions).

Thus, if a system is consistent, it must be that $r \leq n$.

(ii) Suppose the system is consistent and $r = n$. Then the row-echelon matrix is in the form

$$\left[\begin{array}{cccccc|c}1 & * & * & \dots & * & & * \\ 0 & 1 & * & \dots & * & & * \\ 0 & 0 & 1 & \dots & * & & * \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & \dots & 1 & & * \\ \vdots & \multicolumn{5}{l}{\text{(0 rows if any)}} & \end{array}\right]$$

Let's use mathematical induction on n to show that the system has a unique solution. Let the row echelon augmented matrix be $[r_{ij}]$. If $n = 1$, then the row-echelon matrix is $\left[1 \mid r_{12}\right]$ which corresponds to $x_1 = r_{12}$, a unique solution.

Assume that any consistent system of $n-1$ variables with $n-1$ non-zero rows in the augmented row echelon matrix has a unique solution. When a consistent system has n variables and $r = n$, we have $x_n = r_{n,n+1}$, and we can substitute $r_{n,n+1}$ into x_n in all of the remaining $n-1$ equations to obtain

$$\begin{aligned} x_1 + r_{12}x_2 + \ldots + r_{1n}(r_{n,n+1}) &= r_{1,n+1} \\ x_2 + \ldots + r_{2n}(r_{n,n+1}) &= r_{2,n+1} \\ &\vdots \\ x_{n-1} + r_{n-1,n}(r_{n,n+1}) &= r_{n-1,n} \end{aligned}$$

Put all the constants into right hand side to obtain a system of n equations:

$$\begin{aligned} x_1 + r_{12}x_2 + \ldots &= r_{1,n+1} - r_{1n}(r_{n,n+1}) \\ x_2 + \ldots &= r_{2,n+1} - r_{2n}(r_{n,n+1}) \\ &\vdots \\ x_{n-1} &= r_{n-1,n} - r_{n-1,n}(r_{n,n+1}) \end{aligned}$$

This system corresponds to a new row echelon augmented matrix with $n-1$ variables and $n-1$ non-zero rows. By induction assumption, this system of $n-1$ equations has a unique solution. Since $x_n = r_{n,n+1}$ is the only possible solution to the n th equation, the system of n equation also has a unique solution.

By mathematical induction, if $r = n$, then the system has a unique solution.

(iii) Start from the n th equation. If it has non-leading variables, use them as free variables (say, $t_1, \ldots, t_k$), and express the leading variable x_{n-k} in terms of them. Substitute $x_{n-k}, t_1, \ldots, t_n$ into $x_{n-k}, \ldots, x_n$ of the equation one row above ($n-1$ th equation). If there are more non-leading variables, assign them as new free variables $t_{k+1}, \ldots$ or something. Repeat this process until every x_i is expressed in terms of free variables.

Note that there is at least one free variable since the row echelon augmented matrix has at least one non-leading entries (as $r < n$). Since free variable can be any real number, there are infinite many choices and thus infinitely many solution.

□

Theorem 1.28. If the row echelon augmented matrix of a linear system does not contain a row where all except the last entries are zero, then the system is consistent.

Proof. Let r be number of non-zero rows and n be number of variables. Note that $r \leq n$. If $r = n$, the system must have a unique solution by previous theorem. If $r < n$, repeating the process of the previous proof (iii) must give a solution since there is at least one free variable, and at least one more variable is involved for each substitution of rows, so there is no room for contradiction. (I can't think of a better reasoning now.) $\square$

Theorem 1.29. If a linear system has more variables than equations, then it either has no solution or infinite solution.

Proof. If a linear system has more variables than equations, elementary operations may still lead to false equation. Example: $x + y + z = 1 \ \cap \ x + y + z = 2$, which is two equations in three variables. Subtracting one from other gives $0 = 1$.

Let the linear system have m equations and n variables where $m < n$. If the row echelon augmented matrix does not contain a row of $\begin{bmatrix} 0 & 0 & \dots & 0 & | & b \end{bmatrix}$ ($b \neq 0$), then it is consistent. Moreover, the number of leading entries r is at most m, i.e. $r \leq m$ since the augmented matrix has only m rows, and each leading entry occupies a distinct row. Thus $r \leq m < n$ which means $r < n$, so the system has infinite solution. $\square$

Note: However, when a linear system has more or equal equations than variables, it may still have no solutions, one solution or infinitely many solutions.

To summarize the conditions for number of solutions of a linear system:

1. If the row echelon augmented matrix contains a row of $\begin{bmatrix} 0 & 0 & \dots & 0 & | & b \end{bmatrix}$ (where $b \neq 0$), then the system is inconsistent (the system has no solutions).
2. Otherwise, the system is consistent.
   (a) If the number of leading entries is same as the number of variables, then the system has a unique solution.
   (b) Otherwise, the number of leading entries is less than the number of variables, so the system has infinitely many solution.

Theorem 1.30 (Uniqueness of reduced echelon form)**.** Each matrix is row equivalent to only one row echelon matrix.

In other words, a matrix has a unique reduced echelon form.

Proof. **Uniqueness:** [7]

If a matrix is row reduced two reduced echelon matrices R and S, then we need to show $R = S$. Suppose $R \neq S$ to the contrary. Then select the leftmost column at which R and S differ and also select all leading 1 columns to the left of this column, resulting in two matrices R' and S'. For example, if

$$R = \begin{bmatrix} 1 & 2 & 0 & 3 & 5 \\ 0 & 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad S = \begin{bmatrix} 1 & 2 & 0 & 7 & 9 \\ 0 & 0 & 1 & 8 & 9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

then

$$R' = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad S' = \begin{bmatrix} 1 & 0 & 7 \\ 0 & 1 & 8 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that the leading 1 columns of R' abd S' above, without the zero rows at the bottom, form an identity matrix I_2.

In general, either

$$R' = \left[\begin{array}{c|c} I_n & \mathbf{r}' \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \qquad \text{and} \qquad S' = \left[\begin{array}{c|c} I_n & \mathbf{s}' \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$$

or

$$R' = \left[\begin{array}{c|c} I_n & \mathbf{0} \\ \hline & 1 \\ \mathbf{0} & \vdots \\ & 0 \end{array}\right] \qquad \text{and} \qquad S' = \left[\begin{array}{c|c} I_n & \mathbf{s}' \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$$

or

$$R' = \left[\begin{array}{c|c} I_n & \mathbf{r}' \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \qquad \text{and} \qquad S' = \left[\begin{array}{c|c} I_n & \mathbf{0} \\ \hline & 1 \\ \mathbf{0} & \vdots \\ & 0 \end{array}\right]$$

It follows that R' and S' are row equivalent since deletion of columns does not affect row equivalent (because each corresponding columns of R and S are row equivalent). Also, R' and S' are in reduced echelon form but $R' \neq S'$.

Now interpret R' and S' as augmented matrices. The system for R' either has a unique solution $\mathbf{r}'$ or is inconsistent (it cannot have infinitely many solutions since it has no non-leading variables). Similarly, the system for S' either has a unique solution $\mathbf{s}'$ or is inconsistent.

Since $R' \sim S'$, the systems for R' and S' are equivalent (by Theorem 1.24), which means they have the same solutions . If both systems have unique solutions, then $\mathbf{r}' = \mathbf{s}'$, but this contradicts the above construction that the last column of R' and S' are different. If one system has unique solution and one is inconsistent, this also contradicts the fact that the two systems have the same solutions.

Either way there is a contradiction so we must have $R = S$.

□

1.6.4 Matrix vector equation

Sometimes it is convenient for us to view a system of linear equation as a product of matrix and vector.

Definition 1.37. A **matrix vector equation** is of the form

$$A\mathbf{x} = \mathbf{b}$$

where A is an $m \times n$ coefficient matrix, $\mathbf{x}$ is the $n \times 1$ variable vector, and $\mathbf{b}$ is a $m \times 1$ constant vector.

Theorem 1.31. Every linear system in the simultaneous equation form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_n \end{aligned}$$

is equivalent to the matrix vector equation

$$A\mathbf{x} = \mathbf{b}$$

where $A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$ is the coefficient matrix, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$.

In other words, the simultaneous equation form and the matrix vector equation form have the same solutions.

Proof. Since A is $m \times n$ and $\mathbf{x}$ is $n \times 1$, $\mathbf{b}$ is $m \times 1$. By definition of matrix product,

$$(A\mathbf{x})_i = \sum_{k=1}^{n} a_{ik}x_k = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n$$

. Writing this out as column vector, we have

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Two vectors are equal if and only if all of their entries are equal. Thus, we get a system of m equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_n \end{aligned}$$

□

Theorem 1.32. Let a linear system of n equations in n variables be in the form

$$A\mathbf{x} = \mathbf{b}$$

If the coefficient matrix A is invertible, then the system has the unique solution

$$\mathbf{x} = A^{-1}\mathbf{b}$$

Proof. Since A^{-1} exists, $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution to $A\mathbf{x} = \mathbf{b}$, as $A(A^{-1}\mathbf{b}) = I\mathbf{b} = \mathbf{b}$. So at least one solution exists.

To show uniqueness of solution, suppose $\mathbf{u}$ is a solution to the linear system. We can multiply both sides of the equation by A^{-1} at the left to get

$$\begin{aligned} A^{-1}A\mathbf{u} &= A^{-1}\mathbf{b} \\ I\mathbf{u} &= A^{-1}\mathbf{b} \\ \mathbf{u} &= A^{-1}\mathbf{b} \end{aligned}$$

Thus, if $\mathbf{x}$ is any solution, it must be in the form of $A^{-1}\mathbf{b}$. Since the inverse of A is unique, the solution vector $\mathbf{x} = A^{-1}\mathbf{b}$ is necessarily unique as well.

□

1.6.5 Homogeneous equations

Definition 1.38. A linear system is called **homogeneous** if all of its constant terms are zero:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= 0 \end{aligned}$$

Written as matrix vector equation:

$$A\mathbf{x} = \mathbf{0}$$

Definition 1.39. A **trivial solution** to a homogeneous linear system is $x_1 = 0$, $x_2 = 0$, …, $x_n = 0$. A **non-trivial solution** is a solution in which at least one variable has a non-zero value.

We can check that the trivial solution is indeed a solution:

$$\begin{aligned} a_{11}0 + a_{12}0 + \ldots + a_{1n}0 &= 0 \\ a_{21}0 + a_{22}0 + \ldots + a_{2n}0 &= 0 \\ &\vdots \\ a_{m1}0 + a_{m2}x_2 + \ldots + a_{mn}0 &= 0 \end{aligned}$$

Theorem 1.33. Every homogeneous linear system is consistent. If a homogeneous system has a unique solution, then it must be the trivial solution.

Proof. Note that $x_1 = \ldots = x_n = 0$ is always a solution to any homogeneous linear system. If a homogeneous system has a unique solution, then it cannot have any other solution except the trivial solution. $\square$

Theorem 1.34. If a homogeneous linear system has more variables than equations, then it has infinitely many non-trivial solution.

Proof. First, note that for a homogeneous linear system, the entries in the rightmost column of the augmented matrix are all zero after any elementary row operations (since $0 + k0 = 0$, $0 \leftrightarrow 0$ is still 0, and $k0 = 0$).

By Theorem 1.29, a linear system with more variables than equations has either no solution or infinitely many solution. But a homogeneous linear system cannot have no solutions since the row echelon augmented matrix cannot contain a row of $\left[\begin{array}{cccc|c} 0 & 0 & \ldots & 0 & b \end{array}\right]$ $(b \neq 0)$. Thus, it must have infinitely many solution. $\square$

1.7 Invertible matrix theorem

Theorem 1.35 (Invertible Matrix Theorem part I)**.** The following conditions are equivalent for an $n \times n$ matrix A:

1. A is invertible.
2. The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
3. A is row equivalent to the identity matrix I_n.
4. The system $A\mathbf{x} = \mathbf{b}$ has at least one solution $\mathbf{b}$ for every choice of column $\mathbf{b}$.
5. There exists an $n \times n$ matrix C such that $AC = I_n$.

Proof. (Copied from [8])

We show that each of these conditions implies the next, and that (5) implies (1).

(1) $\Rightarrow$ (2): If A^{-1} exists, then $A\mathbf{x} = \mathbf{0}$ gives $\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}$.

(2) $\Rightarrow$ (3): Assume (2) is true. Let R be the reduced echelon form of A. It suffices to show that $R = I_n$. Suppose this is not the case. Then R has a row of zero (since R is square matrix). Now consider the augmented matrix $\left[A \mid \mathbf{0}\right]$ of the system $A\mathbf{x} = \mathbf{0}$. Then $\left[A \mid \mathbf{0}\right] \longrightarrow \left[R \mid \mathbf{0}\right]$ is the reduced form, and $\left[R \mid \mathbf{0}\right]$ also has a row of zeros. Since R is square there must be at least one non-leading variable, and hence at least one parameter. Thus the system $A\mathbf{x} = \mathbf{0}$ has infinitely many solutions, contrary to (2). So $R = I_n$ must be true.

(3) $\Rightarrow$ (4): Consider the augmented matrix $\left[A \mid \mathbf{b}\right]$ of the system $A\mathbf{x} = \mathbf{b}$. Using (3), let $A \longrightarrow I_n$ by a sequence of elementary row operations. Then these same operations bring $\left[A \mid \mathbf{b}\right] \longrightarrow \left[I_n \mid \mathbf{b}\right]$ for some column $\mathbf{c}$. Hence the system $A\mathbf{x} = \mathbf{b}$ has a solution (in fact unique) by Gaussian elimination. This proves (4).

(4) $\Rightarrow$ (5): Write $I_n = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \end{bmatrix}$ where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the columns of I_n. For each $j = 1, 2, \ldots, n$, the system $A\mathbf{x} = \mathbf{e}_j$ has a solution $\mathbf{c}_j$ by (4), so $A\mathbf{c}_j = \mathbf{e}_j$. Now let $C = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n \end{bmatrix}$ be the $n \times n$ matrix with these $\mathbf{c}_j$ as its columns. Then Theorem 1.5 (column of matrix product) gives (5):

$$AC = A\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \ldots & \mathbf{c}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{c}_1 & A\mathbf{c}_2 & \ldots & A\mathbf{c}_n \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \end{bmatrix} = I_n$$

(5) $\Rightarrow$ (1): Assume that (5) is true so that $AC = I_n$ for some matrix C. Then the system $C\mathbf{x} = \mathbf{0}$ only has the solution $\mathbf{x} = 0$ (because $\mathbf{x} = I_n\mathbf{x} = AC\mathbf{x} = A\mathbf{0} = \mathbf{0}$). Thus condition (2) holds for the matrix C rather than A. Hence the argument above that (2) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5) shows that a matrix C' exists such that $CC' = I_n$. But then

$$A = AI_n = A(CC') = (AC)C' = I_nC' = C'$$

Thus $CA = CC' = I_n$ which, together with $AC = I_n$, shows that C is the inverse of A. This proves (1). $\square$

(5) $\Rightarrow$ (1) shows that left/right inverse is equivalent to 'full inverse'. Thus, when we want to check if C is the inverse of A, we only need to check either $AC = I$ or $CA = I$, and the other will be automatically true.

Corollary 1.36 (left/right inverse implies inverse)**.** If A and C are square matrices such that $AC = I$, then $CA = I$, and A, C are inverses of each other. i.e. $A^{-1} = C$ and $C^{-1} = A$.

Corollary 1.37. The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if A is invertible.

Proof. [9] If A is invertible, then $\mathbf{x} = A^{-1}\mathbf{b}$ so $\mathbf{x}$ is necessarily unique.

We now show that if $A\mathbf{x} = \mathbf{b}$ has a unique solution, then A is invertible. Let $\mathbf{u}$ be the unique solution. Suppose A is not invertible. Then by item (i) and (ii) of Invertible Matrix Theorem (1.35), the homogeneous system $A\mathbf{x} = \mathbf{0}$ has a non-trivial solution, call it $\mathbf{z}$. We have $\mathbf{z} \neq 0$.

Let $\mathbf{y} = \mathbf{u} - \mathbf{z}$ so that $\mathbf{z} = \mathbf{u} - \mathbf{y}$. Since $\mathbf{z}$ is non-zero, $\mathbf{u} \neq \mathbf{y}$. Then

$$A\mathbf{z} = 0 \implies A(\mathbf{u} - \mathbf{y}) = 0 \implies A\mathbf{u} - A\mathbf{y} = 0 \implies A\mathbf{u} = A\mathbf{y}$$

Since $A\mathbf{u} = \mathbf{b}$, we also have $A\mathbf{y} = \mathbf{b}$. But $\mathbf{u}$ is the unique solution to $A\mathbf{x} = \mathbf{b}$, so $\mathbf{u} = \mathbf{y}$ necessarily, contradicting $\mathbf{u} \neq \mathbf{y}$.

Thus, A must be invertible. □

Theorem 1.38. A square matrix is invertible if and only if it is a product of elementary matrices.

Proof. If a matrix A is invertible, then by item (3) of Invertible Matrix Theorem (1.35), A is row equivalent to I. This means I is also row equivalent to A. By elementary matrix corollary (1.21), there is a sequence of elementary matrices $E_1, E_2, \ldots, E_k$ such that

$$A = E_k \ldots E_2 E_1 I = E_k \ldots E_2 E_1$$

Conversely, if $A = E_k \ldots E_2 E_1$, then $A = E_k \ldots E_2 E_1 I$ so by corollary 1.21, A is row equivalent to I, which means A is invertible. □

Theorem 1.39. Matrices A and B are row equivalent if and only if there exists an invertible matrix P such that $B = PA$.

Proof. If A and B are row equivalent, then $B = E_k E_{k-1} \ldots E_1 A$ for some elementary matrices $E_1, \ldots E_k$ by elementary matrix corollary (1.21). Since elementary matrices are invertible (by corollary 1.20), their product $E_k E_{k-1} \ldots E_1$ is also invertible by Theorem 1.15 (generalized reversed product rule). Thus there exists an invertible matrix $P = E_k E_{k-1} \ldots E_1$ such that $B = PA$.

Conversely, if there exists invertible matrix P such that $B = PA$, then by the previous theorem, invertible matrix P can be expressed as a product of elementary matrices, say $P = E_k E_{k-1} \ldots E_1$. Thus we can write $B = E_k E_{k-1} \ldots E_1 A$, so by elementary matrix corollary, A and B are row equivalent. □

1.8 Determinant

The determinant is a useful function that takes in a matrix and outputs a number. It can be used to determine whether a matrix has inverse, and much more other stuff.

1.8.1 Definition

Notation:

Let A be a matrix. Let $A_{(ij)}$ denote the submatrix obtained by deleting the i th row and j th column of A. (This notation with mini bracket is non-standard, as other sources often just omit the mini bracket, but it will be confused with the (i, j) th entry of A, so I add mini bracket.)

Example: Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$. Then

$$A_{(32)} = \begin{bmatrix} a_{11} & \cancel{a_{12}} & a_{13} & a_{14} \\ a_{21} & \cancel{a_{22}} & a_{23} & a_{24} \\ \cancel{a_{31}} & \cancel{a_{32}} & \cancel{a_{33}} & \cancel{a_{34}} \\ a_{41} & \cancel{a_{42}} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{13} & a_{14} \\ a_{21} & a_{23} & a_{24} \\ a_{41} & a_{43} & a_{44} \end{bmatrix}$$

Note: If A is 1×1, then $A_{(11)}$ is the "empty matrix", and we define $\det A_{(11)} = 1$.

Definition 1.40. The **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is a scalar denoted by $\det A$ (or $|A|$), defined recursively as:

$$\det A = \begin{cases} a_{11} & \text{if } n = 1 \\ \displaystyle\sum_{i=1}^{n} (-1)^{1+i} a_{i1} \det A_{(i1)} & \text{if } n \geq 2 \end{cases}$$

This definition is called **cofactor expansion** along the first column.
Determinants are only defined for square matrix, and every square matrix has a determinant.
We use vertical bar instead of square bracket to denote determinant of a matrix with written out entries:

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

Theorem 1.40. The determinant of a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is

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

The determinant of a 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ is

$$\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Proof. For 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, let's unpack the definition:

$$\begin{aligned} \det A &= \sum_{i=1}^{2} (-1)^{1+i} a_{i1} \det A_{(i1)} \\ &= (-1)^2 a_{11} \det A_{(11)} + (-1)^3 a_{21} \det A_{(21)} \\ &= a_{11} \det[a_{22}] - a_{21} \det[a_{12}] \\ &= a_{11}a_{22} - a_{12}a_{21} \qquad \text{(determinant of } 1 \times 1 \text{ matrix is the entry itself)} \end{aligned}$$

For 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, let's unpack the definition again:

$$\begin{aligned}
\det A &= \sum_{i=1}^{3} (-1)^{1+i} a_{i1} \det A_{(i1)} \\
&= (-1)^2 a_{11} \det A_{(11)} + (-1)^3 a_{21} \det A_{(21)} + (-1)^4 a_{31} \det A_{(31)} \\
&= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{21}(a_{12}a_{33} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22}) \\
&= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}
\end{aligned}$$

□

There are two ways to memorize the 3×3 determinant:

1. Rule of Sarrus:

Write the first and second column to the right of the matrix. Sum up the products along the blue lines, and subtract the products along the red line to get the determinant.

$$\begin{array}{ccc:cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array} \qquad \begin{aligned} \det A = \; & a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & -a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \end{aligned}$$

2. Cofactor expansion

Expand the determinant into 2×2 determinants along first row or first column like in the proof above. The two ways of cofactor expansion lead to the same determinant (which we will eventual show for $n \times n$ matrix). I expand along first row because it's slightly easier.

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

Example: $$\begin{aligned}
\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 7 \\ 8 & 9 & 15 \end{vmatrix} &= 1 \begin{vmatrix} 5 & 7 \\ 9 & 15 \end{vmatrix} - 2 \begin{vmatrix} 4 & 7 \\ 8 & 15 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 8 & 9 \end{vmatrix} \\
&= 5(15) - 7(9) - 2(4(15) - 7(8)) + 3(4(9) - 5(8)) = -8
\end{aligned}$$

For 4×4 determinant or larger, rule of Sarrus no longer works so we have to use cofactor expansion.

Definition 1.41. The (i, j)-**minor** of matrix A is

$$M_{ij} = \det A_{(ij)}$$

The (i, j)-**cofactor** of matrix A is

$$C_{ij} = (-1)^{i+j} \det A_{(ij)}$$

They can also be called the minor/cofactor of entry a_{ij}.

Note: The context should make it clear that M_{ij} and C_{ij} refers to minor and cofactor instead of (i, j) th entry of matrix named C or M. To specify it is the minor/cofactor of matrix A, we can write M_{Aij} and C_{Aij} . Note that each minor/cofactor is a scalar.

The signs of C_{ij} form a checkerboard pattern in the matrix positions:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

The definition of determinant (for $n \geq 2$) rewritten using minor or cofactor is

$$\det A = \sum_{i=1}^{n} (-1)^{1+i} a_{i1} M_{i1} \quad \text{or} \quad \det A = \sum_{i=1}^{n} a_{i1} C_{i1}$$

1.8.2 Effect of elementary row operations

Theorem 1.41 (Effect of elementary row operations on determinant)**.** For an $n \times n$ matrix,

(i). Multiplying one row by k multiplies the determinant by k.

(ii). Interchanging two rows multiplies the determinant by -1 (for $n \geq 2$).

(iii). Adding the multiple of one row to another row does not change the determinant.

Proof. **(i)** [10] Let $B = [b_{ij}]$ be the matrix obtained by multiplying row p of A by k. We want to show that $\det B = k \det A$ for all matrix sizes n.

We use mathematical induction on n. The statement is trivially true for $n = 1$.

For $n = 2$, let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $B = \begin{bmatrix} ka & kb \\ c & d \end{bmatrix}$. Then $\det A = ad - bc$ and $\det B = (ka)d - (kb)c = k(ad - bc) = k \det A$. If $B = \begin{bmatrix} a & b \\ kc & kd \end{bmatrix}$ instead, then $\det B = a(kd) - b(kc) = k(ad - bc) = k \det A$. Thus, the statement is true for $n = 2$. (It is actually not necessarily to check for $n = 2$ since $n = 1$ is a valid base case. I just check it to gain some intuition.)

Assume the statement is true when the matrix size is $n - 1$. When the matrix size is n, by definition,

$$\det A = \sum_{i=1}^{n} a_{i1} (-1)^{i+1} \det A_{(i1)}$$
$$\det B = \sum_{i=1}^{n} b_{i1} (-1)^{i+1} \det B_{(i1)}$$

Let's compare each i th term of both summations.

When $i \neq p$, then $b_{i1} = a_{i1}$. Note that $A_{(i1)}$ and $B_{(i1)}$ are matrices of sizes $(n-1)$, and $B_{(i1)}$ contains row p which is the row that got multiplied by k, so we can use induction assumption to have $\det B_{(i1)} = k \det A_{(i1)}$. Thus, $b_{i1}(-1)^{i+1} \det B_{(i1)} = a_{i1}(-1)^{i+1} k \det A_{(i1)}$.

When $i = p$, then $b_{i1} = ka_{i1}$ since b_{p1} is in row p, the row that got multiplied by k. Note that $A_{(i1)} = B_{(i1)}$ since the p th row of A and B gets deleted to form these matrices, and other rows remain unchanged. Thus, $b_{i1}(-1)^{i+1} \det B_{(i1)} = ka_{i1}(-1)^{i+1} \det A_{(i1)}$.

Since each term in the summation of $\det B$ is k times the corresponding term in the summation of $\det A$, we conclude $\det B = k \det A$ when the matrix size is n.

By mathematical induction, $\det B = k \det A$ is true for all matrix sizes n.

(ii) [11] First, let's consider interchanging **adjacent rows** because it's easier. Let $B = [b_{ij}]$ be the matrix obtained by interchanging row p and row $p + 1$ of A (where $1 \leq p \leq n - 1$). We want to show that $\det B = -\det A$.

We use mathematical induction again. For $n = 2$, let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then $\det A = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$ and $\det B = \begin{vmatrix} c & d \\ a & b \end{vmatrix} = bc - ad = -(ad - bc) = -\det A$. So the statement is true for $n = 2$.

Assume the statement is true for matrix size $n - 1$. Compare each i th term of summations of $\det A$ and $\det B$ again.

When $i \in \{1, \ldots, n\} \backslash \{p, p + 1\}$ (namely, $i \neq p$ and $i \neq p + 1$), then $a_{i1} = b_{i1}$, and note that $A_{(i1)}$ and $B_{(i1)}$ contain both interchanged rows, so by induction assumption, $\det B_{(i1)} = -\det A_{(i1)}$. Thus, $b_{i1}(-1)^{i+1} \det B_{(i1)} = a_{i1}(-1)^{i+1}(-\det A_{(i1)})$.

When $i \in \{p, p+1\}$, then $a_{p1} = b_{p+1,1}$ and $a_{p+1,1} = b_{p1}$ since row $p, p+1$ are rows that get interchanged, so we compare the p th term of $\det A$ with $p + 1$ th term of $\det B$, and $p + 1$ th term of $\det A$ with p th term of $\det B$. Let's visualize the matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \\ a_{p+1,1} & a_{p+1,2} & \cdots & a_{p+1,n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \qquad B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{p+1,1} & a_{p+1,2} & \cdots & a_{p+1,n} \\ a_{p1} & a_{p2} & \cdots & a_{pn} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

The shaded area is the entries involved in the p th term of $\det A$ and $p+1$ term of $\det B$. We see that $a_{p1} = b_{p+1,1}$, and also $A_{(p1)} = B_{(p+1,1)}$, since $A_{(p1)}$ is formed by deleting p th row of A which is the row with entries a_{pj}, and $B_{(p+1,1)}$ is formed by deleting $p+1$ th row of B, which is also the row with entries a_{pj} because $b_{p+1,j} = a_{pj}$ after row interchange. Thus, the only difference between the p th term of $\det A$ and $p+1$ th term of $\det B$ is in the power of negative one, in particular, $(-1)^{(p+1)+1} = -(-1)^{p+1}$.
By similar reasoning, we note that the p th term of $\det B$ is -1 times the $p+1$ th term of $\det A$.

Since each term in the summation of $\det B$ is -1 times some corresponding term in the summation of $\det A$, we conclude that $\det B = -\det A$ when the matrix size is n and adjacent rows are interchanged.

Now let's consider the general case of interchanging rows. Let row p and row q be interchanged in A, where $p < q$. We can do that by interchanging two adjacent rows $2(q-p)-1$ times: First swap row i and $i+1$, then row $i+1$ and $i+2$, and so on. After swapping row $q-1$ and q (now we have swapped $q-p$ times so far), we have original p th row in position of q th and original l th row in position of $l-1$ th for $p+1 \le l \le q$. Then proceed backwards swapping adjacent rows for $q-p-1$ until every row except row p and q is in original place.
Let's illustrate this with an example. To interchange 2nd and 5th row, we swap $(2,3) \to (3,4) \to (4,5) \to (3,4) \to (2,3)$:

$$\begin{bmatrix}1\\2\\3\\4\\5\\6\end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix}1\\3\\2\\4\\5\\6\end{bmatrix} \xrightarrow{R_3 \leftrightarrow R_4} \begin{bmatrix}1\\3\\4\\2\\5\\6\end{bmatrix} \xrightarrow{R_4 \leftrightarrow R_5} \begin{bmatrix}1\\3\\4\\5\\2\\6\end{bmatrix} \xrightarrow{R_3 \leftrightarrow R_4} \begin{bmatrix}1\\3\\5\\4\\2\\6\end{bmatrix} \xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix}1\\5\\3\\4\\2\\6\end{bmatrix}$$

Since $2(q-p)-1$ is an odd number, interchanging adjacent rows for $2(q-p)-1$ times will multiply the determinant by $(-1)^{2(q-p)-1} = -1$, and we have $\det B = -\det A$ when any two rows p and q are interchanged.

(iii) [10] Let's first prove a lemma:

Lemma 1.42. Let A, B and C be $n \times n$ matrices that are identical except that each entry in the p th row of A is the sum of the corresponding entries in p th rows of B and C. i.e. $\text{row}_p(A) = \text{row}_p(B) + \text{row}_p(C)$. Then

$$\det A = \det B + \det C$$

Proof of lemma: [10]

We use mathematical induction on n. The statement is trivially true for $n = 1$. Assume the statement is true for matrix size $n-1$. When the matrix size is n, consider a_{i1} and $A_{(i1)}$:

Case 1: If $i \neq p$,

$$a_{i1} = b_{i1} = c_{i1} \qquad \text{and} \qquad \det A_{(i1)} = \det B_{(i1)} + \det C_{(i1)}$$

, the latter by induction assumption because $A_{(i1)}, B_{(i1)}, C_{(i1)}$ are identical except that one row of $A_{(i1)}$ is the sum of the corresponding rows of $B_{(i1)}$ and $C_{(i1)}$.

Case 2: If $i = p$,

$$a_{p1} = b_{p1} + c_{p1} \qquad \text{and} \qquad A_{(p1)} = B_{(p1)} = C_{(p1)}$$

, the latter being true since the row p in A, B, C is deleted to form $A_{(i1)}, B_{(i1)}, C_{(i1)}$, and other rows remain unchanged. This means $\det A_{(p1)} = \det B_{(p1)} = \det C_{(p1)}$.

Now write out the defining sums for $\det A$, splitting off the p th term for special attention.

$$\begin{aligned}
\det A &= (\sum_{i \neq p} a_{i1}(-1)^{i+1} \det A_{(i1)}) + a_{p1}(-1)^{p+1} \det A_{(p1)} \\
&= \sum_{i \neq p} a_{i1}(-1)^{i+1} (\det B_{(i1)} + \det C_{(i1)}) + (b_{p1} + c_{p1})(-1)^{p+1} \det A_{(p1)} \qquad \text{(by Case 1 and Case 2)} \\
&= \sum_{i \neq p} a_{i1}(-1)^{i+1} \det B_{(i1)} + \sum_{i \neq p} a_{i1}(-1)^{i+1} \det C_{(i1)} + b_{p1}(-1)^{p+1} \det A_{(p1)} + c_{p1}(-1)^{p+1} \det A_{(p1)} \\
&= \sum_{i \neq p} b_{i1}(-1)^{i+1} \det B_{(i1)} + \sum_{i \neq p} c_{i1}(-1)^{i+1} \det C_{(i1)} + b_{p1}(-1)^{p+1} \det B_{(p1)} + c_{p1}(-1)^{p+1} \det C_{(p1)} \\
&\quad \text{(since } a_{i1} = b_{i1} = c_{i1} \text{ for } i \neq p) \qquad \qquad \text{(since } \det A_{(p1)} = \det B_{(p1)} = \det C_{(p1)}) \\
&= \left(\sum_{i \neq p} b_{i1}(-1)^{i+1} \det B_{(i1)} + b_{p1}(-1)^{p+1} \det B_{(p1)} \right) + \left(\sum_{i \neq p} c_{i1}(-1)^{i+1} \det C_{(i1)} + c_{p1}(-1)^{p+1} \det C_{(p1)} \right) \\
&= \det B + \det C
\end{aligned}$$

By mathematical induction, $\det A = \det B + \det C$ for all matrix sizes n. ■ (QED symbol)

We also need another lemma:

Lemma 1.43. If A contains two identical rows, then $\det A = 0$.

Proof. If rows p and q of A are identical, let B be obtained from A by interchanging these rows. Then $B = A$ so $\det A = \det B$. But $\det B = -\det A$ by property (ii) of Theorem 1.41, so $\det A = -\det A$. This implies $\det A = 0$ □

Lemma 1.44. If A contains a row that is a non-zero scalar multiple of another row, then $\det A = 0$.

Proof. Let A have another row p that is k times row q. Let B be the matrix obtained from A by multiplying row q by k.

Note that row p and row q of B are identical, so $\det B = 0$. By property (i), $\det B = k \det A$, so $\det A = \frac{1}{k}(0) = 0$.

□

Now we prove property (iii). Let B be the matrix obtained by adding k times row q of A to row p of A, and we want to show that $\det B = \det A$.

Let C be the matrix obtained from A by replacing row p with k times row q. Then since

$$\mathrm{row}_p(B) = \mathrm{row}_p(A) + k \cdot \mathrm{row}_q(A) = \mathrm{row}_p(A) + \mathrm{row}_p(C)$$

, Lemma 1.42 applies to B to show that $\det B = \det A + \det C$. Since one row of C is a scalar multiple of another, $\det C = 0$ by the previous Lemma. Thus, $\det A = \det B$, as desired.

□

1.8.3 Determinant properties

Theorem 1.45 (Conditions of zero determinant)**.** For an $n \times n$ matrix,

(i). A matrix with a row of zeros has zero determinant.

(ii). A matrix with a row that is a non-zero scalar multiple of another row has zero determinant.

(iii). A zero matrix has zero determinant.

Proof. (i) Let A be any square matrix. Let B be the matrix obtained by multiplying a row of A by 0. Then by property (i) of previous theorem, $\det B = 0 \det A = 0$.

(Note: Even though multiplying a row by 0 is not an elementary row operation, there is nothing in the proof of property (i) that stops us from substituting 0 for k. Thus the formula $\det B = k \det A$ still works for 0.)

(ii) This is just Lemma 1.44 restated.
(iii) A zero matrix contains at least a row of zeros, so by (ii), $\det \mathbf{0} = 0$. □

Theorem 1.46 (Determinant of identity matrix)**.** .

(i). $\det I = 1$

Proof. (i) Let $I = [e_{ij}]$. Note that $I_1 = 1$. Assume $I_{n-1} = 1$ for matrix size $n-1$. Note that $e_{i1} = 1$ if $i = 1$, else $e_{i1} = 0$. So

$$\begin{aligned}
\det I &= \sum_{i=1}^{n} e_{i1}(-1)^{i+1} \det I_{(i1)} \\
&= e_{11}(-1)^{1+1} \det I_{(11)} + e_{21}(-1)^{2+1} \det I_{(21)} + \ldots + e_{n1}(-1)^{n+1} \det I_{(n1)} \\
&= 1 \det I_{n-1} + 0 + \ldots + 0 \\
&= 1(1) = 1
\end{aligned}$$

□

Lemma 1.47 (Product rule for elementary matrices)**.** If A is a square matrix and E is an elementary matrix, then

$$\det(EA) = \det E \det A$$

Thus, if $E_1, E_2, \dots, E_k$ are all elementary matrices, then

$$\det(E_k \dots E_2 E_1 A) = \det E_k \dots \det E_2 \det E_1 \det A$$

Proof. Recall that by elementary matrix theorem (Theorem 1.19), if E is an elementary matrix obtained by doing one elementary row operation to I_n, then doing that operation to A results in EA.

By Theorem 1.41 (effect of row operation on det), the elementary row operation multiplies the determinant of the matrix operated on by either k, -1, or 1, let's say it is λ.
Then the matrix obtained after the elementary row operation on A, namely EA, has determinant $\lambda \det A$.
Similarly, the matrix obtained after the elementary row operation on I, namely E, has determinant $\lambda(1)$.

Thus, $\det(EA) = \lambda \det A = \det E \det A$, and by applying this repeatedly we also have,

$$\begin{aligned}\det(E_k \dots E_2 E_1 A) &= \det(E_k \dots E_2 E_1) \det A \\ &= \det(E_k \dots E_2) \det E_1 \det A \\ &= \dots \\ &= \det E_k \dots \det E_2 \det E_1 \det A\end{aligned}$$

□

Theorem 1.48 (Determinant determines invertibility)**.** Matrix A is invertible if and only if $\det A \neq 0$.

Proof. Let's break it down into two statements:

(i) If A is invertible, then $\det A \neq 0$.

(ii) If A is non-invertible, then $\det A = 0$.

Proof of (i):

If A is invertible, then A is row equivalent to I by item (3) of Invertible Matrix Theorem (Theorem 1.35). By elementary matrix theorem (Theorem 1.19), there is a sequence of elementary matrices $E_1, E_2, \dots, E_k$ such that $E_k \dots E_2 E_1 A = I$. Taking determinant on both sides, we have

$$\det(E_k \dots E_2 E_1 A) = \det I$$

By previous lemma (1.47), we can apply determinant product rule on elementary matrices, so

$$\det E_k \dots \det E_2 \det E_1 \det A = 1$$

Since the right hand side is 1, any factor on the left hand side cannot be 0, meaning $\det A \neq 0$.

Proof of (ii):

Let $A \longrightarrow R$ where R is reduced row echelon form. Then by elementary matrix theorem there is a sequence of elementary matrix $E_1, E_2, \dots, E_k$ such that $E_k \dots E_2 E_1 A = R$. By previous lemma (1.47), we have

$$\det E_k \dots \det E_2 \det E_1 \det A = \det R$$

If A is non-invertible, then R must have a row of zeros, because otherwise, we must have $R = I$ which means A would be row equivalent to I, and by Invertible Matrix Theorem this would mean A is invertible, contradicting the initial assumption that A is non-invertible. Since R has a row of zeros, $\det R = 0$ by Theorem 1.45 (conditions of zero determinant). So the equation above becomes

$$\det E_k \dots \det E_2 \det E_1 \det A = 0$$

Since right hand side is 0, at least one factor in left hand side is 0. However, note that the determinant of an elementary matrix cannot be zero, since elementary row operation either multiply determinant of I by non-zero k, -1 or 1. Thus, $\det A = 0$, and our statement is proven. □

Theorem 1.49 (Product rule of determinants)**.** If A and B are $n \times n$ matrices, then

$$\det(AB) = \det A \det B$$

Proof. Let's consider two cases.

Case 1: A has no inverse.

Then AB also has no inverse (otherwise $A(B(AB)^{-1}) = I$ which means A is invertible by Corollary 1.36). Hence the above theorem (determinant determines invertibility) gives

$$\det(AB) = 0 \quad \text{and} \quad \det A = 0$$

which means $\det(AB) = 0 \det B = \det A \det B$. So the property is satisfied.

Case 2: A has an inverse.

Then A is a product of elementary matrices by Theorem 1.38, say $A = E_1 E_2 \dots E_k$. Then product rule for elementary matrices (substituting "A" with I in equation of Lemma 1.47) gives

$$\det A = \det(E_1 E_2 \dots E_k) = \det E_1 \det E_2 \dots \det E_k$$

But then substituting "A" with B in Lemma 1.47 gives

$$\det(AB) = \det((E_1 E_2 \dots E_k)B) = \det E_1 \det E_2 \dots \det E_k \det B = \det A \det B$$

So the property also holds in this case.

□

Thus, by applying the product rule repeatedly, we also have

$$\det(A_1 A_2 \dots A_k) = \det A_1 \det A_2 \dots \det A_k$$

for square matrices $A_1, A_2, \dots, A_k$.

Lemma 1.50. If E is an elementary matrix, then $\det E^T = \det E$

Proof. Let $E = [e_{ij}]$ and $I = [u_{ij}]$. If E is obtained by scaling or interchanging rows of I, then $E^T = E$, since scaling only changes an entry on the main diagonal, so E is still symmetric. For interchanging rows, suppose row p and q are interchanged to give E. Then $e_{qp} = u_{pp} = 1$ and $e_{pq} = u_{qq} = 1$, which means $e_{pq} = e_{qp}$. All other entries outside main diagonal remain zero. Thus E is symmetric and $E^T = E$.

If E is obtained by adding a multiple of a row to another row of I (type III elementary matrix), then determinant remains unchanged by Theorem 1.41 so $\det E = \det I = 1$. Note that E^T is also of type III elementary matrix, since E differs from I by only one nonzero entry outside main diagonal, and E^T only flips this entry to the other side of main diagonal, which mean E^T can also be obtained by a single type III elementary row operation. So $\det E^T = \det E$.

Hence, $\det E^T = \det E$ for every elementary matrix E. □

Theorem 1.51. If A is a square matrix, then $\det A^T = \det A$.

Proof. Let A be a square matrix. If A is not invertible, then neither is A^T (by Corollary 1.14), so $\det A^T = \det A = 0$ by Theorem 1.48.

On the other hand, if A is invertible, then $A = E_k \dots E_2 E_1$, where the E_i are elementary matrices (by Theorem 1.38). Hence, $A^T = (E_k \dots E_2 E_1)^T = E_1^T E_2^T \dots E_k^T$ so the product rule gives

$$\begin{aligned}
\det A^T &= \det(E_1^T E_2^T \dots E_k^T) \\
&= \det E_1^T \det E_2^T \dots \det E_k^T \qquad \text{(product rule: Theorem 1.49)} \\
&= \det E_1 \det E_2 \dots \det E_k \qquad \text{(since } \det E^T = \det E) \\
&= \det E_k \dots \det E_2 \det E_1 \qquad \text{(rearrange)} \\
&= \det(E_k \dots E_2 E_1) \qquad \text{(product rule again)} \\
&= \det A
\end{aligned}$$

□

Let A be a square matrix. For every elementary row operation on row p and q of A, there is a corresponding elementary column operation on column p and q.

Corollary 1.52. The effect of elementary column operations on the determinant of a matrix is the same as the corresponding row operations.

Proof. Suppose B is obtained by performing an elementary column operation on A. Then B^T is obtained from A^T by the corresponding elementary row operation. Let the row operation multiplies the determinant by λ. Then by the above theorem,

$$\det B = \det B^T = \lambda \det A^T = \lambda \det A$$

Thus, elementary column operations have the same effect as elementary row operations. $\square$

Now here comes the almighty Cofactor Expansion Theorem which states that the determinant is the same by cofactor expansion along any rows or columns.

Theorem 1.53 (Cofactor Expansion Theorem)**.** If $A = [a_{ij}]$ is an $n \times n$ matrix, then

(i). $\det A = \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \det A_{(ij)}$ (cofactor expansion along column j)

(ii). $\det A = \sum_{j=1}^{n} a_{ij}(-1)^{i+j} \det A_{(ij)}$ (cofactor expansion along row i)

Proof. **(i)** Given an $n \times n$ matrix $A = [a_{ij}]$, let $B = [b_{ij}]$ be obtained by moving column j to the left side, using $j-1$ interchange of adjacent columns:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix} \longrightarrow B = \left[\begin{matrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{matrix} \; \underbrace{\begin{matrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{matrix}}_{\text{skips column } j} \right]$$

Then $\det B = (-1)^{j-1} \det A$ by previous corollary. Note that $B_{(i1)} = A_{(ij)}$ and $b_{i1} = a_{ij}$ for all i, as illustrated by the shaded area when $i = 1$:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & \dots & a_{2j} & \dots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{nj} & \dots & a_{nn} \end{bmatrix} \longrightarrow B = \left[\begin{matrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{matrix} \; \underbrace{\begin{matrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{matrix}}_{\text{skips column } j} \right]$$

Thus we have

$$\begin{aligned} \det A &= (-1)^{j-1} \det B \\ &= (-1)^{j-1} \sum_{i=1}^{n} b_{i1}(-1)^{i+1} \det B \\ &= \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \det A_{(ij)} \end{aligned}$$

This is the cofactor expansion of $\det A$ along column j.

(ii) To prove row expansion, redefine $B = A^T$. Then $b_{ij} = a_{ji}$ and $B_{(ij)} = (A_{(ji)})^T$ for all i, j (since $(A_{(ji)})^T$ is obtained by deleting row i and column j of A^T). Expanding $\det B$ along column j gives

$$\begin{aligned} \det A = \det A^T = \det B &= \sum_{i=1}^{n} b_{ij}(-1)^{i+j} \det B_{(ij)} \qquad \text{(expand along } i\text{th column of } B) \\ &= \sum_{i=1}^{n} a_{ji}(-1)^{j+i} \det (A_{(ji)})^T \\ &= \sum_{i=1}^{n} a_{ji}(-1)^{j+i} \det A_{(ij)} \end{aligned}$$

This is the required expansion of $\det A$ along row j (with the variables swapped from original formula). $\square$

Let's include the important determinant properties in one theorem.

Theorem 1.54 (Determinant properties)**.** Let A be an $n \times n$ matrix. Then

(i). $\det(kA) = k^n \det A$

(ii). $\det(A^{-1}) = \dfrac{1}{\det A}$ if A is invertible.

(iii). $\det A^T = \det A$

(iv). $\det(AB) = (\det A)(\det B)$

Proof. (i) kA multiplies each row of A by k, so it is equivalent to doing the elementary row operation of scaling a row by k for n times. Thus, by Theorem 1.41 (effect of row operation on det), $\det(kA) = k^n \det A$.

(ii)

$$\begin{aligned}
\det I &= 1 \\
\det(AA^{-1}) &= 1 \\
(\det A)\det(A^{-1}) &= 1 \qquad \text{(product rule)} \\
\det(A^{-1}) &= \frac{1}{\det A}
\end{aligned}$$

(iii) Theorem 1.51

(iv) Theorem 1.49 (product rule) □

1.8.4 Finding determinant by row reduction

Note that cofactor expansion is pretty inefficient for large n, as it is a recursive algorithm. We can find the determinant in a more efficient way: by row reducing it into triangular form, and using the following property:

Theorem 1.55. If A is a triangular matrix, then $\det A$ is the product of entries on the main diagonal of A.

Proof. There are two cases: upper triangular and lower triangular. Let A be an $n \times n$ triangular matrix.

Upper triangular:

Let's use induction on the matrix size n. When $n = 1$, $\det A = a_{11}$, so the statement is trivially true.

Assume that the determinant is product of main diagonal when matrix size is $n - 1$. When matrix size is n, the matrix has the form $\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{bmatrix}$. Note that $a_{i1} = 0$ for $i \geq 2$. So

$$\begin{aligned}
\det A &= \sum_{i=1}^{n} a_{i1}(-1)^{i+1} \det A_{(i1)} \\
&= a_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2n} \\ 0 & a_{33} & \dots & a_{3n} \\ 0\vdots & 0\vdots & \ddots & \vdots \\ 0 & 0 & 0\dots & a_{nn} \end{vmatrix} - 0 + \ldots + 0
\end{aligned}$$

Note that the determinant above has size $(n-1)$ and is upper triangular. By induction assumption, it evaluates to $a_{22}a_{33}\ldots a_{nn}$. Thus, $\det A = a_{11}a_{22}\ldots a_{nn}$, which is the product of entries on the main diagonal.

Lower triangular:

Let's use induction on the matrix size n. When $n = 1$, the statement is trivially true.

Assume that the determinant is product of main diagonal when matrix size is $n-1$. When matrix size is n, the matrix has the form $\begin{bmatrix} a_{11} & 0 & 0\dots & 0 \\ a_{21} & a_{22} & 0\dots & 0 \\ \vdots & \vdots & \ddots & 0\vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$, and

$$\det A = \sum_{i=1}^{n} a_{i1}(-1)^{i+1} \det A_{(i1)}$$

$$= a_{11}\begin{vmatrix} a_{22} & 0 & 0\dots & 0 \\ a_{32} & a_{33} & 0\dots & 0 \\ \vdots & \vdots & \ddots & 0\vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} - a_{12}\begin{vmatrix} 0 & 0 & 0\dots & 0 \\ a_{32} & a_{33} & 0\dots & 0 \\ \vdots & \vdots & \ddots & 0\vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} + a_{13}\begin{vmatrix} 0 & 0 & 0\dots & 0 \\ a_{22} & 0 & 0\dots & 0 \\ \vdots & \vdots & \ddots & 0\vdots \\ a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} + \text{rest of the terms}$$

Note that for terms of $i \geq 2$, each determinant has the top row being all zero since it consists of $a_{12}, a_{13}, \dots, a_{1n}$. Thus, by Theorem 1.45, they all evaluate to zero.

Apply induction assumption on the determinant next to a_{11}, and we get $\det A = a_{11}a_{22}\dots a_{nn}$.

□

Example:

Given $A = \begin{bmatrix} 1 & 2 & 3 & 2 \\ 1 & -3 & 2 & 1 \\ 2 & 1 & 2 & 5 \\ 3 & -4 & 1 & 2 \end{bmatrix}$, to evaluate $\det A$:

$$\begin{vmatrix} 1 & 2 & 3 & 2 \\ 1 & -3 & 2 & 1 \\ 2 & 1 & 2 & 5 \\ 3 & -4 & 1 & 2 \end{vmatrix} \xrightarrow[\substack{R_3-2R_1 \\ R_4-3R_1}]{R_2-R_1} \begin{vmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & -3 & -4 & 1 \\ 0 & -10 & -8 & -4 \end{vmatrix} \xrightarrow[R_4-2R_2]{\substack{\det\times 5 \\ 5R_3-3R_5}} \begin{vmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & 0 & -17 & 8 \\ 0 & 0 & -6 & -2 \end{vmatrix} \xrightarrow{R_4-\frac{6}{17}R_3} \begin{vmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & 0 & -17 & 8 \\ 0 & 0 & 0 & -\frac{82}{17} \end{vmatrix}$$

$$\det A = 1(-5)(-17)(-\frac{82}{17}) \times \frac{1}{5} = -82$$

Note: We scaled a row by 5 in the process, so we have to multiply by $\frac{1}{5}$ to get the original determinant.

1.8.5 Cramer's rule

Cramer's rule is a way of solving a linear system having unique solution using determinants.

Notation: Let $A_j(\mathbf{b})$ denote the matrix obtained from A by replacing column j by the vector $\mathbf{b}$.

Example: Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$. Then

$$A_j(\mathbf{b}) = A = \begin{bmatrix} a_{11} & b_1 & a_{13} & a_{14} \\ a_{21} & b_2 & a_{23} & a_{24} \\ a_{31} & b_3 & a_{33} & a_{34} \\ a_{41} & b_4 & a_{43} & a_{44} \end{bmatrix}$$

Theorem 1.56 (Cramer's rule)**.** Let $A\mathbf{x} = \mathbf{b}$ be a linear system where A is an invertible $n \times n$ matrix. Then the solution $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ has entries given by

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A}\ ,\ x_2 = \frac{\det A_2(\mathbf{b})}{\det A}\ ,\ \dots\ ,\ x_n = \frac{\det A_n(\mathbf{b})}{\det A}$$

where each $A_j(\mathbf{b})$ is obtained by replacing the j th column of A by the vector $\mathbf{b}$.

Proof. Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots \mathbf{a}_n \end{bmatrix}$ be an invertible matrix, where each $\mathbf{a}_i$ is a column of A. Denote the columns of the $n \times n$ identity matrix I by $\mathbf{e}_1, \dots, \mathbf{e}_n$, and denote $I_j(\mathbf{x}) = \begin{bmatrix} \mathbf{e}_1 & \dots & \mathbf{x} & \dots & \mathbf{e}_n \end{bmatrix}$. If $A\mathbf{x} = \mathbf{b}$, consider the product $A \cdot I_j(\mathbf{x})$:

$$\begin{aligned} A \cdot I_j(\mathbf{x}) &= A \begin{bmatrix} \mathbf{e}_1 & \dots & \mathbf{x} & \dots & \mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} A\mathbf{e}_1 & \dots & A\mathbf{x} & \dots & A\mathbf{e}_n \end{bmatrix} \qquad \text{(column of matrix product: Theorem 1.5)} \\ &= \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{b} & \dots & \mathbf{a}_n \end{bmatrix} \qquad \text{(since } A\mathbf{e}_j = \mathbf{a}_j \text{ by Theorem 1.7 and } A\mathbf{x} = \mathbf{b}) \\ &= A_j(\mathbf{b}) \end{aligned}$$

Taking determinants of both sides and using product rule (1.49):

$$(\det A)(\det I_j(\mathbf{x})) = \det A_j(\mathbf{b})$$

Note that $\det I_j(\mathbf{x})$ has the form

$$\begin{vmatrix} 1 & 0 & \dots & x_1 & \dots & 0 \\ 0 & 1 & \dots & x_2 & \dots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & 0 & \dots & x_j & \dots & 0 \\ \vdots & & & & \ddots & \vdots \\ 0 & 0 & \dots & x_n & \dots & 1 \end{vmatrix}$$

This can be obtained from I by performing the scaling operation $R_j \to x_j R_j$ on row j which multiplies determinant by x_j, and row replacement operation $R_p \to R_p + x_p R_j$ on all other rows p, which doesn't change the determinant. Thus, $\det I_j(\mathbf{x}) = x_j 1 = x_j$ by Theorem 1.41 (effect of row opeartion on det).

Thus, the equation above becomes

$$(\det A) x_j = \det A_j(\mathbf{b})$$

Since A is invertible, $\det A \neq 0$ by Theorem 1.48 (determinant determines invertibility). So we can divide both sides by $\det A$ to obtain

$$x_j = \frac{\det A_j(\mathbf{b})}{\det A}$$

which is the desired formula. □

Example: To solve the linear system

$$\begin{aligned} 3x - 2y + z &= 13 \\ -2x + y - 4z &= 11 \\ x + 4y - 5z &= -31 \end{aligned}$$

Let $A\mathbf{x} = \mathbf{b}$ where $A = \begin{bmatrix} 3 & -2 & 1 \\ -2 & 1 & 4 \\ 1 & 4 & -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 13 \\ 11 \\ -31 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

Calculate the required determinants:

$$\det A = \begin{vmatrix} 3 & -2 & 1 \\ -2 & 1 & 4 \\ 1 & 4 & -5 \end{vmatrix} = 3 \begin{vmatrix} 1 & 4 \\ 4 & -5 \end{vmatrix} - (-2) \begin{vmatrix} -2 & 4 \\ 1 & -5 \end{vmatrix} + \begin{vmatrix} -2 & 1 \\ 1 & 4 \end{vmatrix} = -60$$

$$\det A_x(\mathbf{b}) = \begin{vmatrix} 13 & -2 & 1 \\ 11 & 1 & 4 \\ -31 & 4 & -5 \end{vmatrix} = 13(-5 - (4)(4)) - (-2)(11(-5) - 4(-31)) + 11(4) - (-31) = -60$$

$$\det A_y(\mathbf{b}) = \begin{vmatrix} 3 & 13 & 1 \\ -2 & 11 & 4 \\ 1 & -31 & -5 \end{vmatrix} = 3(11(-5) - 4(-31)) - 13((-2)(-5) - 4) + (-2)(-31) - 11 = 180$$

$$\det A_z(\mathbf{b}) = \begin{vmatrix} 3 & -2 & 13 \\ -2 & 1 & 11 \\ 1 & 4 & -31 \end{vmatrix} = 3(-31 - 11(4)) - (-2)((-2)(-31) - 11) + 13(-2(4) - 1) = -240$$

Thus, by Cramer's rule,

$$x = \frac{\det A_x(\mathbf{b})}{\det A} = \frac{-60}{-60} = 1$$

$$y = \frac{\det A_y(\mathbf{b})}{\det A} = \frac{180}{-60} = -3$$

$$z = \frac{\det A_z(\mathbf{b})}{\det A} = \frac{-\ 240}{-60} = 4$$

The solution is $(x, y, z) = (1, -3, 4)$.

1.9 Finding inverse matrix

There are two main ways of finding the inverse of a matrix.

1.9.1 Gauss-Jordan method

Process 1.42 (Gauss-Jordan method)**.** Given a square matrix A, we can find A^{-1} (if any) as follows:

1. Place A and I side by side to form the block matrix $\begin{bmatrix} A & I \end{bmatrix}$.
2. Row reduce A into reduced echelon form R. At each step, use the same row operation on I. (So that the the same row operation is performed on the whole row of $\begin{bmatrix} A & I \end{bmatrix}$.)
3. If R contains a zero row, A^{-1} does not exist.
4. Otherwise, $R = I$, and the block matrix will become $\begin{bmatrix} I & A^{-1} \end{bmatrix}$. The right half of the block matrix is exactly A^{-1}.

Example: Let $A = \begin{bmatrix} 2 & 7 & 1 \\ 1 & 4 & 1 \\ 1 & 3 & 0 \end{bmatrix}$. To find A^{-1}, start with $\begin{bmatrix} A & I \end{bmatrix}$:

$$\left[\begin{array}{ccc|ccc} 2 & 7 & -1 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_1 \leftrightarrow R_2} \left[\begin{array}{ccc|ccc} 1 & 4 & -1 & 0 & 1 & 0 \\ 2 & 7 & 1 & 1 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow[R_3 - R_1]{R_2 - 2R_1} \left[\begin{array}{ccc|ccc} 1 & 4 & -1 & 0 & 1 & 0 \\ 0 & -1 & 3 & 1 & -2 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 \end{array}\right] \xrightarrow[\text{then } -1R_2]{R_3 - R_2}$$

$$\left[\begin{array}{ccc|ccc} 1 & 4 & -1 & 0 & 1 & 0 \\ 0 & 1 & -3 & -1 & 2 & 0 \\ 0 & 0 & -2 & -1 & 1 & 1 \end{array}\right] \xrightarrow[\text{then } -\frac{1}{2}R_3]{R_2 - \frac{3}{2}R_3} \left[\begin{array}{ccc|ccc} 1 & 4 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{3}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right] \xrightarrow{R_1 - 4R_2 + R_3} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{3}{2} & -\frac{3}{2} & \frac{11}{2} \\ 0 & 1 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{3}{2} \\ 0 & 0 & 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{array}\right]$$

Thus, $A^{-1} = \begin{bmatrix} -\frac{3}{2} & -\frac{3}{2} & \frac{11}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{3}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$.

The method relies on the following fact.

Theorem 1.57. If a sequence of elementary row operations transforms A into I, then the same sequence of operations transforms I into A^{-1}.

Proof. Suppose a sequence of elementary row operations transforms A into I, and the same sequence transforms I into C. Then we can let $E_k \dots E_2 E_1 A = I$ where E_i are elementary matrices, and $E_k \dots E_2 E_1 I = C$. This means $(E_k \dots E_2 E_1)A = CA = I$. By Corollary 1.36 (left inverse implies inverse), C is the inverse of A. Thus $C = A^{-1}$. □

1.9.2 Inverse matrix formula

Definition 1.43. For a square matrix A, the **cofactor matrix**, denoted C_A or $\operatorname{cof} A$, is the matrix containing the cofactor of each entry of A.

$$C_A = [C_{ij}]$$

where $C_{ij} = (-1)^{i+j} \det A_{(ij)}$ is the (i, j)-cofactor of A.

The **adjugate** of A, denoted $\operatorname{adj} A$, is the transpose of the cofactor matrix of A.

$$\operatorname{adj} A = C_A^T$$

Theorem 1.58 (Inverse matrix formula / Adjugate formula)**.** Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A$$

Writing the entries out:

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$$

or more explicitly,

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} +\det A_{(11)} & -\det A_{(12)} & \dots & (-1)^{n+1} \det A_{(1n)} \\ -\det A_{(21)} & +\det A_{(22)} & \dots & (-1)^{n} \det A_{(2n)} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1} \det A_{(n1)} & (-1)^{n} \det A_{(n2)} & \dots & +\det A_{(nn)} \end{bmatrix}^T$$

Proof. For an invertible $n \times n$ matrix $A = [a_{ij}]$, let $A^{-1} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix}$ where each $\mathbf{x}_j$ is the j th column of A^{-1}. Let $\mathbf{e}_j$ be the j th column of identity matrix I. We have

$$\begin{aligned} AA^{-1} &= I \\ A \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix} \\ \begin{bmatrix} A\mathbf{x}_1 & A\mathbf{x}_2 & \dots & A\mathbf{x}_n \end{bmatrix} &= \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \end{bmatrix} \end{aligned}$$

This means

$$A\mathbf{x}_j = \mathbf{e}_j$$

for each column j. Note that the i th entry of $\mathbf{x}_j$ (denoted $(\mathbf{x}_j)_i$) is the (i, j)-entry of A^{-1}. By Cramer's rule (1.56),

$$(A^{-1})_{ij} = (\mathbf{x}_j)_i = \frac{\det A_i(\mathbf{e}_j)}{\det A} \qquad \dots(1)$$

where $A_i(\mathbf{e}_j)$ is the matrix obtained by replacing the i th column of A by the j th column of I. Note that $A_i(\mathbf{e}_j)$ has the form

$$A_i(\mathbf{e}_j) = \begin{bmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ a_{12} & \dots & 0 & \dots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{j1} & \dots & 1 & \dots & a_{jn} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & \underbrace{0}_{i\text{th column}} & \dots & a_{nn} \end{bmatrix}$$

By Cofactor Expansion Theorem (1.53), cofactor expansion along any row or column leads to the same determinant. A cofactor expansion down the i th column of $A_i(\mathbf{e}_j)$ shows that

$$\begin{aligned} \det A_i(\mathbf{e}_j) &= \sum_{k=1}^{n} (A_i(\mathbf{e}_j))_{ki} (-1)^{k+i} \det A_{(ki)} \\ &= 0 + (-1)^{j+i} \det A_{(ji)} + 0 \\ \det A_i(\mathbf{e}_j) &= C_{ji} \qquad \dots(2) \end{aligned}$$

where C_{ji} is the (j, i)-cofactor of A. By equation (1) and (2), the (i, j)-entry of A^{-1} is

$$\begin{aligned} (A^{-1})_{ij} &= \frac{C_{ji}}{\det A} \\ &= \frac{1}{\det A} (C^T)_{ij} \qquad \text{(Taking transpose means swapping indices)} \end{aligned}$$

Thus, writing A^{-1} out,

$$\begin{aligned} A^{-1} &= \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix} \\ &= \frac{1}{\det A}\operatorname{adj} A \end{aligned}$$

$\square$

Theorem 1.59. (i) If $A = [a_{11}]$ is a 1×1 non-zero matrix, then $A^{-1} = [\frac{1}{a_{11}}]$.

(ii) If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is an invertible 2×2 matrix, then

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(iii) If $A = [a_{ij}]$ is an invertible 3×3 matrix, then

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} +\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ -\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \\ +\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} & -\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} & +\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix}^T$$

Proof. We're gonna use the inverse matrix formula (1.58) introduced in the previous theorem.

(i) The cofactor C_{11} of a 1×1 matrix is defined to be 1. So $A^{-1} = \dfrac{1}{\det A}[C_{11}] = \dfrac{1}{a_{11}}[1] = [\dfrac{1}{a_{11}}]$.

(ii) If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}^T = \frac{1}{ad - bc}\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}^T = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

(iii) Just use the inverse matrix formula for $n = 3$.

$\square$

Example: To find the inverse of $A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix}$,

$\det A = 2\begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} + 3\begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} = 14$

$$\begin{aligned} A^{-1} &= \frac{1}{\det A}\begin{bmatrix} +\begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} & +\begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} \\ -\begin{vmatrix} 1 & 3 \\ 4 & -2 \end{vmatrix} & +\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} & -\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} \\ +\begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} & -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} & +\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} \end{bmatrix}^T \\ &= \frac{1}{14}\begin{bmatrix} -2 & 3 & 5 \\ 14 & -7 & -7 \\ 4 & 1 & -3 \end{bmatrix}^T = \frac{1}{14}\begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} \end{aligned}$$

$$A^{-1} = \begin{bmatrix} -\frac{1}{7} & 1 & \frac{2}{7} \\ \frac{3}{14} & -\frac{1}{2} & \frac{1}{14} \\ \frac{5}{14} & -\frac{1}{2} & -\frac{3}{14} \end{bmatrix}$$

1.10 Solving linear systems (summary)

Theorem 1.60. A linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if $\det A \neq 0$.

If $\det A = 0$, then the system either has no solutions or infinitely many solutions.

Proof. By Theorem 1.48 (determinant determines invertibility), A is invertible if and only if $\det A \neq 0$. By Corollary 1.37, the linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only A is invertible. Thus, $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if $\det A \neq 0$. □

To summarize, there are several ways to solve a linear system $A\mathbf{x} = \mathbf{b}$.

(i). Gaussian elimination / Gauss-Jordan elimination

1. Row reduce the augmented matrix into row echelon form / reduced echelon form.
2. If a row $\left[\begin{array}{cccc|c} 0 & 0 & \dots & 0 & b \end{array}\right]$ appears where $b \neq 0$, the system is inconsistent.
3. Otherwise, the system is consistent.
   (a) If number of non-zero rows r = number of variables n, then the system has a unique solution. Use back substitution / direct substitution to determine the value of each variable.
   (b) If number of non-zero rows r < number of variables n, then the system has infinitely many solution. Assign free variables to the non-leading variables and use back substitution to express leading variables in terms of free variables.

(ii). Determinant and Cramer's rule

1. Find the determinant of the coefficient matrix, i.e. $\det A$.
2. If $\det A \neq 0$, then the system has a unique solution. Use Cramer's rule to find the value of each variable.
3. Otherwise, if $\det A = 0$, the system either is inconsistent or has infinitely many solutions. Use Gaussian elimination in this case.

(iii). Inverse matrix

1. Find A^{-1} by Gauss-Jordan method or Inverse matrix formula.
2. Find the solution vector $\mathbf{x} = A^{-1}\mathbf{b}$.

Method (iii) is usually the most troublesome, but it is useful when A^{-1} has already been found.

2 Vector spaces

2.1 Definition

Definition 2.1 (Addition and scalar multiplication on a set)**.** .

- An *addition* on a set V is a function that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$.
- A *scalar multiplication* on a set V is a function that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbb{F}$ and each $\lambda \in V$.

Definition 2.2 (Vector space)**.** A **vector space** is a non-empty set V along with an addition on V and a scalar multiplication on V such that the following properties hold.

(i). **commutativity**

$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v} \in V$.

(ii). **associativity**

$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$ and $(cd)\mathbf{v} = c(d\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and for all $c, d \in \mathbb{F}$.

(iii). **additive identity**

There exists an element $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.

(iv). **additive inverse**

For every $\mathbf{v} \in V$, there exists $w \in V$ such that $\mathbf{v} + \mathbf{w} = \mathbf{0}$.

(v). **multiplicative identity**

$1\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$

(vi). **distributive property**

$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ and $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$ for all $c, d \in \mathbb{F}$ and all $\mathbf{u}, \mathbf{v} \in V$.

From the way addition and scalar multiplication on V are defined, the vector space must be closed under addition and scalar multiplication, meaning if $u, v \in V$, then $u + v \in V$; and if $v \in V$, then $\lambda v \in V$.

Elements of a vector space are called vectors (as expected).
A vector space over $\mathbb{R}$ is called a **real vector space**, and a vector space over $\mathbb{C}$ is called a **complex vector space**.

Theorem 2.1. (i). A vector space has a unique additive identity.

(ii). Every element in a vector space has a unique additive inverse.

Proof. (i) Suppose $\mathbf{0}$ and $\mathbf{0}'$ are both additive identities foro some vector space V. Then

$$\mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}$$

where the first equality holds because $\mathbf{0}$ is an additive identity, the second equality comes from commutativity, and the third equality holds because $\mathbf{0}'$ is an additive identity. Thus $\mathbf{0}' = \mathbf{0}$, proving that V has only one additive identity.

(ii) Suppose $\mathbf{w}$ and $\mathbf{w}'$ are additive inverses of $\mathbf{v} \in V$. Then

$$\mathbf{w} = \mathbf{w} + \mathbf{0} = \mathbf{w} + (\mathbf{v} + \mathbf{w}') = (\mathbf{w} + \mathbf{v}) + \mathbf{w}' = 0 + \mathbf{w}' = \mathbf{w}'$$

Thus $\mathbf{w} = \mathbf{w}'$, as desired. $\square$

Since additive inverse is unique, we can unambiguously use $-\mathbf{v}$ to denote $\mathbf{u}$, and use $\mathbf{u} - \mathbf{v}$ to denote $\mathbf{u} + (-\mathbf{v})$.
We also use V to denote vector space without explicitly stating V is a vector space.

Theorem 2.2. .

(i). $0\mathbf{v} = \mathbf{0}$ for every $\mathbf{v} \in V$.

(ii). $\lambda\mathbf{0} = \mathbf{0}$ for every $\lambda \in \mathbb{F}$.

(iii). $(-1)\mathbf{v} = -\mathbf{v}$ for every $\mathbf{v} \in V$.

(iii) says that if an element of V is multiplied by the scalar -1, then the result is the additive inverse of the element of V.

Proof. (i) For $\mathbf{v} \in V$, we have

$$\begin{aligned} 0\mathbf{v} &= (0+0)\mathbf{v} \\ 0\mathbf{v} &= 0\mathbf{v} + 0\mathbf{v} \qquad \text{(distributive property)} \end{aligned}$$

Adding the additive inverse of $0\mathbf{v}$ (namely $-(0\mathbf{v})$) to both sides of the equation, we get

$$\begin{aligned} 0\mathbf{v} - (0\mathbf{v}) &= 0\mathbf{v} + 0\mathbf{v} - (0\mathbf{v}) \\ \mathbf{0} &= 0\mathbf{v} \end{aligned}$$

as desired.

(ii) For $\lambda \in \mathbb{F}$, we have

$$\lambda\mathbf{0} = \lambda(\mathbf{0}+\mathbf{0}) = \lambda\mathbf{0} + \lambda\mathbf{0}$$

Adding the additive inverse of $\lambda\mathbf{0}$ to both sides gives $\mathbf{0} = \lambda\mathbf{0}$, as desired.

(iii) For $\mathbf{v} \in V$, we have

$$\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = (1+(-1))\mathbf{v} = 0\mathbf{v} = \mathbf{0}$$

This equation says that $(-1)\mathbf{v}$, when added to $\mathbf{v}$, gives $\mathbf{0}$. Thus $(-1)\mathbf{v}$ is the additive inverse of $\mathbf{v}$, as desired. ☐

References

[1] Sheldon Axler, “Linear algebra done right.” [Online]. Available: https://link.springer.com/content/pdf/10.1007/978-3-031-41026-0.pdf

[2] David C. Lay, “Linear algebra and its applications.” [Online]. Available: https://home.cs.colorado.edu/~alko5368/lecturesCSCI2820/mathbook.pdf

[3] W. Keith Nicholson, “Linear algebra with applications.” [Online]. Available: https://math.emory.edu/~lchen41/teaching/2020_Fall/Nicholson-OpenLAWA-2019A.pdf

[4] Proof Wiki, “Matrix multiplication is associative.” [Online]. Available: https://proofwiki.org/wiki/Matrix_Multiplication_is_Associative

[5] W. Keith Nicholson, “Linear algebra with applications ch.2.5.” [Online]. Available: https://math.emory.edu/~lchen41/teaching/2020_Fall/Section_2-5.pdf

[6] Matthew Towers, “Math0005 algebra 1: 3.10 rref existence and uniqueness.” [Online]. Available: https://www.ucl.ac.uk/~ucahmto/0005_2021/Ch3.S10.html

[7] W.H. Holzmann, “Uniqueness of reduced row echelon form.” [Online]. Available: https://www.cs.uleth.ca/~holzmann/notes/reduceduniq.pdf

[8] W. Keith Nicholson, “Linear algebra with applications ch.2.4.” [Online]. Available: https://math.emory.edu/~lchen41/teaching/2020_Fall/Section_2-4.pdf

[9] Emily, “Answer to: How do we know that if ax=b has a unique solution, a is invertible?” [Online]. Available: https://math.stackexchange.com/questions/791750/how-do-we-know-that-if-ax-b-has-a-unique-solution-a-is-invertible

[10] W. Keith Nicholson, “Linear algebra with applications ch.3.6.” [Online]. Available: https://math.emory.edu/~lchen41/teaching/2020_Fall/Section_3-6.pdf

[11] Ken Kuttler, “3.2: Properties of determinants.” [Online]. Available: https://math.libretexts.org/Bookshelves/Linear_Algebra/A_First_Course_in_Linear_Algebra_(Kuttler)/03%3A_Determinants/3.02%3A_Properties_of_Determinants