# Introduction to Linear Algebra

Chang, Yung-Hsuan

June 18, 2024

# Contents

# Preface

This note is summarized by Yung-Hsuan Chang as he took the course MIT 18.16 Linear Algebra instructed by Gilbert Strang.

# 1 System of Linear Equations

The fundamental problem of linear algebra is to solve $n$ linear equations in $n$ unknowns. For example, the case when $n = 2$,

$$\begin{cases} \quad 2x - y = 0; \\ -x + 2y = 3. \end{cases} \tag{1.1}$$

There are three main ways to see this problem:

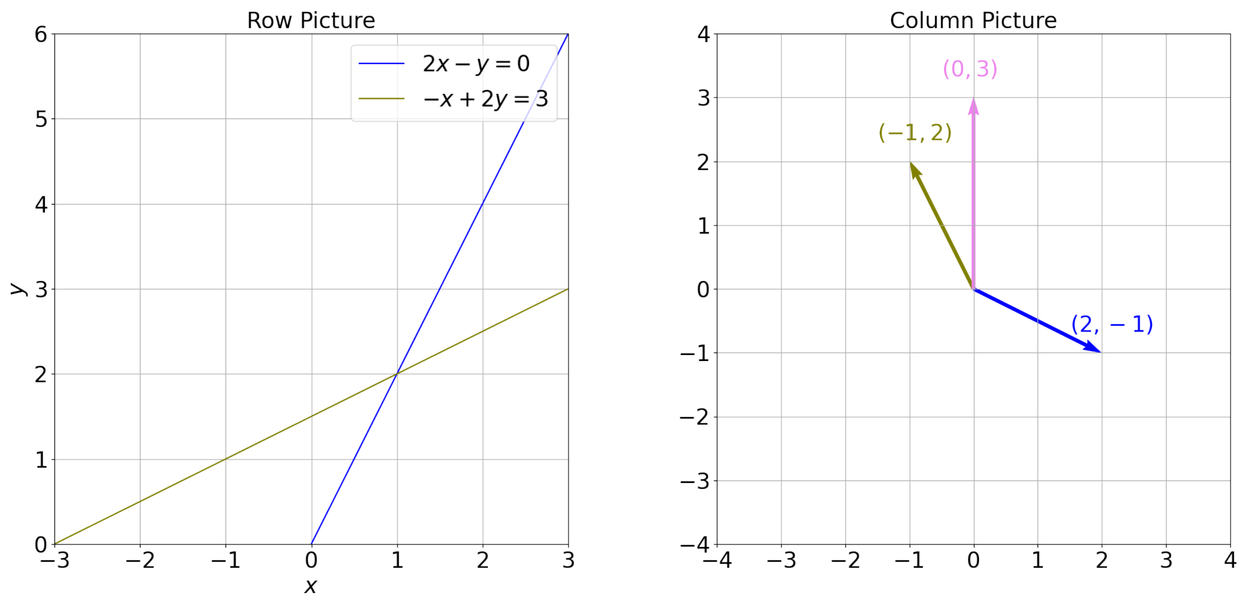
1. row picture,
2. column picture, and
3. matrix picture.

Row picture describes the relationship among equations. Take (1.1) for an example, the graph shows the concept of row picture.

Column picture sees unknowns as scalars of vectors. In (1.1), the system can be written as

$$x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}. \tag{1.2}$$

Column picture can be illistrated as below.



One can imagine that, after some stretch (being multiplied by a scalar), the sum of the two vectors, $(2,-1)$ with scalar $x$ and $(-1,2)$ with scalar $y$, is $(0,3)$. The true answer for (1.1) is $(x,y)=(1,2)$. One can easily verify and check the solution. Note that the high and thin notation representing a vector in (1.2) and the horizontal and with comma notation represent the same thing, i.e., both

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \text{and} \qquad (2,-1)$$

represent the vector with first component 2 and the second component $-1$. They are both called the "column vector." The high and thin notation coincides with the matrix picture, which will be discussed.

In matrix picture, we write the system as

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}. \tag{1.3}$$

In this case, we call $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$ the coefficient matrix and $\begin{bmatrix} x \\ y \end{bmatrix}$ the vector of unknowns. We can simply write

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

where $A$ is the coefficient matrix and $\mathbf{x}$ is the vector of unknowns. The benefit of this form is that there might be some beautiful properties for the matrix $A$. We are going to discuss those properties in this book.

## 1.1 Elimination with Matricies

**Question**. How to solve the equation

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

in a systematic way?

We can solve the equation by transforming the matrix $A$ into an upper triangular matrix.

**Notation** (Matrix). We usuallly use a capital latter to represent a matrix. For example, just as we see, $A$. In addition, we might use subscript so indicate the numbers of rows $m$ and the number of columns $n$ by writing $A_{m\times n}$. Moreover, we use subscript with two numbers $i$ and $j$ to indicate the component $A_{ij}$ on the $i$-th row and the $j$-th column of the matrix $A$ if there is no confusion.

**Example 1.1.1**. Let

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

Then, $A_{11} = 4$, $A_{21} = -1$, $A_{12} = 0$, and $A_{22} = 2$.

**Definition 1.1.2** (Upper Triangular). We say a square matrix $A_{n\times n}$ is upper triangular if $A_{ij} = 0$ for all $n \geq i > j > 0$, i.e.,

$$A = \begin{bmatrix} & & \\ 0 & & * \\ \vdots & \ddots & \\ 0 & \cdots & 0 \end{bmatrix},$$

where the asterisk denotes any possible situation.

**Example 1.1.3**. Let

$$B = \begin{bmatrix} 4 & -1 \\ 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 6 & 0 \\ 7 & -13 \end{bmatrix}.$$

Then, $B$ is upper triangular, and $C$ is not upper triangular since $A_{21} \neq 0$.

**Row Operation and Elementary Matrix**

If we have an upper triangular coefficient matrix, the solution for the last equation is straightforward. We can then solve the equation above it, followed by the one above that, and so on. To transform a matrix $A$ into an upper triangular matrix $U$, we simply need to multiply $A$ by an appropriate sequence of elementary matrices.

**Definition 1.1.4** (Elementary Matrix). The effect of elementary matrix is to do row operations. There are three types of elementary matrix:

1. row switching,
2. row multiplication, and
3. row addition.

Row switching exchanges two rows, row multiplication makes a specific row being scaled by a non-zero constant, and row addition replace a row with the sum of it and another row with a scalar. Symbolically, we write

$$R_i \leftrightarrow R_j$$

to indicate row switching between row $i$ and row $j$,

$$kR_i \to R_i$$

to indicate row $i$ is scaled by $k$, and

$$R_i + kR_j \to R_i$$

to indicate row $i$ is being added by $R_j$ scaled by $k$.

Take a matrix

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix}$$

for example, we add $-3$ times of the first row to the second row, which makes the matrix $A$ become

$$E_{21}A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix},$$

where the matrix $E_{21}$ works on the first row and the second row, without knowing what $E_{21}$ is now. To make the matrix upper triangular, we add $-2$ times of the second row the the third row, which makes the matrix $A$ become an upper triangular matrix

$$U = E_{32}E_{21}A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix},$$

where the matrix $E_{32}$ works on the decond row and the third row, without knowing what $E_{32}$ is now as well. Sometimes, we need an $E_{31}$; however, since the first element of the third row is 0, we do not need an $E_{31}$.

What is $E_{21}$ and $E_{32}$? We have better to look into some properties of matrices first. We will answer this question after some inquires.

If we have a matrix $A_{m\times n}$ and a column vector $\mathbf{v}_{n\times 1}$, what does $A\mathbf{v}$ mean? We can of course apply the matrix multiplication on it, the answer should be obvious.

**Definition 1.1.5** (Matrix Multiplication). Let $A$ be an $m \times n$ matrix and let $B$ be an $n \times p$ matrix. Let $C = AB$. Then,

$$C_{ij} = \sum_{\ell=1}^{n} A_{i\ell}B_{\ell j}$$

for $i = 1, 2, \ldots, m$ and for $j = 1, 2, \ldots, p$.

However, we can actually observe that

$$A\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{a}_i,$$

where $v_i$ is the $i$-th component of $\mathbf{v}$ and $\mathbf{a}_i$ is the $i$-th column of $A$. Through either way can we find out that the product $A\mathbf{v}$ is a column vector and is also an $n \times 1$ matrix.

How about multiplying a row vector $\mathbf{r}_{1\times m}$ at the left side of $A_{m\times n}$? We can still have another perspective to see this operation aside from applying the multiplication rule directly, which is

$$\mathbf{r}A = \sum_{i=1}^{m} r_i \boldsymbol{\alpha}_i,$$

where $r_i$ is the $i$-th component of $\mathbf{r}$ and $\boldsymbol{\alpha}_i$ is the $i$-th row of $A$. Therefore, $\mathbf{r}A$ will be a row vector and also a $1 \times m$ matrix.

Let's get back to the question mentioned just now, what is $E_{21}$ and $E_{32}$? We have

$$E_{21}\begin{bmatrix}1 & 2 & 1\\ 3 & 8 & 1\\ 0 & 4 & 1\end{bmatrix}=\begin{bmatrix}1 & 2 & 1\\ 0 & 2 & -2\\ 0 & 4 & 1\end{bmatrix},$$

and $E_{21}$ is the operation that adds $-3$ times of the first row to the second row on $A$. We first write the system of equation with the concept

$$\mathbf{r}A=\sum_{i=1}^{m} r_i\boldsymbol{\alpha}_i,$$

having

$$\begin{cases}\boldsymbol{\epsilon}_1 A=[1 \quad 2 \quad 1];\\ \boldsymbol{\epsilon}_2 A=[0 \quad 2 \quad -2];\\ \boldsymbol{\epsilon}_3 A=[0 \quad 4 \quad 1],\end{cases}$$

where $\boldsymbol{\epsilon}_i$ is the $i$-th row of $E_{21}$. It is clear that $\boldsymbol{\epsilon}_1=[1 \quad 0 \quad 0]$ and $\boldsymbol{\epsilon}_3=[0 \quad 0 \quad 1]$ since the first row and the third row of $E_{21}A$ are identical to those of $A$. The hard part is $\boldsymbol{\epsilon}_2$. Since we add $-3$ times of the first row to the second row, the second row of $E_{21}A$ is

$$[3 \quad 8 \quad 1]+(-3)[1 \quad 2 \quad 1].$$

Hence,

$$\boldsymbol{\epsilon}_2 A=(-3)[1 \quad 2 \quad 1]+(1)[3 \quad 8 \quad 1]+(0)[0 \quad 4 \quad 1],$$

which makes $\boldsymbol{\epsilon}_2=[-3 \quad 1 \quad 0]$. We therefore obtain our matrix

$$E_{21}=\begin{bmatrix}1 & 0 & 0\\ -3 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

How about $E_{32}$? The matrix $E_{32}$ is the operation that adds $-2$ times of the second row to the third row on $E_{21}A$. By the same fashion, we obtain

$$E_{32}=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -2 & 1\end{bmatrix},$$

which makes

$$E_{32}E_{21}A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix}.$$

Aside from the elementary matrix, we can have more perspective about the multiplication. From the concept of

$$A\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{a}_i,$$

we can actually extend this kind of multiplication between two matrices. We just consider

$$V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_p]_{n\times p}$$

and

$$C = AV = [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_p]_{m\times p},$$

and we can find out that each column of $C$ is a combination of columns of $A$. This kind of multiplication is thought as columns time columns.

From the concept of

$$\mathbf{r}A = \sum_{i=1}^{m} r_i \boldsymbol{\alpha}_i,$$

we can also extend this kind of multiplication. We consider

$$R = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_k \end{bmatrix}_{k\times m}$$

and

$$D = RA = \begin{bmatrix} \mathbf{r}_1 A \\ \mathbf{r}_2 A \\ \vdots \\ \mathbf{r}_k A \end{bmatrix}_{k\times n},$$

we can fund out that each row of $D$ is a combination of rows of $A$. This kind of multiplication is thought as rows time rows.

We have one more critical skill about matrix multiplication, the block multiplication. This is for big matrices. Say, we have $P_{20\times 20}Q_{20\times 20} = R_{20\times 20}$. We can separate these big matrices, for example, as

$$P_{20\times 20} = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix},$$

where $P_1, P_2, P_3, P_4$ are all $10 \times 10$. In the same fashion, we have

$$P_{20\times 20} = \begin{bmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{bmatrix}$$

and

$$R_{20\times 20} = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix}.$$

Since $P_{20\times 20}Q_{20\times 20} = R_{20\times 20}$, we will have $R_1 = P_1Q_1 + P_2Q_3$. Nobody can see this instantly work, but it works.

We have some rules about matrix multiplications.

**Theorem 1.1.6** (Matrix Multiplication). Let $A_{m\times n}, B_{n\times p}, C_{n\times p}, D_{k\times m}$ be matrices. Then,

1. $A(B + C) = AB + AC$,
2. $(B + C)D = BD + CD$, and
3. $(AB)C = A(BC)$.

**Inverse Matrix**

For the equation $A_{m\times n}\mathbf{x}_{n\times 1} = \mathbf{b}_{m\times 1}$, if we can find some matrix $M$ such that $MA\mathbf{x} = \mathbf{x}$, then we just obtain the solution $\mathbf{x} = M\mathbf{b}$. How simple linear algebra is! In order to obtain the solution, we might ask for a stronger condition. If we have $MA\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, then $MA$ is an $n \times n$ identity matrix, where we write $MA = I_n$. In this case, we call $M$ a left inverse of $A$. For the opposite case, $AN = I_m$, we call $N$ a right inverse of $A$.

**Definition 1.1.7** (Invertible and Inverse Matirx). Let $A$ be an $n \times n$ matrix. We say $A$ is invertible if there exists a unique matrix $B$ such that $AB = BA = I_n$. If such matrix $B$ exists, $A^{-1}$ denotes $B$ and is called the inverse (matrix) of $A$.

**Theorem 1.1.8** (Inverse Matrix for Square Matrix). Let $A$ be an $n \times n$ matrix. If $AB = I_n$, then $BA = I_n$, and vise versa.

Now, the question is that how we can find the inverse and in what condition we can find the inverse?

**Question**. What conditions do we have that suggest a matrix is invertible? Are they sufficient or necessary?

**Theorem 1.1.9**. Let $A$ be an matrix. The equation $A\mathbf{x} = \mathbf{0}$ has a solution other than $\mathbf{x} = \mathbf{0}$ if and only if $A$ is not ivertible.

We now obtain a sufficient and necessary condition about whether a matrix is invertible or not. We can now find the ways to obtain inverses.

**Theorem 1.1.10**. Let $A$ be invertible. We consider the augmented matrix $[A \quad I]$ and use the Gauss-Jordan method on it until we have $[I \quad B]$. Then, $B$ is the inverse of $A$.

**Proof**. Assume that $A$ can be decomposed into the identity matrix $I$. Then, we will have

$$E_k E_{k-1} \cdots E_2 E_1 A = I$$

for $E_i$ elementary matrices. Note that elementary matrices are all invertible. This means the product of $E_i$'s are the inverse of $A$ by definition. Hence, we just need to do the same operation on $I$ to obtain $B$, which is now known as the inverse of $A$. That is,

$$\begin{aligned} E_k E_{k-1} \cdots E_2 E_1 I &= B \\ &= A^{-1}. \end{aligned}$$

$\square$

In addition, since we are doing row operations, so we can know that inverses are just multiplication of elementary matrices, since we can either

1. switch the rows again,
2. multiplicate a row by the inverse of the scalar again, or
3. add $(-1)$ of row $i$ to row $j$ again

to cancel out the row operation we just done. The question is that how can we find the inverse if we have a matrix made up with a bunch of elementary matrices? We have a nice property about the inverse of the product of matrices.

**Theorem 1.1.11** (Inverse of Product). Let $A_{n\times n}, B_{n\times n}$ be invertible matrices. Then,

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

To be more general, let $\{E_i\}_{i=1}^m$ be a finite sequence of invertible $n \times n$ matrices. Then,

$$(E_1E_2\cdots E_{m-1}E_m)^{-1} = {E_m}^{-1}{E_{m-1}}^{-1}\cdots {E_2}^{-1}{E_1}^{-1}.$$

**LU Decomposition**

Combining the concept of row operation and inverse matrix, it occurs to us that after transforming a matrix $A$ into an upper triangular matrix $U$, what kind of relation does $A$ and $U$ have? The answer is quite simple. We have the factorization

$$A = PLU, \tag{1.4}$$

where $P$ is the so-called permutation matrix. Permutations are used for row exchanges. For example, there are six $3 \times 3$ permutation matrices,

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}, \begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}, \begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}, \begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}, \begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}.$$

The first one changes nothing (no row exchange), which is the identity matrix. The second one exchanges the first and the second row. The third exchanges the first and the third row. The fourth exchanges the second and the third row. The fifth exchanges the second and the third row first, and then exchanges the newer first and the second row. We can see that

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

The sixth exchanges the first and the third row first, and then exchanges the newer first and second row. We can see that

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

In fact, these matrices are just different arrangements of rows $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$. It can be known that there will be a number $n!$ of $n \times n$ permutation matrices since we can put the first row in $n$ ways (you have $n$ choices in the beginning), put the second row in $(n-1)$ ways (you already chose the position of the first row), continuing and put the last row in only one way. An interesting property of permutation matrices is that thier transposes are their inverses.

**Definition 1.1.12** (Transpose). Let $A_{m\times n}$ be a matrix. The transpose of $A$, denoted by $A^{\mathrm{T}}$ is defined by ${A^{\mathrm{T}}}_{ij} = A_{ji}$. Thus, $A^{\mathrm{T}}$ will be an $n \times m$ matrix.

**Example 1.1.13**. Let

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

Then,

$$A^{\mathrm{T}} = \begin{bmatrix} 1 & -1 \\ 4 & 0 \end{bmatrix}.$$

We have a theorem about the inverse of the product of matrices, we of course have another theorem about the transpose of the product of matrices, and they are quite similar!

**Theorem 1.1.14** (Transpose of Product). Let $A_{m\times n}, B_{n\times p}$ be matrices. Then, $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$. That is, the transpose of the product is the product of transposes in reversed order.

**Proof**. Let $C = AB$. On the one hand, we have

$$\begin{aligned} {C^T}_{ij} &= C_{ji} \\ &= \sum_{k=1}^{n} A_{jk}B_{ki}. \end{aligned}$$

On the other hand, we have

$$\begin{aligned} (B^{\mathrm{T}}A^{\mathrm{T}})_{ij} &= \sum_{k=1}^{n} (B^{\mathrm{T}})_{ik}(A^{\mathrm{T}})_{kj} \\ &= \sum_{k=1}^{n} B_{ki}A_{jk} \\ &= \sum_{k=1}^{n} A_{jk}B_{ki}. \end{aligned}$$

Therefore, $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$ since ${C^T}_{ij} = (B^{\mathrm{T}}A^{\mathrm{T}})_{ij}$ for all $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, m$. $\square$

Getting back to our discussion about the inverse of permutations, we have the following result.

**Theorem 1.1.15** (Inverse of Permutations). Let $P$ be a permutation matrix. Then, $P^{-1} = P^{\mathrm{T}}$. That is, the inverse of a permutation is the transpose of a permutation.

**Corollary 1.1.16** (Inverse of Permutations). The inverse of a permutation matrix is still a permutation matrix.

Further combining inverse and transpose, we have the following property.

**Theorem 1.1.17** (Transpose of Inverse, Inverse of Transpose). Let $A_{n\times n}$ be an invertible matrix. Then, $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$. That is, the inverse of the transpose is the transpose of the inverse.

Take a small example for (1.4). For simplicity, we first set $P = I$. We have

$$A = \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix}.$$

After row operation, our destination is

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

We add $(-4)$ times of row 1 to row 2, obtaining

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

So, in the language of (1.4), we have

$$\begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix},$$

since the inverse matrix is just doing a backward operation, and the backward operation for adding $(-4)$ times of row 1 to row 2 is adding 4 times of row 1 to row 2.

## 1.2 Four Fundamental Subspaces