

Introduction to Linear Algebra

Chang, Yung-Hsuan

June 13, 2024

Contents

Contents	i
Preface	ii
1 System of Linear Equations	1
1.1 Elimination with Matrices	2
1.2 Four Fundamental Subspaces	9

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 $n = 2$,

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

There are three main ways to see this problem:

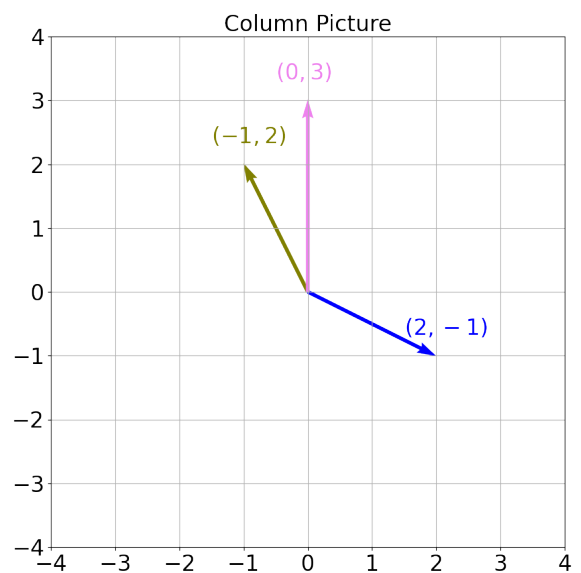
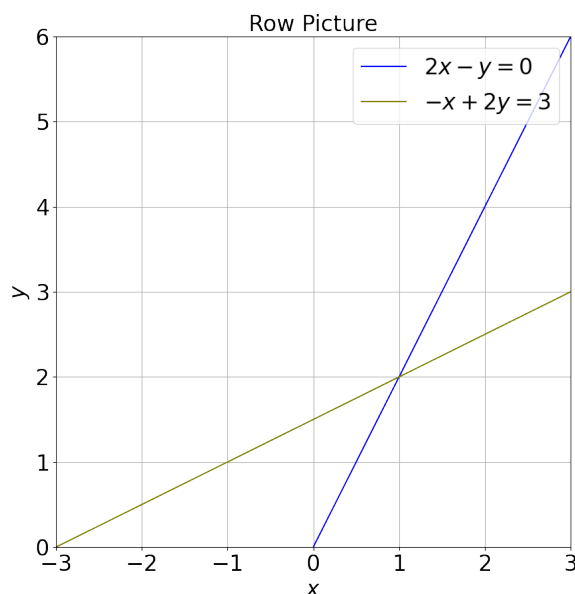
1. row picture,
2. column picture, and
3. matrix picture.

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

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

Column picture can be illustrated 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

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

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

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

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 Matrices

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 usually use a capital letter to represent a matrix. For example, just as we see, A . In addition, we might use subscript to indicate the numbers of rows m and the number of columns n by writing $A_{m \times n}$. Moreover, we use the lowercase of the letter we just chose and with two numbers i and j to indicate the component a_{ij} on the i -th row and the j -th column.

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 $c_{21} \neq 0$.

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 ,

$$kR_i \rightarrow R_i$$

to indicate row i is scaled by k , and

$$R_i + kR_j \rightarrow R_i$$

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 second 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 inquiries.

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

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

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

where $\boldsymbol{\epsilon}_i$ is the i -th row of E_{21} . It is clear that $\boldsymbol{\epsilon}_1 = [1 \ 0 \ 0]$ and $\boldsymbol{\epsilon}_3 = [0 \ 0 \ 1]$ 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

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

Hence,

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

which makes $\boldsymbol{\epsilon}_2 = [-3 \ 1 \ 0]$. 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 find 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 10×10 . In the same fashion, we have

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

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 $R_1 = P_1 Q_1 + P_2 Q_3$. Nobody can see this instantly work, but it works.

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 1.1.6 (Invertible and Inverse Matrix). 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 1.1.7 (Inverse Matrix for Square Matrix). Let A be an $n \times n$ matrix. If $AB = I_n$, then $BA = I_n$, and vice 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 1.1.8. 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 invertible.

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 1.1.9. 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 .

A quick explanation is that we see through the perspective of columns time columns. We are looking for the solution to equations

$$A\mathbf{b}_i = \mathbf{e}_i,$$

for $i = 1, 2, \dots, n$, where \mathbf{b}_i is the i -th column of the inverse of A and \mathbf{e}_i is the i -th column of I_n . Since these equations have the same coefficient matrix, the row operation on them are the same. Hence, we can just do once Gauss-Jordan method on the whole augmented matrix.

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

1.2 Four Fundamental Subspaces