Lecture Notes

for

MA 1140

Elementary Linear Algebra

by

Dipankar Ghosh

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Department of Mathematics
Indian Institute of Technology Hyderabad
Sangareddy, Kandi, Telangana - 502285, India

7th February to 11th March, 2019

Contents

1 Matrices, Linear equations and solvability **7**

1.1 Introduction to system of linear equations 7

1.2 Gaussian Elimination and some examples 9

1.3 System of linear equations (in general) 12

1.4 Row reduced echelon matrices 13

1.5 Elementary matrices 17

2 Vector Spaces **20**

2.1 Vector Space 20

2.2 Subspaces 22

2.3 Basis and Dimension 24

3 Linear Transformations **31**

3.1 Some well known (linear) maps 31

Bibliography **33**

Introduction

Linear algebra[1] is the branch of mathematics concerning linear equations such as

$$a_1x_1 + \cdots + a_nx_n = b$$

linear functions such as

$$(x_1, \ldots, x_n) \mapsto a_1x_1 + \ldots + a_nx_n$$

and their representations through matrices and vector spaces.

Linear algebra is central to almost all areas of mathematics. For instance, linear algebra is fundamental in modern presentations of geometry, including for describing basic objects such as lines, planes and rotations. Also, functional analysis may be basically viewed as the application of linear algebra to spaces of functions. Linear algebra is also used in most sciences and engineering areas, because it allows modeling many natural phenomena, and efficiently computing with such models. For nonlinear systems, which cannot be modeled with linear algebra, linear algebra is often used as a first-order approximation.

Until the 19th century, linear algebra was introduced through systems of linear equations and matrices. In modern mathematics, the presentation through vector spaces is generally preferred, since it is more synthetic, more general (not limited to the finite-dimensional case), and conceptually simpler, although more abstract.

[1] The introduction is noted from Wikipedia [2].

Chapter 1

Matrices, Linear equations and solvability

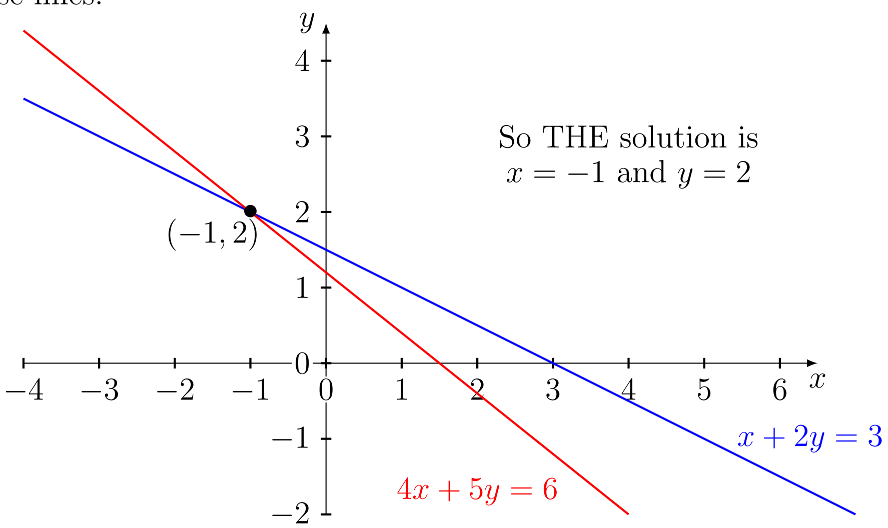
1.1 Introduction to system of linear equations

One of the central problem of linear algebra is '**solving linear equations**'. Consider the following **system of linear equations:**

$$\begin{aligned} x + 2y &= 3 \qquad \text{(1st equation)} \\ 4x + 5y &= 6 \qquad \text{(2nd equation).} \end{aligned} \tag{1.1}$$

Here x and y are the **unknowns**. We want to solve this system, i.e., we want to find the values of x and y in $\mathbb{R}$ such that the equations are satisfied.

1.1 (What does it mean geometrically?). Since the system (1.1) has two unknowns, we consider two dimensional euclidean plane. Draw the lines corresponding to each equation of the system. The set of solutions is nothing but the set of intersection points of these lines.



1.2 (How can we solve the system?). We can solve the system by **Gaussian Elimination**. The original system is

$$\begin{aligned} x + 2y &= 3 \qquad \text{(1st equation)} \\ 4x + 5y &= 6 \qquad \text{(2nd equation)}. \end{aligned} \tag{1.2}$$

We want to change it into an equivalent system, which is comparatively easy to solve. Eliminating x from the 2nd equation, we obtain a **triangulated** system:

$$\begin{aligned} x + 2y &= 3 \qquad \text{(equation1)} \\ -3y &= -6 \qquad \text{(equation 2) - 4(equation 1)}. \end{aligned} \tag{1.3}$$

Observe that both the systems (1.2) and (1.3) have same solutions by verifying that (x, y) is a solution of (1.2) if and only if it is a solution of (1.3). We can solve the 2nd system by **Back-substitution**. What is it? From the 2nd equation of (1.3), we get the value $y = 2$. Then we substitute it in the 1st equation to get $x = -1$. So the solution is $x = -1$ and $y = 2$.

1.3 (Another method to solve the system: Cramer's Rule). The system can be written as

$$\begin{aligned} x + 2y &= 3 \\ 4x + 5y &= 6 \end{aligned} \quad \text{or} \quad \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

The solution depends completely on those six numbers involved in the system. There must be a formula for x and y in terms of those six numbers. Cramer's Rule provides that formula:

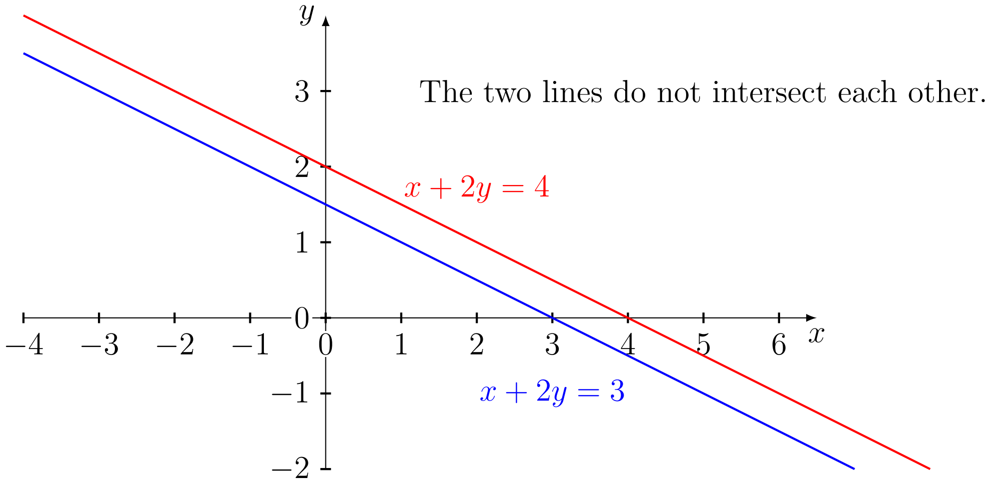
$$x = \frac{\begin{vmatrix} 3 & 2 \\ 6 & 5 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}} = \frac{3 \cdot 5 - 6 \cdot 2}{1 \cdot 5 - 4 \cdot 2} = \frac{3}{-3} = -1 \quad \text{and} \quad y = \frac{\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix}}{\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix}} = \frac{1 \cdot 6 - 4 \cdot 3}{1 \cdot 5 - 4 \cdot 2} = \frac{-6}{-3} = 2.$$

1.4 (Which approach is better?). The direct use of the **determinant formula** for large number of equations and variables would be very difficult. So the better method is Gaussian Elimination. Let's study it systematically. We understand the Gaussian Elimination method by examples.

1.5 (How many solutions do exist for a given system?). A system may have only ONE solution. For example, the system (1.2) has only one solution. A system may NOT have a solution at all. For example, the system:

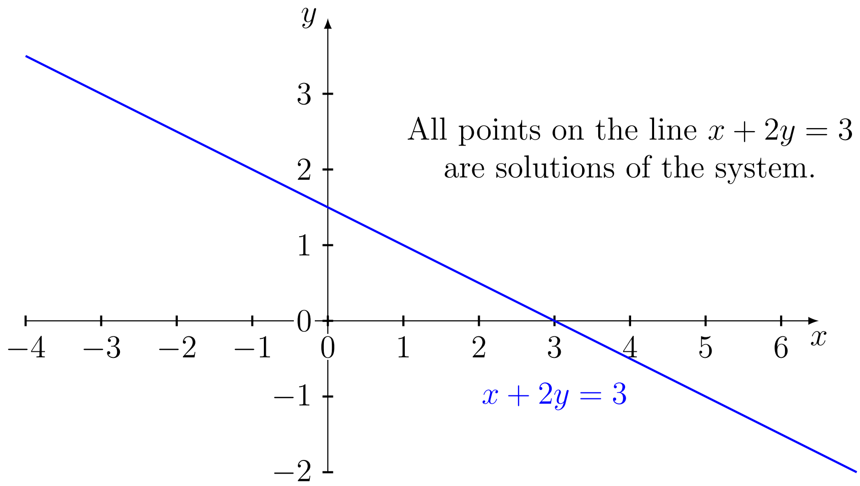
$$\left.\begin{aligned} x + 2y &= 3 \\ x + 2y &= 4 \end{aligned}\right\}. \quad \text{After Gaussian Elimination it becomes} \quad \begin{aligned} x + 2y &= 3 \\ \mathbf{0} &\mathbf{= 1} \end{aligned}$$

which is absurd. So the system does not have solutions. Geometrically,



A system may have INFINITELY many solutions. For example, the system:

$$\left.\begin{aligned} x + 2y &= 3 \\ 2x + 4y &= 6 \end{aligned}\right\} \quad \text{which is equivalent to the system of one equation } x + 2y = 3.$$



1.2 Gaussian Elimination and some examples

The Gaussian Elimination process in short is the following:

Original System

$\Downarrow$ Forward Elimination

Triangular System

$\Downarrow$ Backward Substitution

Solution(s)

Example 1.6. Consider the system:

$$\begin{aligned} v + w &= 5 \\ 4u - 6v &= -2 \\ -2u + 7v + 2w &= 9 \end{aligned} \tag{1.4}$$

There is no harm to interchange the positions of two equations. So the original system is equivalent to the following system:

$$\begin{aligned} 4u - 6v \quad & = -2 \\ v + w \quad & = 5 \\ -2u + 7v + 2w \quad & = 9 \end{aligned} \tag{1.5}$$

What was the aim? We want to change the system so that the coefficient of u in the 1st equation becomes non-zero. We call that non-zero coefficient of u as the **1st pivot**, because using this coefficient, we eliminate u from all other equations. Note that there is no harm if we multiply an equation by a NON-ZERO constant. So we can always make the pivot entry 1 by multiplying by suitable non-zero scalar. We now eliminate u from the 3rd equation. We change the 3rd equation by adding (1/2) times the 1st equation to the 3rd equation, and what we obtain is the following:

$$\begin{aligned} \mathbf{4}u - 6v \quad & = -2 \\ \mathbf{1} \cdot v + w \quad & = 5 \\ 4v + 2w \quad & = 8 \end{aligned} \tag{1.6}$$

One may check that something is a solution of (1.5) if and only if that is a solution of (1.6). Now in the 2nd step, we repeat the same process with the last two equations. Note that we already have the 2nd pivot. So we eliminate v from the 3rd equation by using the 2nd pivot. Change the 3rd equation into (3rd eqn) $-$ 4 (2nd eqn). This will complete the elimination process, and we obtain a **triangular system**:

$$\begin{aligned} \mathbf{4}u - 6v + 0w \quad & = -2 \\ \mathbf{1} \cdot v + w \quad & = 5 \\ \mathbf{-2}w \quad & = -12 \end{aligned} \tag{1.7}$$

Now the system (1.7) can be solved by **backward substitution**, bottom to top. The red colored coefficients are pivots. The last equation gives $w = 6$. Substituting $w = 6$ into the 2nd equation, we find $v = -1$. Substituting $w = 6$ and $v = -1$ into the 1st equation, we get $u = -2$. So we have only one solution: $u = -2$, $v = -1$ and $w = 6$.

Consider the original system (1.4). The **coefficient matrix** of (1.4) is defined to be

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix};$$

while the **augmented matrix** of (1.4) is defined to be

$$(A \,|\, b) = \begin{bmatrix} 0 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{bmatrix}.$$

The forward elimination steps can be described as follows.

$$\begin{bmatrix} 0 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{bmatrix} \overset{R1 \leftrightarrow R2}{\Longrightarrow} \begin{bmatrix} 4 & -6 & 0 & -2 \\ 0 & 1 & 1 & 5 \\ -2 & 7 & 2 & 9 \end{bmatrix} \overset{R3 \to R3+(1/2)R1}{\Longrightarrow}$$
$$\begin{bmatrix} 4 & -6 & 0 & -2 \\ 0 & 1 & 1 & 5 \\ 0 & 4 & 2 & 8 \end{bmatrix} \overset{R3 \to R3-4\cdot R2}{\Longrightarrow} \begin{bmatrix} 4 & -6 & 0 & -2 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & -2 & -12 \end{bmatrix}$$

Hence one can solve the system corresponding to the last augmented matrix by back substitution. In this case, where we have a full set of 3 pivots, there is only one solution.

Next we see some examples where we will have less pivots than 3, i.e., a zero appears in a pivot position. In this case, the system may not have a solution at all, or it can have infinitely many solutions. See the following examples.

Example 1.7. Consider the augmented matrix:

$$\begin{bmatrix} 1 & 1 & 1 & * \\ 2 & 2 & 5 & * \\ 4 & 4 & 8 & * \end{bmatrix}. \quad \text{After applying forward elimination, it becomes} \quad \begin{bmatrix} 1 & 1 & 1 & * \\ 0 & \mathbf{0} & 3 & * \\ 0 & 0 & 4 & * \end{bmatrix}.$$

Now consider some particular values of $*$. For example,

$$\begin{bmatrix} 1 & 1 & 1 & * \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 4 & 7 \end{bmatrix} \overset{R3-(4/3)R2}{\Longrightarrow} \begin{bmatrix} 1 & 1 & 1 & * \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

The corresponding system is

$$\begin{aligned} u + v + w &= * \\ 3w &= 6 \\ 0 &= -1 \end{aligned}$$

This system does not have a solution, because 0 is never equal to -1.

With another values of $*$, we consider

$$\begin{bmatrix} 1 & 1 & 1 & * \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 4 & 8 \end{bmatrix} \overset{R3-(4/3)R2}{\Longrightarrow} \begin{bmatrix} 1 & 1 & 1 & * \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The corresponding system is

$$\begin{aligned} u + \mathbf{v} + w &= * \\ 3w &= 6 \end{aligned}$$

This system has infinitely many solutions. From the last equation, we get $w = 2$. Substituting $w = 2$ to the 1st equation, we have $u + v = *$, which has infinitely many solutions.

1.3 System of linear equations (in general)

Consider a system of m linear equations in n variables $x_1, \ldots, x_n$.

$$
\begin{aligned}
A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\
A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\
&\vdots \\
A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n &= b_m
\end{aligned}
\tag{1.8}
$$

Here $A_{ij}, b_i \in \mathbb{R}$, and $x_1, \ldots, x_n$ are unknown. We denote the system by $AX = b$, where A is called the coefficient matrix, X is the n tuple of unknowns, and b is an n tuple of elements of $\mathbb{R}$. We try to find the values of $x_1, \ldots, x_n$ in $\mathbb{R}$ satisfied by the system. Any n tuple $(x_1, \ldots, x_n)$ of elements of $\mathbb{R}$ which satisfies the system (i.e., which satisfies every equation of the system) is called a **solution** of the system. If $b_1 = \cdots = b_m = 0$, then it is called a **homogeneous system**, which is denoted by $AX = 0$. Every homogeneous system has a trivial solution $x_1 = \cdots = x_n = 0$. On the other hand, a **non-homogeneous** system may or may not have solutions. We always can apply Gaussian Elimination to the system $AX = b$ to check whether solution exists or not, and then solve it if solution exists. In this process, in order to avoid writing equations and variables, we apply elementary row operations on A for a homogeneous system and $(A \,|\, b)$ for a non-homogeneous system.

1.8. Linear combination (cf. Definition 2.5) of equations of the system (1.8) yields another equation, e.g., c(1st eqn) + d(2nd eqn) for some $c, d \in \mathbb{R}$:

$$(cA_{21} + dA_{21})x_1 + (cA_{12} + dA_{22})x_2 + \cdots + (cA_{1n} + dA_{2n})x_n = (cb_1 + db_2).$$

If $(x_1, \ldots, x_n) \in \mathbb{R}^n$ satisfies (1.8), then it satisfies any such linear combination also. Consider another system of linear equations:

$$
\begin{aligned}
B_{11}x_1 + B_{12}x_2 + \cdots + B_{1n}x_n &= b'_1 \\
B_{21}x_1 + B_{22}x_2 + \cdots + B_{2n}x_n &= b'_2 \\
&\vdots \\
B_{m1}x_1 + B_{m2}x_2 + \cdots + B_{mn}x_n &= b'_m
\end{aligned}
\tag{1.9}
$$

Suppose every equation in (1.9) is a linear combination of the equations in (1.8). Then every solution of (1.8) is also a solution of (1.9). Note that in this case a solution of (1.9) need not be a solution of (1.8).

Definition 1.9. Consider the systems (1.8) and (1.9). Suppose every equation in the 2nd system is a linear combination of the equations of the 1st system, and vice versa. Then we say that these two systems are **equivalent**.

Definition 1.10. In total there are three types of elementary row operations:

(1) Interchange of two rows of A, say rth and sth rows, where $r \neq s$.

(2) Multiplication of one row of A by a non-zero scalar $c \in \mathbb{R}$.

(3) Replacement of the rth row of A by (rth row + $c \cdot s$th row), where $c \in \mathbb{R}$ and $r \neq s$.

1.11. All the above three operations are invertible, and each has inverse operation of the same type. The 1st one is it's own inverse. For the 2nd one, inverse operation is 'multiplication of that row of A by $1/c \in \mathbb{R}$'. For the 3rd one, inverse operation is 'replacement of the rth row of A by (rth row $- c \cdot s$th row)'.

Definition 1.12. Let A and B be two $m \times n$ matrices over $\mathbb{R}$. We say that B **is row equivalent to** A if B can be obtained from A by a finite sequence of elementary row operations, i.e.,

$$B = e_r \cdots e_2 e_1(A),$$

where e_i are some elementary row operations.

1.13. 'Row equivalence' is an '**equivalence relation**': 'Row equivalence' is reflexive, i.e., A is row equivalent to A. 'Row equivalence' is symmetric, i.e.,

$$B = e_r \cdots e_2 e_1(A) \quad \Longrightarrow \quad A = (e_1)^{-1}(e_2)^{-1} \cdots (e_r)^{-1}(B).$$

So, we just say that A and B are row equivalent. 'Row equivalence' is transitive, i.e., if B is row equivalent to A and C is row equivalent to B, then C is row equivalent to A.

1.14. Among three elementary row operations, considering row operation of each type, we observe that two matrices $(A \,|\, b)$ and $(A' \,|\, b')$ are row equivalent if and only if the corresponding systems $Ax = b$ and $A'x = b'$ are equivalent. In this case, both the systems have exactly the same solutions. Particularly, in case of homogeneous systems, it follows that A and A' are row equivalent if and only if the corresponding systems $Ax = 0$ and $A'x = 0$ are equivalent.

1.4 Row reduced echelon matrices

Definition 1.15. An $m \times n$ matrix A over $\mathbb{R}$ is called **row reduced** if
(1) the 1st non-zero entry in each non-zero row of A is equal to 1;
(2) each column of A which contains the leading non-zero entry of some row has all its other entries 0.

Example 1.16. Observe that (i), (ii) and (iii) are row reduced matrices; while (iv) and (v) are not row reduced matrices (because of the red colored entries):

$$\text{(i)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{(ii)} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{(iii)} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix} \text{(iv)} \begin{pmatrix} 0 & {\color{red}2} & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{pmatrix} \text{(v)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & {\color{red}-1} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

By a zero row of a matrix, we mean all entries of that row are zero; while a non-zero row means at least one entry is non-zero.

Definition 1.17. An $m \times n$ matrix A over $\mathbb{R}$ is called **row reduced echelon** matrix if
(1) A is row reduced;
(2) every zero row of A occurs below every non-zero row;
(3) if rows $1, \dots, r$ are the non-zero rows, and if the leading non-zero entry of row i occurs in column k_i for $1 \leqslant i \leqslant r$, then $k_1 < k_2 < \cdots < k_r$.

1.18. In Definition 1.17, (i, k_i) are called the pivot positions, and the corresponding variables x_{k_i} are called the pivot variables, while the remaining variables are called free variables. The justification of these nomenclature will be discussed later.

Example 1.19. In Example 1.16, (iv) and (v) are not row reduced, hence they are not row reduced echelon. It can be observed that (ii) and (iii) of Example 1.16 are row reduced, but they are not row reduced echelon.

Exercise 1.20. Let A be an $n \times n$ row reduced echelon matrix over $\mathbb{R}$. Show that A is invertible if and only if A is the identity matrix.

Theorem 1.21. Every $m \times n$ matrix over $\mathbb{R}$ is **row equivalent** to a row reduced echelon matrix over $\mathbb{R}$. More precisely, if A is an $m \times n$ matrix over $\mathbb{R}$, then there is a row reduced echelon matrix B of the same order such that B can be obtained from A by applying finitely many elementary row operations.

Example 1.22. $A = \begin{bmatrix} 0 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{bmatrix}$ is row equivalent to $B = \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 6 \end{bmatrix}$, which is a row reduced echelon matrix.

Proof. $\begin{bmatrix} 0 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{bmatrix} \xRightarrow{R1 \leftrightarrow R2} \begin{bmatrix} 4 & -6 & 0 & -2 \\ 0 & 1 & 1 & 5 \\ -2 & 7 & 2 & 9 \end{bmatrix} \xRightarrow{R1 \to (1/4)R1} \begin{bmatrix} 1 & -3/2 & 0 & -1/2 \\ 0 & 1 & 1 & 5 \\ -2 & 7 & 2 & 9 \end{bmatrix}$

$\xRightarrow{R3 \to R3+2 \cdot R1} \begin{bmatrix} 1 & -3/2 & 0 & -1/2 \\ 0 & 1 & 1 & 5 \\ 0 & 4 & 2 & 8 \end{bmatrix} \xRightarrow{R3 \to R3-4 \cdot R2} \begin{bmatrix} 1 & -3/2 & 0 & -1/2 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & -2 & -12 \end{bmatrix} \xRightarrow{R3 \to (-1/2)R3}$

$\begin{bmatrix} 1 & -3/2 & 0 & -1/2 \\ 0 & 1 & 1 & 5 \\ 0 & 0 & 1 & 6 \end{bmatrix}$ (Corresponds to a triangular system with pivot entries 1)

$\xRightarrow{R2 \to R2-R3} \begin{bmatrix} 1 & -3/2 & 0 & -1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 6 \end{bmatrix} \xRightarrow{R1 \to R1+(3/2)R2} \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 6 \end{bmatrix}$.

So it is just combination of forward and backward eliminations.

Remark 1.23. The non-homogeneous systems corresponding to the augmented matrices A and B have the same solutions. In this case they have only one solution $x_1 = -2$, $x_2 = -1$ and $x_3 = 6$.

Example 1.24. Consider the homogeneous system corresponding to the coefficient matrix

$$\begin{bmatrix} 0 & 1 & -3 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ (which is a row reduced echelon matrix).}$$

Here $(1,2)$ and $(2,4)$ are the pivot positions. They belong to 2nd and 4th columns of the matrix. So x_2 and x_4 are the pivot variables. The remaining variables (i.e., x_1, x_3 and x_5) are called free variables, because their values can be chosen freely to obtain the solutions of the corresponding system:

$$\begin{array}{ccccc} x_2 & -3x_3 & & +(1/2)x_5 & = & 0 \\ & & x_4 & +2x_5 & = & 0 \end{array}$$

which yields that

$$\begin{aligned} x_2 &= 3x_3 - (1/2)x_5 \\ x_4 &= -2x_5 \end{aligned}$$

So THE solutions of the corresponding system are given by

$$\begin{pmatrix} x_1 \\ 3x_3 - (1/2)x_5 \\ x_3 \\ -2x_5 \\ x_5 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 3 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 0 \\ -(1/2) \\ 0 \\ -2 \\ 1 \end{pmatrix}, \quad \text{where } x_1, x_3, x_5 \in \mathbb{R}.$$

1.25 (How to solve a homogeneous system corr. to a row reduced echelon matrix?)**.** Consider $Ax = 0$, where A is a row reduced echelon matrix. Let the rows $1, \ldots, r$ be non-zero, and the leading non-zero entry of row i occurs in column k_i. The system $Ax = 0$ then consists of r non-trivial equations. The variables $x_{k_1}, x_{k_2}, \ldots, x_{k_r}$ are the pivot variables. Let $u_1, \ldots, u_{n-r}$ denote the remaining $n - r$ (free) variables. Then the r non-trivial equations of $Ax = 0$ can be written as

$$x_{k_1} + \sum_{j=1}^{n-r} C_{1j} u_j = 0$$

$$\cdots$$

$$x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j = 0$$

We may assign any values to $u_1, \ldots, u_{n-r}$. Then $x_{k_1}, \ldots, x_{k_r}$ are determined uniquely by those assigned values; see Example 1.24.

Theorem 1.26. Let A be an $m \times n$ matrix over $\mathbb{R}$ with $m < n$. Then the homogeneous system $Ax = 0$ has a non-trivial solution. In fact (over $\mathbb{R}$) it has infinitely many solutions.

Proof. By Theorem 1.21, A is row equivalent to a row reduced echelon matrix B. Then 1.14 yields that $Ax = 0$ and $Bx = 0$ have the same solutions. If r is the number of non-zero rows of B, then $r \leqslant m < n$. The system $Bx = 0$ can be expressed as

$$x_{k_1} + \sum_{j=1}^{n-r} C_{1j} u_j = 0, \ \ldots, \ x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j = 0$$

where $u_1, \ldots, u_{n-r}$ are the free variables. Now assign any values to $u_1, \ldots, u_{n-r}$ to get infinitely many solutions. $\square$

Theorem 1.27. Let A be an $n \times n$ matrix over $\mathbb{R}$. Then A is row equivalent to the $n \times n$ identity matrix if and only if the system $Ax = 0$ has only the trivial solution.

Proof. By Theorem 1.21, A is row equivalent to a row reduced echelon matrix B. Then 1.14 yields that $Ax = 0$ and $Bx = 0$ have the same solutions. If r is the number of non-zero rows of B, then $r \leqslant n$. Moreover, the system $Bx = 0$ can be expressed as

$$x_{k_1} + \textstyle\sum_{j=1}^{n-r} C_{1j} u_j = 0, \ \ldots, \ x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j = 0,$$

where $u_1, \ldots, u_{n-r}$ are the free variables. Hence it can be observed that $B = I_n$ is the identity matrix if and only if $r = n$ if and only if the system has the trivial solution. $\square$

1.28 (How to solve a non-homogeneous system $Ax = b$)**.** Consider the augmented matrix $(A \,|\, b)$ corresponding to a non-homogeneous system $Ax = b$. Apply elementary row operations on $(A \,|\, b)$ to get row reduced echelon form $(B \,|\, c)$. The systems $Ax = b$ and $Bx = c$ are equivalent, and hence they have the same solutions. Let $1, \ldots, r$ be the non-zero rows of B, and the leading non-zero entry of row i occurs in column k_i. The variables $x_{k_1}, x_{k_2}, \ldots, x_{k_r}$ are the pivot variables. Let $u_1, \ldots, u_{n-r}$ denote the remaining $n - r$ (free) variables. Then the system $Bx = c$ can be written as

$$\begin{aligned} x_{k_1} + \sum_{j=1}^{n-r} C_{1j} u_j &= c_1, \\ &\vdots \\ x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j &= c_r \\ 0 &= c_{r+1} \\ &\vdots \\ 0 &= c_m \end{aligned}$$

The system $Ax = b$ (equivalenly, $Bx = c$) has a solution if and only if $c_{r+1} = \cdots = c_m = 0$. Furthermore, in this case, i.e., when $c_{r+1} = \cdots = c_m = 0$, then $r = n$ if and only if the system has a unique solution, and $r < n$ if and only if the system has infinitely many solutions. In the last case, we may assign any values to $u_1, \ldots, u_{n-r}$. Then $x_{k_1}, \ldots, x_{k_r}$ are determined uniquely by those assigned values; see Example 1.29.

Example 1.29. Set $A := \begin{pmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{pmatrix}$. Let $b \in \mathbb{R}^3$ be a column vector. Consider the system $Ax = b$. Applying elementary row operations on the corresponding augmented matrix $(A \,|\, b)$, we get $\begin{pmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(b_1 + 2b_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(b_2 - 2b_1) \\ 0 & 0 & 0 & (b_3 - b_2 + 2b_1) \end{pmatrix}$. In view of 1.28, the system $Ax = b$ has a solution if and only if $b_3 - b_2 + 2b_1 = 0$. In this case, $x_1 = -\frac{3}{5}x_3 + \frac{1}{5}(b_1 + 2b_2)$ and $x_2 = \frac{1}{5}x_3 + \frac{1}{5}(b_2 - 2b_1)$. Assign any value to x_3, and compute x_1, x_2 to get infinitely many solutions of $Ax = b$.

1.5 Elementary matrices

Definition 1.30. An $m \times m$ matrix is called an **elementary matrix** if it can be obtained from the $m \times m$ identity matrix by applying a single elementary row operation.

Example 1.31.
(i) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (obtained by applying the 1st type elementary row operation on I_2).
(ii) $\begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix}$, $c \neq 0$. (iii) $\begin{pmatrix} 1 & 0 \\ 0 & c \end{pmatrix}$, $c \neq 0$. (... 2nd type ...).
(iv) $\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}$, $c \in \mathbb{R}$. (v) $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$, $c \in \mathbb{R}$. (... 3rd type ...).
These are all the 2×2 elementary matrices.

Example 1.32. Consider a matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$. Observe that

(i) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \end{pmatrix}$ (which is same as applying $R1 \leftrightarrow R2$).

(ii) $\begin{pmatrix} c & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} c & 2c & 3c \\ 4 & 5 & 6 \end{pmatrix}$ (which is same as applying $R1 \to c \cdot R1$).

(iii) $\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4 + c & 5 + 2c & 6 + 3c \end{pmatrix}$ (... applying $R2 \to R2 + c \cdot R1$).

Example 1.32 gives us evidence that applying an elementary row operation on a matrix is same as left multiplying by the corresponding elementary matrix. More precisely, we have the following result.

Theorem 1.33. Let e be an elementary row operation. Let E be the corresponding $m \times m$ elementary matrix, i.e., $E = e(I_m)$ (in other words, E is obtained by applying e on I_m), where I_m is the $m \times m$ identity matrix. Then, for every $m \times n$ matrix A,

$$EA = e(A).$$

As an immediate consequence, we obtain the following:

Corollary 1.34. Let A and B be two $m \times n$ matrices. Then A and B are row equivalent if and only if $B = PA$, where P is a product of some $m \times m$ elementary matrices.

Theorem 1.35. Every elementary matrix is invertible.

Proof. Let E be an elementary matrix corresponding to the elementary row operation e, i.e., $E = e(I)$. In view of 1.11, e has an inverse operation, say e'. Set $E' := e'(I)$. Then

$$EE' = e(E') = e(e'(I)) = I \text{ and } E'E = e'(E) = e'(e(I)) = I.$$

So E is invertible (with inverse matrix E'). □

Remark 1.36. It can be observed from the proof of Theorem 1.35 that the inverse of an elementary matrix is again an elementary matrix.

Theorem 1.37. Let A be an $n \times n$ matrix. Then the following are equivalent:
(i) A is invertible.
(ii) A is row equivalent to the $n \times n$ identity matrix.
(iii) A is a product of some elementary matrices.

Proof. In view of Theorem 1.21, let A be row-equivalent to a row-reduced echelon matrix B. Then, by Theorem 1.33, there are some elementary matrices $E_1, \ldots, E_k$ such that

$$B = E_k \cdots E_2 E_1 A. \tag{1.10}$$

Since elementary matrices are invertible, we have

$$E_1^{-1} E_2^{-1} \cdots E_k^{-1} B = A. \tag{1.11}$$

Since elementary matrices are invertible, and product of invertible matrices is invertible, it follows from (1.10) and (1.11) that A is invertible if and only if B is invertible, which is equivalent to that $B = I_n$ by Exercise 1.20. This proves that (i) and (ii) are equivalent. Hence the proof of the theorem is complete once we show the following implications.

(iii) $\Rightarrow$ (i). It follows from Theorem 1.35.

(ii) $\Rightarrow$ (iii). If $B = I_n$, then (1.11) yields that $A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$, which is a product of elementary matrices (by Remark 1.36). □

Theorem 1.38. Let A be an $n \times n$ invertible matrix. If a sequence of elementary row operations reduces A to the identity matrix I_n, then that same sequence of operations when applied to I_n yields A^{-1}.

Proof. If a sequence of elementary row operations reduces A to the identity matrix I_n, then by Theorem 1.33, there are some elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_2 E_1 A = I_n$. So A has left inverse $(E_k \cdots E_2 E_1)$. Therefore $A^{-1} = (E_k \cdots E_2 E_1)$ since A is invertible by Theorem 1.37. Thus $A^{-1} = E_k \cdots E_2 E_1 I_n$. Hence the proof follows from Theorem 1.33. □

1.39 (How to compute inverse of a matrix by elementary row operations)**.** In order to check whether a given $n \times n$ matrix A is invertible or not, we apply elementary row operations on A to make it into a row reduced echelon matrix B. By Theorem 1.37, A is invertible if and only if $B = I_n$. Next, if A is invertible, we compute A^{-1} by Theorem 1.38. The whole process can be done by applying elementary row operations on $(A \mid I_n)$ to make it into $(I_n \mid A^{-1})$ if A is invertible; see Example 1.40.

Example 1.40.
Let $A = \begin{pmatrix} 2 & -1 \\ 1 & 3 \end{pmatrix}$. Note that $(A \mid I_2) = \begin{pmatrix} 2 & -1 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix}$. Applying elementary row operations on $(A \mid I_2)$, we obtain the following:

$$\begin{pmatrix} 2 & -1 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} \overset{\frac{1}{2}R1}{\Longrightarrow} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} \overset{R2 \to R2 - R1}{\Longrightarrow} \begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{7}{2} & -\frac{1}{2} & 1 \end{pmatrix} \overset{\frac{2}{7}R2}{\Longrightarrow}$$
$$\begin{pmatrix} 1 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{7} & \frac{2}{7} \end{pmatrix} \overset{R1 \to R1 + \frac{1}{2}R2}{\Longrightarrow} \begin{pmatrix} 1 & 0 & \frac{3}{7} & \frac{1}{7} \\ 0 & 1 & -\frac{1}{7} & \frac{2}{7} \end{pmatrix}.$$

This process ensures that A is invertible and $A^{-1} = \begin{pmatrix} \frac{3}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{2}{7} \end{pmatrix}$.

Chapter 2

Vector Spaces

MA 1140 is the study of 'vector spaces' and the 'maps' between them. For now, you may consider $\mathbb{R}^n$ as an example of a vector space, where $\mathbb{R}$ is the set of real numbers. Essentially, a vector space means a collection of objects, we call them vectors, where we can add two vectors, and what we get is a vector; we can multiply a vector by a scalar, and what we get is a vector; cf. Figure 2.

2.1 Vector Space

Definition 2.1. A set V of objects (called vectors) along with vector addition '+' and scalar multiplication '$\cdot$' is said to be a **vector space** over a field $\mathbb{F}$ (say, $\mathbb{F} = \mathbb{R}$, the set of real numbers) if the following hold:

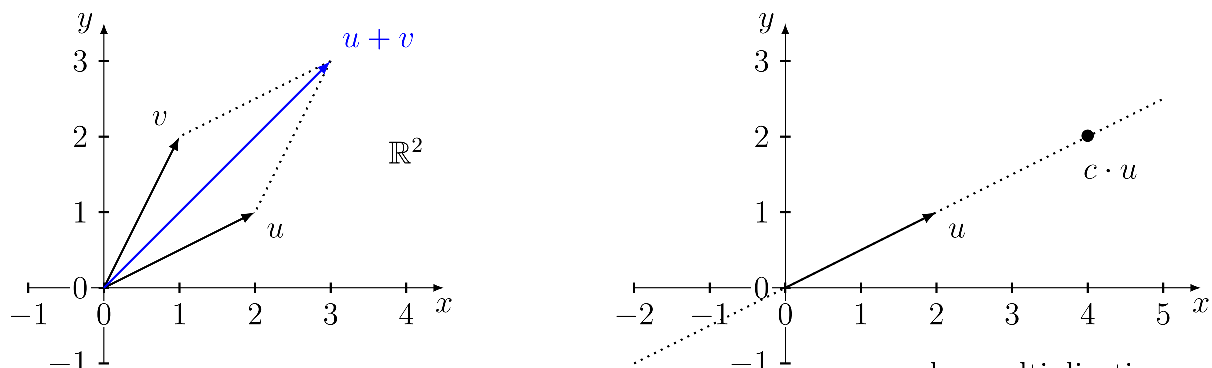
(1) V is closed under '+', i.e. $x + y \in V$ for all $x, y \in V$.

(2) Addition is commutative, i.e. $x + y = y + x$ for all $x, y \in V$.

(3) Addition is associative, i.e. $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$.

(4) Additive identity, i.e. there is $0 \in V$ such that $x + 0 = x$ for all $x \in V$.

(5) Additive inverse, i.e. for every $x \in V$, there is $-x \in V$ such that $x + (-x) = 0$.



vector addition

Figure 2

scalar multiplication

(6) V is closed under '$\cdot$', i.e. $c \cdot x \in V$ for all $c \in \mathbb{F}$ and $x \in V$.

(7) $1 \cdot x = x$ for all $x \in V$.

(8) $(ab) \cdot x = a \cdot (b \cdot x)$ for all $a, b \in \mathbb{F}$ and $x \in V$.

(9) $a \cdot (x + y) = a \cdot x + a \cdot y$ for all $a \in \mathbb{F}$ and $x, y \in V$.

(10) $(a + b) \cdot x = a \cdot x + b \cdot x$ for all $a, b \in \mathbb{F}$ and $x \in V$.

The elements of $\mathbb{F}$ are called **scalars**, and the elements of V are called **vectors**.

Remark 2.2. The first five properties are nothing but the properties of abelian group, i.e. $(V, +)$ is an abelian group.

We simply write cx instead of $c \cdot x$ for $c \in \mathbb{F}$ and $x \in V$ when there is no confusion.

From now, we work over the field $\mathbb{R}$.

Example 2.3. The following are examples of vector spaces.

(1) The n-tuple space, $V = \mathbb{R}^n$, where vector addition and scalar multiplication are defined as follows:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} \quad \text{and} \quad c \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_n \end{pmatrix}.$$

(2) The space $\mathbb{R}^{m \times n}$ of all $m \times n$ matrices

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \quad \text{where } x_{ij} \in \mathbb{R}.$$

The vector addition and scalar multiplication are defined by component wise addition and multiplication as in (1).

(3) Let S be any non-empty set. Let V be the set of all functions from S into $\mathbb{R}$. The sum $f + g$ of two vectors f and g in V is defined to be

$$(f + g)(s) := f(s) + g(s) \text{ for all } s \in S.$$

The scalar multiplication $c \cdot f$ is defined by $(c \cdot f)(s) := cf(s)$. Clearly, V is a vector space. Note that the preceding examples are special cases of this one.

(4) The set $\mathbb{R}[x]$ of all polynomials $a_0 + a_1 x + \cdots + a_m x^m$, where $a_i \in \mathbb{R}$, x is an indeterminate and m varies over non-negative integers. The vector addition and scalar multiplication are defined in obvious way. Then $\mathbb{R}[x]$ is a vector space over $\mathbb{R}$.

Example 2.4. The set $V = \mathbb{R}^{n\times n}$ of all $n \times n$ matrices with vector addition defined by

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} \times \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nn} \end{pmatrix}$$
$$= \begin{pmatrix} -- & -- & -- \\ -- & \left(\sum_{k=1}^{n} x_{ik}y_{kj}\right) & -- \\ -- & -- & -- \end{pmatrix} \quad \text{(matrix multiplication)}$$

and scalar multiplication as before is NOT a vector space. Indeed, the operation '$\times$' is **not commutative** because

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Moreover, every matrix does not necessarily have multiplicative inverse. For example, there does not exist a matrix A such that

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \times A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Definition 2.5. A vector β in V is said to be a **linear combination** of vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ in V if $\beta = c_1\alpha_1 + c_2\alpha_2 + \cdots + c_r\alpha_r$ for some $c_1, c_2, \ldots, c_r \in \mathbb{F}$.

Example 2.6. In $\mathbb{R}^2$, the vector $(1, 2)$ can be written as linear combinations of $\{(1, 0), (0, 2)\}$ and $\{(1, 1), (1, 0)\}$ respectively as

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix} + (-1)\begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

In view of Definition 2.5, one may ask the following questions. Suppose $\alpha_1, \alpha_2, \ldots, \alpha_r$ and β are given vectors in V. Is it possible to write β as a linear combination of $\alpha_1, \alpha_2, \ldots, \alpha_r$? If yes, then is it a unique way to write that? We find the answers to these questions as we proceed further.

Remark 2.7. In Figure 3 below, the set of all linear combinations of $\{u\}$ is given by the dotted line. So v cannot be written as a linear combination of $\{u\}$.

2.2 Subspaces

Definition 2.8. Let V be a vector space over a field $\mathbb{F}$. A subspace of V is a subset W of V which is itself a vector space over $\mathbb{F}$ with the same operations of vector addition and scalar multiplication on V.

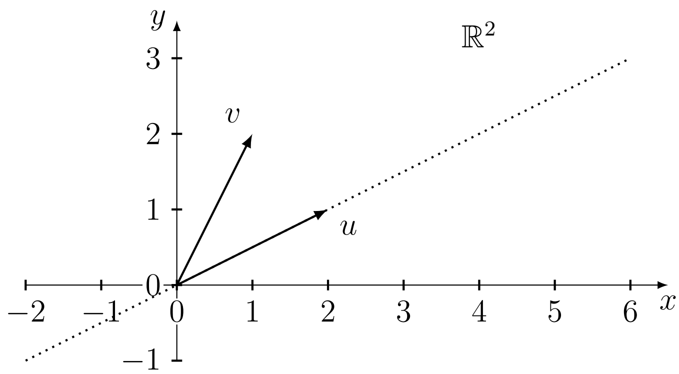


Figure 3

Theorem 2.9. Let W be a non-empty subset of a vector space V over $\mathbb{F}$. Then W is a subspace of V if and only if for each pair of vectors $\alpha, \beta \in W$ and each scalar $c \in \mathbb{F}$, the vector $c\alpha + \beta$ belongs to W.

Proof. Exercise! Note that many properties of W will be inherited from V. $\square$

Theorem 2.10. Let V be a vector space over a field $\mathbb{F}$. The intersection of any collection of subspaces of V is a subspace of V.

Proof. Exercise! Use Theorem 2.9. $\square$

Example 2.11. (1) The subset W consisting of the zero vector of V is a subspace of V.

(2) In $\mathbb{R}^n$, the set of n-tuples $(x_1, \ldots, x_n)$ with $x_1 = 0$ is a subspace; while the set of n-tuples with $x_1 = 1$ is NOT a subspace.

(3) The set of all 'symmetric matrices' forms a subspace of the space of all $n \times n$ matrices. Recall that an $n \times n$ square matrix A is said to be symmetric if (i, j)th entry of A is same as its (j, i)th entry, i.e. $A_{ij} = A_{ji}$ for each i and j.

Definition 2.12. Let S be a set of vectors in a vector space V. The **subspace spanned** by S is defined to be THE smallest subspace of V containing S. We denote this subspace by $\mathrm{Span}(S)$. We write that S spans the subspace $\mathrm{Span}(S)$.

What is the guarantee for existence of a subspace spanned by a given set? The following theorem is giving us that guarantee.

Theorem 2.13. Let S be a set of vectors in a vector space V. The following subspaces are equal.

(1) The intersection of all subspaces of V containing S.

(2) The set of all linear combinations of vectors in S, i.e. $\{c_1v_1 + \cdots + c_rv_r : c_i \in \mathbb{R},\ v_i \in S\}$. (One can check that it is a subspace.)

(3) The subspace $\mathrm{Span}(S)$, i.e. the smallest subspace of V containing S.

Proof. Let W_1, W_2 and W_3 be the subspaces described as in (1), (2) and (3) respectively. Then W_1 is contained in any subspace of V containing S. Therefore, since W_1 is a subspace (by Theorem 2.10), W_1 is the smallest subspace of V containing S, i.e. $W_1 = W_3$.

Using Theorem 2.9, it can be shown that W_2 is a subspace. Therefore, since W_2 contains S, we have $W_1 \subseteq W_2$. Notice that any subspace of V containing S also contains all linear combinations of vectors in S, i.e. any subspace of V containing S also contains W_2. Hence it follows that $W_2 \subseteq W_1$. Therefore $W_1 = W_2$. Thus $W_1 = W_2 = W_3$. □

Remark 2.14. In Figure 3, the subspace spanned by $\{u\}$ can be described by the dotted line; while the subspace spanned by $\{u, v\}$ is $\mathbb{R}^2$; see Example 2.20.

2.3 Basis and Dimension

Definition 2.15. Let V be a vector space over $\mathbb{R}$. A subset S of vectors in V is said to be **linearly dependent** if there exists (distinct) vectors $v_1, v_2, \ldots, v_r$ in S and scalars $c_1, c_2, \ldots, c_r$ in $\mathbb{R}$, not all of which are 0, such that

$$c_1 v_1 + \cdots + c_r v_r = 0.$$

A set S which is not linearly dependent is called **linearly independent**. If $S = \{v_1, v_2, \ldots, v_n\}$ is finite, we say that $v_1, v_2, \ldots, v_n$ are linearly dependent (or independent) instead of saying that S is so.

Remark 2.16. The following statements can be verified easily.

(1) Any set containing the 0 vector is linearly dependent.

(2) A set S of vectors is linearly dependent if and only if there exists a non-trivial relation of vectors of S:

$$c_1 v_1 + \cdots + c_r v_r = 0, \text{ where at least one } c_i \neq 0.$$

This is equivalent to say that there exists at least one vector $v \in S$ which belongs to the subspace spanned by $S \setminus \{v\}$. (For example, $c_1 v_1 + \cdots + c_r v_r = 0$ with $c_1 \neq 0$ implies that $v_1 = (-c_2/c_1)v_2 + \cdots + (-c_r/c_1)v_r$. Conversely, if $v_1 \in \mathrm{Span}(\{v_2, \ldots, v_r\})$, then there is a non-trivial relation $c_1 v_1 + \cdots + c_r v_r = 0$ with $c_1 = 1$.)

(3) Any set containing a linearly dependent subset is again linearly dependent.

(4) Every non-zero vector v in V is linearly independent.

(5) A finite set $\{v_1, \ldots, v_r\}$ is linearly independent if and only if

$$c_1 v_1 + \cdots + c_r v_r = 0 \quad \Longrightarrow \quad c_i = 0 \text{ for all } 1 \leqslant i \leqslant r.$$

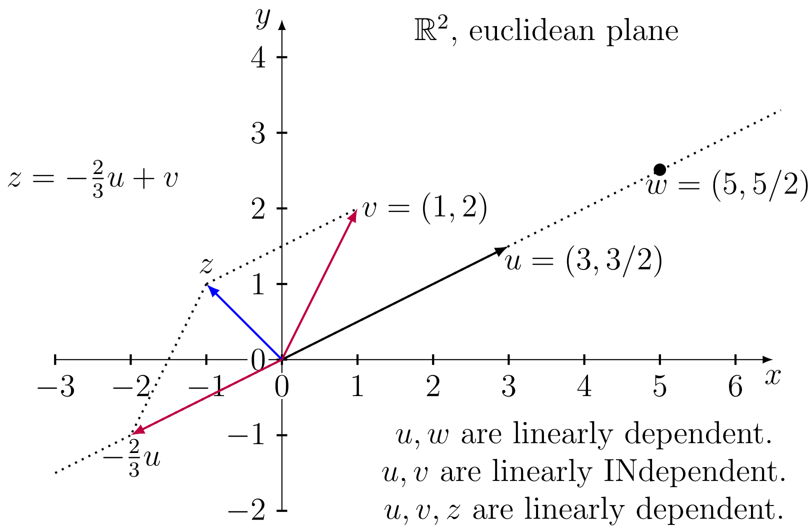
(6) A set S of vectors is linearly independent if and only if every finite subset of S is linearly independent, i.e., if and only if for every subset $\{v_1, \ldots, v_r\} \subseteq S$,

$$c_1v_1 + \cdots + c_rv_r = 0 \quad \Longrightarrow \quad c_i = 0 \text{ for all } 1 \leqslant i \leqslant r.$$

(7) Any subset of a linearly independent set is linearly independent.

Lemma 2.17. Let S be a linearly independent subset of a vector space V. Suppose $v \notin \mathrm{Span}(S)$. Then $S \cup \{v\}$ is linearly independent.

Proof. Let $c_1v_1 + \cdots + c_rv_r + cv = 0$ for some (distinct) vectors $v_1, \ldots, v_r \in S$ and scalars $c_1, \ldots, c_r, c$. If $c \neq 0$, then it follows that $v \in \mathrm{Span}(S)$, which is a contradiction. Therefore $c = 0$, and hence $c_1v_1 + \cdots + c_rv_r = 0$. Since S is linearly independent, it follows that $c_i = 0$ for every $1 \leqslant i \leqslant r$. $\square$



Definition 2.18. Let V be a vector space over $\mathbb{R}$. A set S of vectors in V is called a **basis** of V if S is linearly independent and it spans the space V (i.e., the subspace spanned by S is V).

What is the guarantee that a basis exists? We can prove the existence at least when V is generated (or spanned) by finitely many vectors. How? Start with a finite spanning set S. Then check whether it is linearly independent. If S is linearly dependent, then there is $v \in S$ such that v belongs to the subspace spanned by $S \smallsetminus \{v\}$. One can prove that $S \smallsetminus \{v\}$ spans V. Now repeat the process till we get a linearly independent subset of S which spans V.

For vector space which is not finitely generated, we need the axiom of choice. We will not do that in this course.

The space V is said to be finite dimensional if it has a finite basis. If V does not have a finite basis, then V is said to be infinite dimensional.

Example 2.19. In $\mathbb{R}^2$, the vectors $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$ are linearly dependent because $(7/3)v_1 + (1/3)v_2 + (-1)v_3 = 0$.

Example 2.20. The set $\left\{ v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$ forms a basis of $\mathbb{R}^2$. Indeed, geometrically, it can be observed that v_1, v_2 are linearly independent, and $\{v_1, v_2\}$ spans $\mathbb{R}^2$. Or directly, we see that for EVERY vector $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$, the system

$$x \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{i.e.,} \quad \begin{cases} x + 2y = a \\ 2x + y = b \end{cases} \quad \text{or} \quad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$$

has a UNIQUE solution in x, y because the coefficient matrix $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ is invertible. So every vector in $\mathbb{R}^2$ can be written as a linear combination of $\{v_1, v_2\}$, hence it spans the space $\mathbb{R}^2$. Moreover, when $a = b = 0$, then the system has THE trivial solution $x = y = 0$. Thus $\{v_1, v_2\}$ is linearly independent as well.

Example 2.21. In $\mathbb{R}^n$, let S be the subset consisting of the vectors:

$$\begin{aligned} e_1 &= (1, 0, 0 \ldots, 0) \\ e_2 &= (0, 1, 0, \ldots, 0) \\ &\vdots \\ e_n &= (0, 0, 0, \ldots, 1). \end{aligned}$$

Note that any vector $v = (x_1, \ldots, x_n) \in \mathbb{R}^n$ can be written as a linear combination $x_1e_1 + \cdots + x_ne_n$. So S spans $\mathbb{R}^n$. Moreover, it can be shown that S is linearly independent. Therefore S is a basis of $\mathbb{R}^n$. This particular basis is called the **standard basis** of $\mathbb{R}^n$.

Exercise 2.22. Show that for $\mathbb{R}[x]$, the set of all polynomials over $\mathbb{R}$, the subset

$$S = \{x^n : n = 0, 1, 2, \ldots\}$$

forms a basis.

Lemma 2.23. *Let V be a vector space over $\mathbb{R}$. Suppose $\{v_1, v_2, \ldots, v_n\}$ spans V. Let u be a non-zero vector in V. Then some v_i can be replaced by u to get another spanning set of V, i.e., if necessary, then after renaming the vectors $\{v_1, v_2, \ldots, v_n\}$, we obtain that $\{u, v_2, \ldots, v_n\}$ spans V.*

Proof. Since $\{v_1, v_2, \ldots, v_n\}$ spans V, u can be written as

$$u = c_1v_1 + c_2v_2 + \cdots + c_nv_n \text{ for some } c_1, \ldots, c_n \in \mathbb{R}. \tag{2.1}$$

Hence, since $u \neq 0$, at least one $c_i \neq 0$. Therefore (2.1) yields that

$$v_i = (1/c_i)u + (c_1/c_i)v_1 + \cdots + (c_{i-1}/c_i)v_{i-1} + (c_{i+1}/c_i)v_{i+1} + \cdots + (c_n/c_i)v_n. \quad (2.2)$$

Since any vector v in V can be written as a linear combination of $v_1, v_2, \dots, v_n$, using (2.2) in that linear combination, it follows that v can be written as a linear combination of $v_1, \dots, v_{i-1}, u, v_{i+1}, \dots, v_n$. Thus $\{v_1, \dots, v_{i-1}, u, v_{i+1}, \dots, v_n\}$ spans V. □

Theorem 2.24. Let V be a vector space over $\mathbb{R}$. Suppose V is spanned by a finite set $\{v_1, v_2, \dots, v_n\}$ of n vectors, and $\{u_1, u_2, \dots, u_m\}$ is a linearly independent set of m vectors in V. Then m is finite and $m \leqslant n$.

Proof. If possible, let $n < m$. Note that every $u_i \neq 0$. In view of Lemma 2.23, if necessary, by renaming the vectors $v_1, \dots, v_n$, we have that $\{u_1, v_2, v_3, \dots, v_n\}$ spans V.

In the 2nd step, since $u_2 \in V = \operatorname{Span}\{u_1, v_2, v_3, \dots, v_n\}$, it follows that

$$u_2 = b_1u_1 + b_2v_2 + b_3v_3 + \cdots + b_nv_n \quad \text{for some } b_i \in \mathbb{R}.$$

Then at least one of $\{b_2, \dots, b_n\}$ is non-zero, otherwise if $b_i = 0$ for all $2 \leqslant i \leqslant n$, then $\{u_1, u_2\}$ is linearly dependent, which contradicts the hypotheses. Thus, as in the proof of Lemma 2.23, if necessary, by renaming the vectors $v_2, \dots, v_n$, we have that $\{u_1, u_2, v_3, \dots, v_n\}$ spans V.

Continuing in this way, after n steps, we obtain that $\{u_1, u_2, \dots, u_n\}$ spans V. Hence

$$u_{n+1} \in V = \operatorname{Span}\{u_1, u_2, \dots, u_n\}.$$

This yields that $\{u_1, u_2, \dots, u_{n+1}\}$ is linearly dependent, which contradicts the hypotheses. Therefore $m \leqslant n$. □

Here are some consequences of Theorem 2.24.

Corollary 2.25. If V is a finite dimensional vector space, then any two bases of V have the same number of elements.

Proof. Since V is finite dimensional, it has a finite basis $\{v_1, \dots, v_n\}$. If $\{u_1, \dots, u_m\}$ is a basis of V, then by Theorem 2.24, m is finite and $m \leqslant n$. By the same argument, $n \leqslant m$. Thus $m = n$. □

Corollary 2.25 allows us to define the **dimension** of a finite dimensional vector space as the number of elements in a basis. We denote the dimension of a finite dimensional vector space V by $\dim(V)$. If V is not finite dimensional, then we set $\dim(V) := \infty$. Note that $\dim(V) = 0$ if and only if $V = 0$ is the trivial space.

Corollary 2.26. Let V be a finite dimensional vector space and $d = \dim(V)$. Then
(i) any subset of V containing more than d vectors is linearly dependent.
(ii) A subset of V containing fewer than d vectors cannot span V.
(iii) A linearly independent set $S = \{v_1, \dots, v_d\}$ consisting of d many vectors is a basis.
(iv) A spanning set $S = \{v_1, \dots, v_d\}$ of V consisting of d many vectors is a basis.

Proof. The statements (i) and (ii) follow from Theorem 2.24.

(iii) If possible, suppose S is not a basis. Then S does not span V, i.e., $\mathrm{Span}(S) \neq V$. So there is $v \in V \setminus \mathrm{Span}(S)$. Hence, by Lemma 2.17, $S \cup \{v\}$ is linearly independent, but it has more than d vectors. This contradicts the statement (i).

(iv) If possible, suppose S is not a basis. Then S is linearly dependent. So, by Remark 2.16(2), without loss of generality, we may assume that

$$v_1 \in \mathrm{Span}\{v_2, v_3, \dots, v_d\}. \tag{2.3}$$

Hence, since $V = \mathrm{Span}(S)$, it follows from (2.3) that $V = \mathrm{Span}(\{v_2, v_3, \dots, v_d\})$. This contradicts the statement (ii). $\square$

Exercise 2.27. Find a basis of the space $\mathbb{R}^{m\times n}$ of all $m \times n$ matrices, and hence $\dim(\mathbb{R}^{m\times n})$. Hint: In view of Example 2.21, we have $\dim(\mathbb{R}^n) = n$.

Theorem 2.28. Let W be a subspace of a finite dimensional vector space V, and S be a linearly independent subset of W. Then S is finite, and it is part of a (finite) basis of W.

Proof. Note that S is also a linearly independent subset of V. So S contains at most $\dim(V)$ elements. If S spans W, then S itself is a basis of W, and we are done. If $\mathrm{Span}(S) \neq W$, then there is a vector $v_1 \in W \setminus \mathrm{Span}(S)$. Hence, by Lemma 2.17, $S_1 := S \cup \{v_1\}$ is linearly independent. If $\mathrm{Span}(S_1) = W$, then we are done. Otherwise, there is $v_2 \in W \setminus \mathrm{Span}(S_1)$, and hence by Lemma 2.17, $S_2 := S \cup \{v_1, v_2\}$ is linearly independent. This process stops after some finite steps because at most $\dim(V)$ linearly independent vectors can be there in W. So finally we obtain a set

$$S \cup \{v_1, v_2, \dots, v_m\} \subset W$$

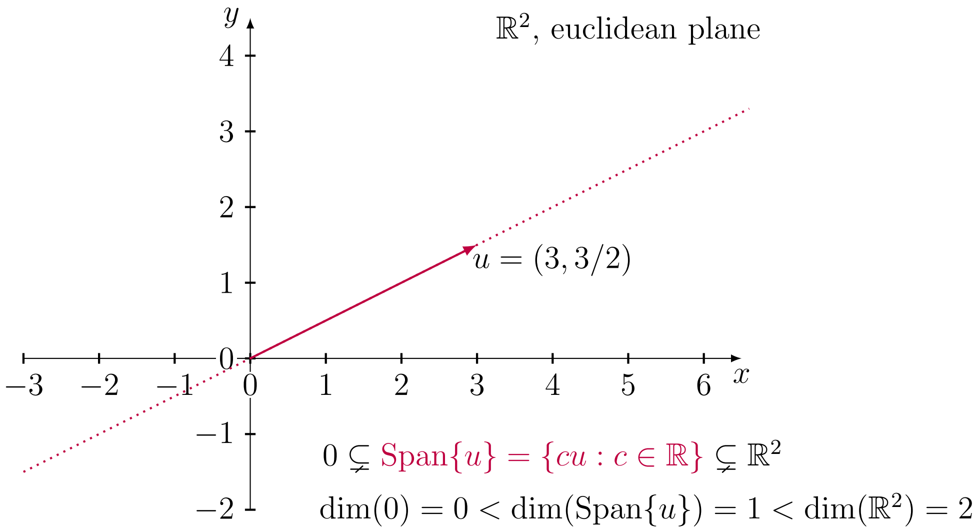
which spans W, i.e., it forms a basis of W. $\square$

As an immediate consequence, we obtain the following.

Corollary 2.29. In a finite dimensional vector space V, every linearly independent set of vectors is part of a basis.

Corollary 2.30. A subspace W of V is PROPER if and only if $\dim(W) < \dim(V)$.

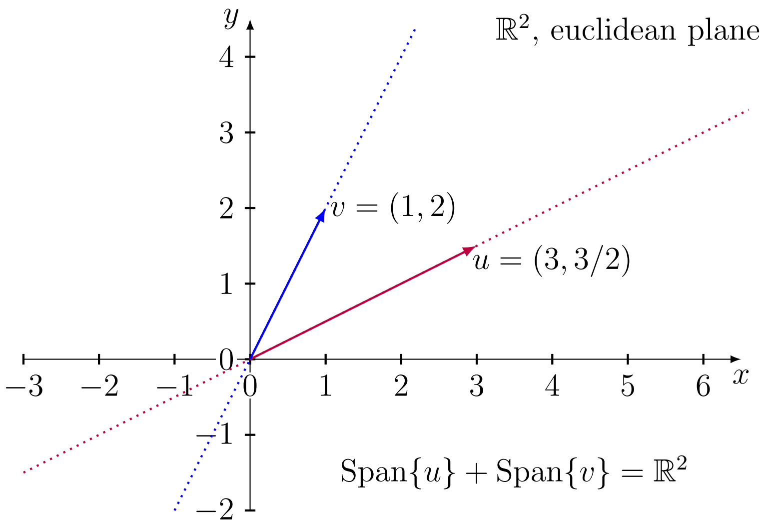
Proof. The 'if' part is trivial, because if $W = V$, then $\dim(W) = \dim(V)$. For the 'only if' part, let $W \subsetneq V$. It follows that $V \neq 0$. So, if $W = 0$, then we are done. Thus we may assume that there is a non-zero vector w in W. Then, by Theorem 2.28, $\{w\}$ can be extended to a finite basis (say S) of W. So, in particular, W is finite dimensional. Since $\mathrm{Span}(S) = W \subsetneq V$, there is a vector $v \in V \setminus \mathrm{Span}(S)$. By Lemma 2.17, $S \cup \{v\}$ is a linearly independent subset of V. Hence, again by Theorem 2.28, $S \cup \{v\}$ can be extended to a basis of V. Therefore $\dim(W) < \dim(V)$. $\square$



Definition 2.31. Let $W_1, W_2, \ldots, W_r$ be subspaces of a vector space V. Then the **sum of these subspaces** is defined to be

$$W_1 + W_2 + \cdots + W_r = \{w_1 + w_2 + \cdots + w_r : w_i \in W_i\}.$$

Remark 2.32. It can be verified that $W_1 + W_2 + \cdots + W_r$ is again a subspace of V. In fact, this is the subspace of V spanned by $W_1 \cup W_2 \cup \cdots \cup W_r$, i.e., the smallest subspace containing all $W_1, W_2, \ldots, W_r$.



Theorem 2.33. Let W_1 and W_2 be finite dimensional subspaces of a vector space V. Then $W_1 + W_2$ is finite dimensional, and

$$\dim(W_1) + \dim(W_2) = \dim(W_1 \cap W_2) + \dim(W_1 + W_2).$$

Proof. Since $W_1 \cap W_2 \subseteq W_1$, it follows that $W_1 \cap W_2$ has a finite basis $\{u_1, \ldots, u_r\}$, which can be extended to a basis

$$\{u_1, \ldots, u_r, v_1, \ldots, v_m\} \text{ of } W_1$$

and a basis

$$\{u_1, \ldots, u_r, w_1, \ldots, w_n\} \text{ of } W_2.$$

Show that $\{u_1, \ldots, u_r, v_1, \ldots, v_m, w_1, \ldots, w_n\}$ is a basis of $W_1 + W_2$. $\square$

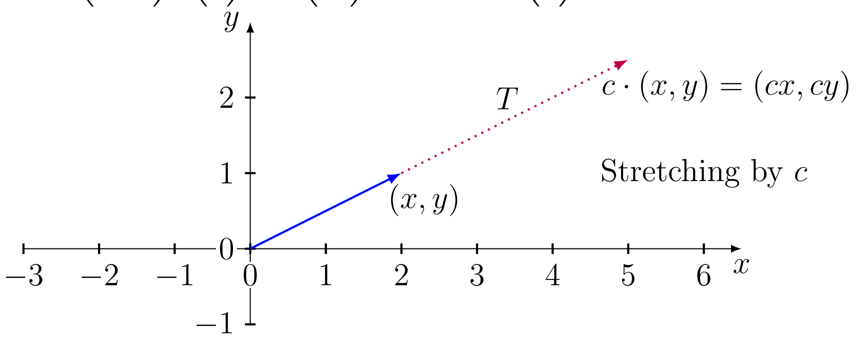
Chapter 3

Linear Transformations

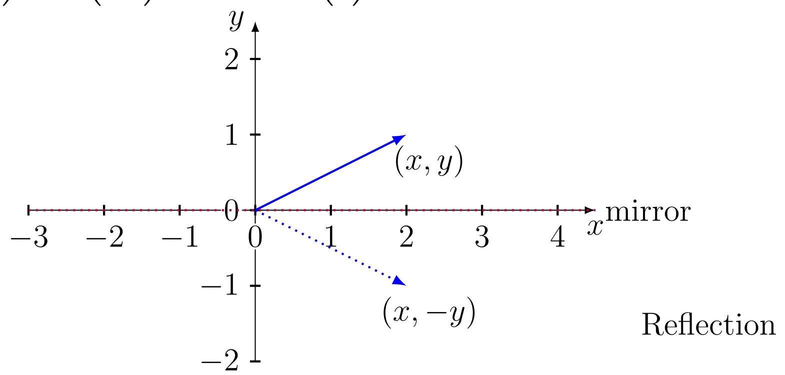
3.1 Some well known (linear) maps

A 'Linear Transformation' is nothing but a map between two vector spaces which preserves the vector space structure. Let us start with some well known maps.

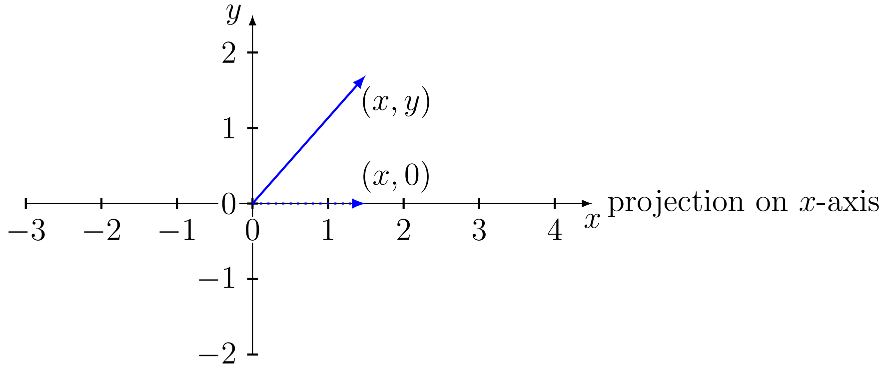
3.1 (Stretching by a scalar). Let $c \in \mathbb{R}$. We have a map $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined by $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto c \begin{pmatrix} x \\ y \end{pmatrix}$ for every $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. Note that this map can be represented by $\begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}$ because $\begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} cx \\ cy \end{pmatrix}$ for every $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$.



3.2 (Reflection with x-axis as mirror). Consider the map $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined by $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ -y \end{pmatrix}$ for every $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. This map can be represented by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.



3.3 (Projection on the x-axis). Consider the map $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ defined to be $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ 0 \end{pmatrix}$ for every $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. This map can be represented by $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.



3.4. In a linear map, domain and codomain need not be same. In view of the projection map $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ in 3.3, one can define another map $S : \mathbb{R}^2 \longrightarrow \mathbb{R}$ by $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto x$. This map can be represented by the 1×2 matrix $\begin{bmatrix} 1 & 0 \end{bmatrix}$ because $\begin{bmatrix} 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x$.

Keeping the above examples in mind, we now give the formal definition of linear transformation.

Definition 3.5. A transformation (or map) between vector spaces which satisfies the rule of linearity is called linear transformation (or linear map). More precisely, let V and W be vector spaces over $\mathbb{R}$. A linear transformation $T : V \to W$ is a function such that

$$T(c_1 v_1 + c_2 v_2) = c_1 T(v_1) + c_2 T(v_2)$$

for all $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2 \in V$.

Example 3.6. Let A be an $m \times n$ matrix over $\mathbb{R}$. Then the map $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(X) := AX$ for all $X \in \mathbb{R}^n$ is a linear transformation. Indeed, $T(X + Y) = A(X + Y)$ $= AX + AY = T(X) + T(Y)$ and $T(cX) = A(cX) = c(AX) = cT(X)$.

Bibliography

[1] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd Edition.

[2] https://www.wikipedia.org

[3] G. Strang, *Linear Algebra and Its Applications*, 4th Edition.