Lectures Math 308: Linear Algebra: System of Linear Equations, Matrix Algebra and Vector Algebra

Gerardo Zelaya Eufemia

May 9, 2019

Contents

1 Introduction

Students more often learn what they read and have struggled with, than what is only told to them. Thus, I ask you to memorize the vocabulary and the definitions, and I will focus in discussing the concepts. You will still need to interiorize these concepts, and for that you need to ask all kind of questions related to the material.

- "OBVIOUS is the most dangerous word in mathematics." E.T.Bell.
- "The greatest obstacle to discovery is not ignorance, it is the (false) illusion of knowledge." Stephen Hawking
- "At some point one does, or does not, make the transition to a deeper involvement where one internalizes and makes automatic the steps that are first learned by rote. At some point one does, or does not, integrate the separate discrete aspects of the activity into a whole.You will know whether you are undergoing this transition." Paul Smith - Math professor at UW - On a sermon about linear algebra.

Some notation that I will use while writing on the board is:

- The abreviation i.e. represent the latin id est, which means *that is*. Similarly, e.g. abreviates exempli gratia, which means *for example*.
- $\forall$ means for all, or for every.
- $\exists$ means it exists. Similarly $\exists!$ means it exists a unique element.
- $\in$ means in, or is an element in a set. An element is contained in a set.
- $\notin$ means not in, or is not an element in a set.
- $\subseteq$ means is a subset of, or is contained in a set. A set is contained in another set.
- $A \subseteq B$ is equivalent to write $B \supseteq A$.
- The domain of the function $f : A \to B$ is A and $\forall a \in A : \exists b \in B : f(a) = b$.
- The codomain of the function $f : A \to B$ is B.

- The range of the function $f : A \to B$ is given by $Range(f) = \{b \in B : \exists a \in A : f(a) = b\}$. Thus the range is always contained in the codomain, but they are in general not equal.
- The natural numbers are $\mathbb{N} = \{0, 1, 2, \ldots\}$. The integers are $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2 \ldots\}$. The rational numbers $\mathbb{Q}$ are all fractions or quotients of integers. The irrational numbers $\mathbb{Q}^c$ are all numbers one thinks of that are not rational, like π, e, etc. The real numbers $\mathbb{R}$ is the union of rational and irrational numbers. The complex numbers are $\mathbb{C} = \{a + bi/a, b \in \mathbb{R}; i = \sqrt{-1}\}$.

Finally, let me remark that these set of notes do not intent to replace the textbook, but rather to complement it. Thus, I ask you to read the text along these notes. Trying to read a later chapter in a textbook without understanding its notation and its style is really difficult. Hence, start reading the textbook since the beggining and along the lectures.

2 Linear System of Equations

An equation is a mathematical equality where one does not know the value of one or more variables. In high school, we study **polynomial equations in one variable**: linear equations of the standard form $ax + b = 0$, quadratic equations of the standard form $ax^2 + bx + c = 0$ or in their vertex and intercept form $a(x - v)^2 + k$. All of these could be generalized to polynomials of higher degrees: cubic equations which always have a real solution and often lead to complex solutions, quartic or biquadratic equations, etc. However, we are going to study **linear equations on several variables**, whose standard form is:

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

Example 1 $x_1 + 2x_2 - 3x_3 = 10$

In general, one calls the numbers $a_1, a_2, \ldots, a_n$ the coefficients. One calls $x_1, x_2, \ldots, x_n$ the variables. And one calls the number b to be the constant term.

Remark 1
1. *However, not all variables need to appear. Some of them may have coefficient* 0 *in a particular equation, and thus one does not write them.*
2. *We will only consider equations with real coefficients and real constant terms, and we will look only for real solutions. Although the algebra of complex linear equations is not much different than this, we will only cover it until the last chapter of the quarter when we will cover Eigenvalues and Eigenvectors.*
3. *Additionally, rather than using random names for the variables, we will try to always use the variable names* $x_1, x_2, \ldots$ *and we will rewrite all equations so the variables appear in order.*

In the study of the graph of an line in the plane of the form $y = ax + b$, one uses ordered pairs (x_0, y_0) to represent a point in the line. We will rename ordered pairs and call them 2-tuples. More general, an n-tuple is an ordered list $(s_1, s_2, \ldots, s_n)$ of n numbers.

A **particular solution** to linear equation of the form $a_1x_1 + \ldots + a_nx_n = b$ is a tuple of values $(s_1, s_2, \ldots, s_n)$ that satisfies the equation, i.e. if one plugs the values $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ into the equation, then one obtains an equality $a_1s_1 + a_2s_2 + \ldots + a_ns_n = b$.

Example 2 *The equation* $x_1 + 2x_2 - 3x_3 = 10$ *has a particular solution* $(0, 5, 0)$*. Another particular solution is* $(1, 6, 1)$.

However, one is interested not only in a particular solution, but in the set of all solutions. For a finite set, we can list the elements, but for an infinite set, one needs a convention on how to write it or how to define it. A set is written in the form $\{F : P\}$ or $\{F|P\}$ or even $\{F/P\}$ where F is the form of the elements of the set, and P is the list of properties that the elements satisfy.

Example 3 *For example,* $\{3x + 1 : x = 1, 2, 3\} = \{4, 7, 10\}$

Notice that in order to find a particular solution of a single equation, one can assign any values to all but one variable, and then solve for that last variable. Hence, the **solution set** of a single linear equation can be represented with the use of $n - 1$ parameters:

$$\left\{ \left(\frac{b - a_2s_1 - a_3s_2 - \ldots - a_{n-1}s_{n-1}}{a_1}, s_1, s_2, \ldots, s_{n-1} \right) : s_1, s_2, \ldots, s_{n-1} \in \mathbb{R} \right\}$$

Remark 2 *Notice how one requires to order the variables, and how one tends to assign parameters to all but the first variable. Alternative, one can just assume that $x_2, x_3, \ldots, x_n$ will be assigned any value, and then one will solve for the first variable only, thus the solution set is:*

$$\left\{\left(\frac{b - a_2x_2 \ldots - a_nx_n}{a_1}, x_2, x_3, \ldots, x_n\right) : x_1, x_2, \ldots, x_{n-1} \in \mathbb{R}\right\}$$

Example 4 *The solution set of the equation $x_1 + 2x_2 - 3x_3 = 10$ is*

$$\left\{\left(\frac{10 - 2x_2 + 3x_3}{1}, x_2, x_3\right) : x_2, x_3 \in \mathbb{R}\right\}$$

The form of the elements $\left(\frac{b-a_2s_1-a_3s_2-\ldots-a_{n-1}s_{n-1}}{a_1}, s_1, s_2, \ldots, s_{n-1}\right)$ in the solution set of a linear equation $a_1x_1 + \ldots a_nx_n = b$ is called the **general solution**.

However, one will be interested not only in a solution to an equation, but in a common solution to several linear equations. Hence, a system of linear equations is a list of equations. In general, it is represented as:

$$\begin{array}{ccccccccc}
a_{1,1}x_1 & + & a_{1,2}x_2 & + & \ldots & + & a_{1,n}x_n & = & b_1 \\
a_{2,1}x_1 & + & a_{2,2}x_2 & + & \ldots & + & a_{2,n}x_n & = & b_2 \\
\vdots & & \vdots & & \ddots & & \vdots & & \vdots \\
a_{m,1}x_1 & + & a_{m,2}x_2 & + & \ldots & + & a_{m,n}x_n & = & b_m
\end{array}$$

Or, for the students familiarized with the Sum-Sigma notation, it is abbreviated as $\forall 1 \leq i \leq n : \sum_{j=1}^{n} a_{ij}x_j = b_j$.

Remark 3 *Notice that one implies an order of the equations, and the coefficient of the variable x_j in the i-th equation is denoted by a_{ij}. Moreover, please do not attach i and j to a given role, as every book chooses one role and it rarely sticks to it all the time. For example, one could say that a_{ji} is the coefficient of x_i in the j-th equation, and this still agrees with what one said above.*

A **particular solution** of a linear system of equations is an n-tuple $(s_1, \ldots, s_n)$ of values that satisfies all of the equations in the system. In other words, it is a common solution. The **solution set** of a linear system of equations is the set of points $(s_1, \ldots, s_n)$ that satisfy all equations in the system. The form of the elements in the solution set of a linear system of equations is called the **general solution** of the system.

Theorem 1 *The solution set of a linear system of equations is either empty, a unique solution or infinitely many solutions. Hence, one indicates that the system is inconsistent when it does not have solutions, or consistent if it has solutions. If the system has only one solution, one writes the solution in curly brackets to denote the solution set. And if the system has infinitely many solutions, one determines how many variables are free to choose its value and thus these variables can be consider parameters, and one solves the rest of variables in terms of these parameters.*

First let's study some types of systems that are easy to solve: A **Diagonal System** has the same number of equations as that of variables. Each equation contains only one variable and each variable appears in exactly one equation. The order of the equations is choose to match that of the variables, and the system has exactly one solution. In an abstract manner, it is defined by saying that a_{ij} is zero whenever $i \neq j$ and non-zero whenever $i = j$.

$$\begin{array}{ccccccc}
a_{1,1}x_1 & & \ldots & & & = & b_1 \\
& a_{2,2}x_2 & \ldots & & & = & b_2 \\
& & \ddots & & & = & \vdots \\
& & & & a_{n,n} & = & b_n
\end{array}$$

Thus, the unique solution is $\left(\frac{b_1}{a_{11}}, \frac{b_2}{a_{22}}, \ldots, \frac{b_n}{a_{nn}}\right)$.

Example 5 *The linear system of equations*

$$\begin{array}{cccc}
2x_1 & & = & 7 \\
& 3x_2 & = & 6
\end{array}$$

is a diagonal system whose unique solution is $(7/2, 2)$. However, the linear system

$$\begin{array}{ccccc}
2x_1 & & & = & 7 \\
& 0x_2 & = & 0 & \\
& & x_3 & = & -2
\end{array}$$

will not be considered a diagonal system because although we have written x_2 in the second equation, it has coefficient 0.

A **Triangular systems** has the same number of equations as that of variables. The first variable to appear in the i-th equation is the i-th variable, giving it a stair-shaped presentation. It has a unique solution, which is obtained by the process of **backward substitution**, i.e. find the value of x_n from the last equation. Then, having found x_n, find the value of x_{n-1} from the $n-1$-st equation; and so on.

$$\begin{array}{ccccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & a_{1,3}x_3 & + & \dots & + & a_{1,n}x_n & = & b_1 \\ & & a_{2,2}x_2 & + & a_{2,3}x_3 & + & \dots & + & a_{2,n}x_n & = & b_2 \\ & & & & a_{3,3}x_3 & + & \dots & + & a_{3,n}x_n & = & b_2 \\ & & & & & & \ddots & & \vdots & & \vdots \\ & & & & & & & & a_{m,n}x_n & = & b_m \end{array}$$

Example 6 *The linear system of equations*

$$\begin{array}{ccccc} 2x_1 & + & 3x_2 & = & 7 \\ & & 5x_2 & = & 6 \end{array}$$

is a triangular system. One first solves the last equation $x_2 = 6/5$. Then one plugs in the previos equation to obtain $2x_1 + 3(6/5) = 7$ and solves for the previous variable $x_1 = 17/5$. Hence the unique solution is $(17/5, 6/5)$.

Remark 4 *Diagonal systems and triangular systems are the simplest types of systems. However, not all systems have the same number of equations as that of variables. And one needs a similar type of system that resembles triangular systems, which will allows one to identify which variables are free to choose their values (and thus can be replaced by parameters), and which variables must be solved in terms of these parameters.*

Remark 5 *Please keep in mind that the terms diagonal and triangular will have a slightly different meaning for matrices.*

An **Echelon Form System**'s number of equations is at most its number of variables. Each variable is the leading variable of at most one equation, and it has the same stair shape as a triangular system, i.e.: the leading variable of an equation is to the right of the leading variable of the previous equation. The textbook allows an echelon system to have zero equations, which are equality statements of the form $0 = 0$, irrelevant to the solution of the system. In this class, one will not allow the system to have such equalities, one will say that one has an echelon system and some equality statements. The list of variables which are the first to appear in at least one equation of an echelon system are the **leading variables**. The rest of variables are called **free variables.**

Remark 6 *One can always speak of the leading variable of a single equation. However, one reserves the labels of* ***leading variables*** *and* ***free variables*** *to the variables in an* ***echelon form system.***

Example 7 *The following system is an echelon form system*

$$\begin{array}{ccccccccc} 2x_1 & + & 3x_2 & & & - & 7x_4 & = & 7 \\ & & 5x_2 & - & x_3 & + & 2x_4 & = & 10 \\ & & & & & & x_4 & = & 1 \end{array}$$

whose leading variables are x_1, x_2, x_3, and its only free variable is x_3 because it is not leading any equation.

2.1 Equivalent System, Jordan-Gauss Elimination, and Row operations

The purpose of this chapter is to find a method on how to solve any system of linear equations, i.e. to determine whether the system is inconsistent, or if it is consistent with one solution or with infinitely many solutions. However, the form of a general system is not always as nice as that of a diagonal system, a triangular system or an echelon system. So one needs to reduce the system to an equivalent echelon system, plus maybe some equality statements and from them determine the consistency of the system.

First, one defines three elementary row operations allowed to perform in the process of reducing the system.

1. **Interchange** two equations: Take two equations, say these of index i and j

$$\left\{\begin{array}{ll} \ldots & \\ i\textbf{-th Equation} & a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i \\ \ldots & \\ j\textbf{-th Equation} & a_{j1}x_1 + a_{j2}x_2 + \ldots + a_{jn}x_n = b_j \\ \ldots & \end{array}\right\}$$

   And change them, so the new i-the equation is the old j-th equation and viceversa. No other equation is changed in any manner.

$$\left\{\begin{array}{ll} \ldots & \\ i\textbf{-th Equation} & a_{j1}x_1 + a_{j2}x_2 + \ldots + a_{jn}x_n = b_j \\ \ldots & \\ j\textbf{-th Equation} & a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i \\ \ldots & \end{array}\right\}$$

2. **Multiply** an equation by a nonzero constant. Take one equation, say this of index i, and multiply it by a constant $cneq0$, which in this class must be real.

$$\left\{\begin{array}{ll} \ldots & \\ i\textbf{-th Equation} & a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = b_i \\ \ldots & \end{array}\right\}$$

   Notice that by the distribution of multiplication over sum, all coefficients and the constant term are multiplied by the constant c

$$\left\{\begin{array}{ll} \ldots & \\ i\textbf{-th Equation} & ca_{i1}x_1 + ca_{i2}x_2 + \ldots + ca_{in}x_n = cb_i \\ \ldots & \end{array}\right\}$$

3. **Add a multiple** of one equation to another equation. Take two equations, say these of index i and j. Multiply the i-th equation by $cneq0$, and add the result to the j-th equation

$$\left\{\begin{array}{lccccccccc} \ldots & & & & & & & & & \\ i\textbf{-th Equation} & a_{i1}x_1 & + & a_{i2}x_2 & + & \ldots & + & a_{in}x_n & = & b_i \\ \ldots & & & & & & & & & \\ j\textbf{-th Equation} & a_{j1}x_1 & + & a_{j2}x_2 & + & \ldots & + & a_{jn}x_n & = & b_j \\ \ldots & & & & & & & & & \end{array}\right\}$$

   Notice that the coefficients of the new j-th equation are of the form $ca_{ik} + a_{jk}$ and the constant term is of the form $cb_i + b_j$. Moreover, notice that neither i-th equation nor any other equation is modified.

$$\left\{\begin{array}{lccccccc} \ldots & & & & & & & \\ i\textbf{-th Equation} & a_{i1}x_1 & + & \ldots & + & a_{in}x_n & = & b_i \\ \ldots & & & & & & & \\ j\textbf{-th Equation} & (a_{j1} + ca_{i1})x_1 & + & \ldots & + & (a_{jn} + ca_{in})x_n & = & (b_j + cb_i) \\ \ldots & & & & & & & \end{array}\right\}$$

Two systems are called **equivalent systems** is one can be transformed to the other by a finite sequence of row elementary operations. The goal is to transform any system of equations into an Echelon form system, called its **Echelon Form System**, plus maybe some equality statements that must be written at the bottom.

Example 8 *Consider the system of equations*

$$\begin{array}{rcrcrcr} x_1 & - & 3x_2 & + & 2x_3 & = & -1 \\ 2x_1 & - & 5x_2 & - & x_3 & = & 2 \\ -4x_1 & + & 13x_2 & - & 12x_3 & = & 11 \end{array}$$

. After performing the operations (1) Subtract twice the first row to the second one, (2) Add four times the first row to the third one, (3) Subtract the second row from the third one, one obtains the Echelon form equivalent

$$\begin{array}{rcrcrcr} x_1 & - & 3x_2 & + & 2x_3 & = & -1 \\ & & x_2 & - & 5x_3 & = & 4 \\ & & & & x_3 & = & 3 \end{array}$$

Notice that there are no equality statements, and the resulting system is a triangular system, so it has a unique solution: $(50, 19, 3)$.

The **Pivot Positions** are ordered pairs (i, j) indicating that the location of the leading variables. Saying that (i, j) is a pivot position means that x_j is the leading variable of the i-th equation. Moreover, the pivot columns are the indices of the leading variables. Hence, saying that (i, j) is a pivot position means that j is a pivot column.

Example 9 *In the system*

$$\begin{array}{ccccccc} x_1 & - & 3x_2 & + & 2x_3 & = & -1 \\ & & x_2 & - & 5x_3 & = & 4 \\ & & & & x_3 & = & 3 \end{array}$$

the pivot positions are $(1, 1), (2, 2), (3, 4)$. *Hence, the pivot columns are* $1, 2, 4$.

Remark 7 *Pivot positions are also called leading positions, and pivot columns correspond to the indices of the leading variables. The word "column" comes from the writing of the system, where one preserves the terms with the same variable in a single column.*

Remark 8 *In order to solve any system of equations, one needs to find a systematic general method of performing row elementary operations that guarantees to find its equivalent echelon form. This method is called Gauss Elimination.*

The process of **Gauss Elimination** in order to obtain the equivalent Echelon system to any given system is:

- Exchange the first row for another row, if necessary and possible, so the leading variable in the first equation is x_1. Then add multiples of the first equation to all other equations to cancel their x_1 terms. Hence, the only x_1 term will be in the first equation.
- Ignore the first equation. Exchange the second row for another row, if necessary and possible, so the leading variable in the second equation is x_2. And then add multiples of the second equation to all other equations below to cancel their x_2 terms.
- Continue in this manner until all equations are exhausted. It may be possible to obtain equality statements in the bottom, and these may be true and of the form $0 = 0$, or they may be false and of the form $0 = b$ where b is a nonzero constant.

Remark 9 *Notice that Gauss Elimination does not tell us how to choose what equations one must exchange. These choices will influence the answer. Hence,* ***the echelon form of any given system is not unique.*** *Moreover, many of the examples in this class will start with integers coefficients. This does not imply that the solution will only contain integers, in general the answer will contain fractions or rational numbers. In order to avoid bigger denominators, in the steps of exchanging, one aims to choose the equation whose leading coefficient has the smallest absolute value.*

Example 10 *Consider the system*

$$\begin{array}{rrrrrrl} & & 6x_3 & & +19x_5 & 11x_6 & = -27 \\ 3x_1 & +12x_2 & +9x_3 & -6x_4 & +26x_5 & +31x_6 & = -63 \\ x_1 & +4x_2 & +3x_3 & -2x_4 & +10x_5 & +9x_6 & = -17 \\ -x_1 & -4x_2 & -4x_3 & +2x_4 & -13x_5 & -11x_6 & = 22 \end{array}$$

Perform the following row elementary operations (1) Exchange equations 1 and 3; (2) Subtract equation 1 to equation 2; (3) Add equation 1 to equation 4; (4) Exchange equations 2 and 4; (5) Add 6 times equation 2 to equation 3; (5) Add 4 times equation 3 to equation 4. Then one obtains its Echelon form system

$$\begin{array}{rrrrrrl} x_1 & +4x_2 & +3x_3 & -2x_4 & +10x_5 & +9x_6 & = -17 \\ & & -x_3 & & -3x_5 & -2x_6 & = 5 \\ & & & & x_5 & -x_6 & = 3 \\ & & & & & 0 & = 0 \end{array}$$

. *The leading variables are* x_1, x_3, x_5. *The free variables are* x_2, x_4, x_6. *The pivot positions are* $(1, 1), (2, 3), (3, 5)$. *The pivot columns are* $1, 3, 5$. *Its echelon form system has a true equality statement* $0 = 0$.

One can further simplify a system of equations by finding its reduced echelon form by performing **Jordan-Gauss elimination**. Gauss elimination is called the **forward phase** of this process, which transforms a given system to its echelon form. The new part is called the **backward phase**, which transforms the echelon system to its reduced echelon form.

A **reduced echelon system** is a linear system of equations in echelon form with the additional conditions that all the leading coefficients are 1, and each leading variable only appears in exactly one equation, that which it leads. Thus the coefficients in any pivot column are all 0 but one, which is 1.

The backward phase of the **Jordan Gauss elimination** consists of:

- Multiply each equation by the reciprocal of its leading coefficient.
- If the leading variable of the last equation is x_i, then add multiples of the last equation to all other equations above in order to cancel their x_i terms.
- Move to the next equation above. If this equation has leading variable x_i, then add multiples of this equation to all other equations above in order to cancel their x_i terms.
- Continue in this manner, until one reaches the second equation of the system.

Example 11 *Continuing with the previous example, where the system*

$$\begin{array}{rrrrrrl} & & 6x_3 & & +19x_5 & 11x_6 & = -27 \\ 3x_1 & +12x_2 & +9x_3 & -6x_4 & +26x_5 & +31x_6 & = -63 \\ x_1 & +4x_2 & +3x_3 & -2x_4 & +10x_5 & +9x_6 & = -17 \\ -x_1 & -4x_2 & -4x_3 & +2x_4 & -13x_5 & -11x_6 & = 22 \end{array}$$

has echelon form

$$\begin{array}{rrrrrrl} x_1 & +4x_2 & +3x_3 & -2x_4 & +10x_5 & +9x_6 & = -17 \\ & & -x_3 & & -3x_5 & -2x_6 & = 5 \\ & & & & x_5 & -x_6 & = 3 \\ & & & & & 0 & = 0 \end{array}$$

One performs: Multiply equation 2 by -1; add -3 times equation 2 to equation 1; add -3 times equation 3 to equation 2; and add -1 times equation 3 to equation 1. Then one obtains its reduced echelon form

$$\begin{array}{rrrrrrl} x_1 & +4x_2 & & -2x_4 & & +4x_6 & = -5 \\ & & x_3 & & & 5x_6 & = -14 \\ & & & & x_5 & -x_6 & = 3 \\ & & & & & 0 & = 0 \end{array}$$

Notice that the leading variables, free variables, pivot positions and pivot columns are the same as these in the echelon form. The solution set is

$$\{(-4s_2 + 2s_4 - 4s_6 - 5, s_2, -5s_6 - 14, s_4, s_6 + 3, s_6) : s_6, s_4, s_2 \in \mathbb{R}\}$$

.

Another important type of systems are homogeneous systems. A **Homogeneous** linear system of equations is that which has all constant terms equal to 0

$$\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \ldots & + & a_{1,n}x_n & = & 0 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \ldots & + & a_{2,n}x_n & = & 0 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \ldots & + & a_{m,n}x_n & = & 0 \end{array}$$

It is always consistent because of the **trivial solution** $(0, 0, \ldots, 0)$, which clearly satisfies it. So one is usually interested in finding other possible solutions.

Theorem 2 *(Same as theorem 1, just restated) A linear system of equations has either no solutions (inconsistent), or a unique solution (consistent with 1 solution), or infinitely many solutions (consistent with ∞-ly many solutions)*

Proof: A system of linear equations can be transformed to an equivalent echelon form system through the process of Gauss-elimination. After the reduction, if there are false equality statements (of the form $0 = b$ for a nonzero constant b), then the system is inconsistent. Otherwise, all equality statements are true (of the form $0 = 0$) and the system is consistent. If the resulting echelon form system is triangular, then the system has a unique solution. Otherwise, the number of equations is less than the number of variables, and there are free variables, which can be replaced by parameters, and thus the system has ∞-ly many solutions.

2.2 Geometric Interpretation

Before one asks for the solution of an equation, one needs to understand where these solutions are. Recall that an equation is of the form $a_1x_1 + \ldots + a_nx_n = b$, and that a particular solution is an n-tuple $(s_1, \ldots, s_n)$ of real numbers. The solution set of a single equation is called a **Hyperplane.**

In trying to develop the concept of a hyperplane, consider first a linear equation in one variable $ax + b = 0$. The solution $x = b/a$ is a point found in the real line. Now, consider a linear equation in two variables $ax + by = c$. The solution is a line in the real plane. Somehow, the solution to an equation in n-variables lives in an n-dimensional space, which one will call the euclidean space, and a hyperplane is an $n-1$ dimensional linear object living in that n-dimensional space.

Moreover, the solution of a system of linear equations is the common solutions to all equations, i.e. the intersection of the hyperplanes determined by each single equation of the system.

Example 12 *Consider the system*

$$\begin{array}{rcrcr} x_1 & + & 2x_2 & = & 7 \\ x_1 & - & x_2 & = & -2 \\ 2x_1 & + & x_2 & = & 5 \end{array}$$

Geometrically, one is studying three lines or 1-hyperplanes in the plane or 2-space, each one determined by each equation. The easiest lines to study are these of the forms $x_1 = a$ or vertical lines, and $x_2 = b$ or horizontal lines. If you don't understand this, change x_1 by x and x_2 by y and draw lines of the form $x = a$ and $y = b$. The echelon form system is

$$\begin{array}{rcrcr} x_1 & + & 2x_2 & = & 7 \\ & & x_2 & = & 3 \\ & & 0 & = & 0 \end{array}$$

Thus Gauss elimination is an attempt to change some of these lines by easier lines while preserving the solutions. Notice that one did not need three lines to determine the solution set, but only two lines. This is because the echelon system has two equations and a true equality statement of the form $0 = 0$. The reduced echelon system is

$$\begin{array}{rcrcr} x_1 & & & = & 1 \\ & & x_2 & = & 3 \\ & & 0 & = & 0 \end{array}$$

, which are a horizontal and a vertical line.

2.3 From linear systems to matrices

First of all, let's define what a matrix is. A **matrix** is a rectangular array of numbers, which one calls **entries**, i.e. a list of m lists of n numbers each of them.

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

Remark 10 *The use of square brackets [] or parenthesis () is optional, we will treat them indistinguishable here in this class.*

Remark 11 *For the moment, we don't need to understand anything else about matrices. Later we will give them a better structure to the set of matrices*

Recall that one tries to write the equations of a system in a column, in an orderly manner so the terms with the same variable all lie in the same column. At that moment, one may be interested only in the coefficients of the variables, and one know to which variable they belong just by looking at their relative position.

Given a linear system of equations

$$\begin{array}{ccccccc} a_{1,1}x_1 & + & \ldots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & \ldots & + & a_{2,n}x_n & = & b_2 \\ \vdots & & \ddots & & \vdots & & \vdots \\ a_{m,1}x_1 & + & \ldots & + & a_{m,n}x_n & = & b_m \end{array}$$

One associates two matrices to this system. The first one is called the **coefficient matrix** of the system

$$\begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ a_{2,1} & \dots & a_{2,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}$$

which arranges only the coefficients of the variables ignoring the constant terms. The other matrix is called the **augmented matrix**

$$\left(\begin{array}{ccc|c} a_{1,1} & \dots & a_{1,n} & b_1 \\ a_{2,1} & \dots & a_{2,n} & b_2 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & \dots & a_{m,n} & b_n \end{array}\right)$$

One should write the vertical line to indicate that this matrix comes from a system in standard form, and thus there is an equal sign there.

Remark 12 *At this moment, a matrix is used only to simplify notation. Rather than writing a lengthy linear system, one doesn't write the variable names, which for us will always be standard names $x_1, \dots, x_n$, and one uses the augmented matrix of the system. One performs the same row elementary operations, Jordan-Gauss elimination, and from them one deduces the solution to the system.*

2.4 Key differences between linear system of equations and matrices

The first major differences are the described types. But let's first introduce a new word in our vocabulary. The dimensions of a matrix is the number of rows times the number of columns. For example

$$A = A_{m\times n} = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ a_{2,1} & \dots & a_{2,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}_{m\times n}$$

The matrix above has dimensions $m \times n$. We can call the first dimension to the number of rows, and the second dimension to the number of columns. Sometimes one calls the pair just the dimension (singular word) of the matrix. It is denoted by writing the dimension as a sub-index on the name of the matrix $A_{m\times n}$ or as a sub-index in the representation of the matrix. It is not mandatory to write it, but it is a good exercise to always try to identify these dimensions.

A **diagonal matrix** is of the form

$$\begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

In other words, all entries outside the main diagonal are zero. However, the elements in the diagonal may or may not be zero. Thus, while the coefficient matrix of a diagonal linear system is a diagonal matrix, it is a diagonal matrix with non-zero entries in the diagonal.

Also, a triangular matrix can be of two types, both of which are square matrices, and thus one writes them as A_n rather than $A_{ntimesn}$: an **upper triangular matrix** is of the form

$$A_n = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}_{n\times n}$$

In other words, all entries below the main diagonal are zero, i.e. $a_{ij} = 0$ whenever $i > j$. A **lower triangular matrix** is of the form

$$A_n = \begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ a_{31} & a_{32} & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}_{n\times n}$$

In other words, all entries above the main diagonal are zero, i.e.: $a_{ij} = 0$ whenever $i < j$. Yet in both cases, there may be zeros in the diagonal. Hence, we say that a triangular system of equations is one whose coefficient matrix is upper triangular and there are no zero coefficients in the main diagonal.

Another relevant difference occurs while performing Gauss elimination. After performing Gauss elimination on a system, one obtains its echelon form equivalent and some equality statements. Here, we call the echelon form equivalent to the equations only, and we say that it results in an echelon form system PLUS some equality statements. However, in matrices, the echelon form matrix is the whole matrix with the same dimensions than the original matrix including the zero rows.

2.5 Vocabulary

Linear equation, linear system of equations, particular solution, solution set, general solution, diagonal system of equations, triangular system of equations, homogeneous system of equations, echelon form system of equations, reduced echelon form system of equations, coefficient matrix of the system, augmented matrix of the system, dimensions of the matrix, diagonal matrix, upper triangular matrix, lower triangular matrix.

3 Vectors and the Euclidean Space

Recall that we write solutions of a linear equation with n-variables as n-tuples $(s_1, s_2, \dots, s_n)$ of real numbers and parameters, which were motivated in the ordered pairs (x_0, y_0) studied in high-school. Such n-tuples are going to be called **row vectors**. Notice that in matrix notation, they are matrices of dimensions $1 \times n$. However, in order to enjoy the richness of matrix algebra, one needs to define vertical n-tuples or $ntimes1$ matrices refered as **column vectors** or simply **vectors**. For example, a general vector will be written:

$$\vec{v} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots a_{n1} \end{pmatrix}_{n\times 1}$$

Remark 13 *One could omit the double index in the entries of the vector, but again in order to enjoy the richness of matrix algebra, one must remember that all entries belong to a matrix of a single column, thus one may choose to write double index with a* 1 *as the second index for all entries.*

If one fixes the dimension n or the number of entries of the vectors, then the set of all vectors of dimension n is called the n-dimensional (real) **Euclidean Space**, also denoted $\mathbb{R}^n$.

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : x_1, x_2, \dots, x_n \in \mathbb{R} \right\}$$

Remark 14 *Notice that we will be using column vectors and not row vectors, which will make more sense when we study matrix algebra and in particular matrix multiplication. The difference of these choices is because there is no commutativity in matrix multiplication. However, the choice was already made when one did row elementary operations and not column elementary operations.*

Remark 15 *Within the Euclidean space, one can study the points with integral entries or $\mathbb{Z}^n$, which are all vectors whose entries are all in the integers. This is called a lattice, but we will not study them in this class. Similarly, $\mathbb{C}^n$ is the set of n-dimensional vectors whose entries are complex numbers. But again, we will study these complex Euclidean spaces until much more later.*

Remark 16 *The **entries** of a vector are also called **components** or **coordinates** of the vector. Rather than writing a column vector with entries $a_1, \ldots, a_n$, I will type $(a_1, \ldots, a_n)^T$, which will mean the column vector with these entries.*

However, a set by itself is usually irrelevant. In Mathematics, one needs to define a structure in a set in order to exploit all its characteristics. Hence, one is interested in equality of vectors, addition (and subtractions) of vectors, and scalar multiplication of vectors.

1. **Equality of Vectors**: Two vectors are equal $(x_1, \ldots, x_n)^T = (y_1, \ldots, y_n)^T$ if and only if its corresponding entries are equal, i.e. $x_1 = y_1, \ldots, x_n = y_n$. For example $(1, 2, 3)^T \neq (1, 2, -3)^T$.

   Remark 17 *Equality is reflexive and symmetric, i.e. a vector is equal to itself, and if a vector is equal to another vector, then this other vector is equal to the former one. Two vectors of different dimensions cannot be equal to each other*

2. **Addition of Vectors**: The sum or difference of two vectors of the same dimension $(x_1, \ldots, x_n)^T \pm (y_1, \ldots, y_n)^T$ is the vector of the same common dimension whose entries are the sum or difference of the corresponding entries, i.e. $(x_1 \pm y_1, \ldots, x_n \pm y_n)^T$. For example $(1, 2, 3)^T + (-2, 4, -3)^T = (-1, 6, 0)^T$.

   Remark 18 *The addition or subtraction of two vectors of different dimensions is undefined. Addition of vectors is commutative, i.e. the order of the summands is irrelevant.*

3. **Scalar Multiplication of Vectors**: One can multiply a vector $(x_1, \ldots, x_n)^T$ by a scalar or (real) number c. One usually writes the scalar first, but this product is also commutative. Then one distributes the scalar on all entries of the vector $c \cdot (x_1, \ldots, x_n)^T = (cx_1, \ldots, cx_n)^T$. For example $\pi \cdot (1, 2, 3)^T = (\pi, 2\pi, 3\pi)^T$.

Hence, one understands that the n-dimensional Euclidean space is not just the set of all n-dimensional vectors, but this set together with the operations defined above, and its properties. A combination of these properties gives us:

$$c(x_1, \ldots, x_n)^T + d(y_1, \ldots, y_n)^T = (cx_1 + dy_1, \ldots, cx_n + dy_n)^T$$

Proposition 1 *Some algebraic properties for scalars $a, b \in \mathbb{R}$, and vectors $\vec{X} = (x_1, \ldots, x_n)^T, \vec{Y} = (y_1, \ldots, y_n)^T, \vec{Z} = (z_1, \ldots, z_n)^T \in \mathbb{R}^n$ are:*

- *Commutativity of sum: $\vec{X} + \vec{Y} = \vec{Y} + \vec{X}$.*
- *Distributivity of scalar multiplication over sum: $a(\vec{X} + \vec{Y}) = a\vec{X} + a\vec{Y}$ and $(a + b)\vec{X} = a\vec{X} + b\vec{X}$.*
- *Associativity of sum: $(\vec{X} + \vec{Y}) + \vec{Z} = \vec{X} + (\vec{Y} + \vec{Z})$.*
- *Associativity of scalar multiplication: $a(b\vec{X}) = (ab)\vec{X}$.*
- *The identity of sum is the zero vector $\vec{0}$ whose entries are all zeros, thus $\vec{X} + \vec{0} = \vec{0} + \vec{X} = \vec{X}$. Sometimes I will write 0 in place of $\vec{0}$, thus 0 is the zero number, and the corresponding zero vector in the Euclidean Space in discussion.*
- *The inverse of a vector is -1 times the vector, thus $\vec{X} + (-\vec{X}) = \vec{0}$.*
- *The identity scalar for scalar multiplication is 1, thus $1 \cdot \vec{X} = \vec{X}$.*

A **Linear Combination** of vectors $\vec{v_1}, \ldots, \vec{v_n}$ is the resulting vector of the form $a_1\vec{v_1} + \ldots + a_n\vec{v_n}$ where $a_1, \ldots, a_n \in \mathbb{R}$. For example a linear combination of the vector $\vec{x} = (x_1, x_2)^T$ and the vector $\vec{y} = (y_1, y_2)^T$ is $a\vec{x} + b\vec{y} = (ax_1 + by_1, ax_2 + by_2)^T$. For example, one may understand a linear combination as a linear expression where the vectors play the role of variables.

Remark 19 *One only talks about linear combination of vectors of the same dimension.*

A different interpretation of a linear system of equations is to write it as a vector equation. Given the system of equations:

$$\left(\begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \ldots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \ldots & + & a_{2,n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \ldots & + & a_{m,n}x_n & = & b_m \end{array}\right)$$

The **Associated Vector Equation** is:

$$x_1 \underbrace{\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}}_{\vec{a}_1} + \ldots + x_n \underbrace{\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}}_{\vec{a}_n} = \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{\vec{b}}$$

Remark 20 *Thus, the system of linear equations is consistent if and only if $\vec{b}$ is a linear combination of the coefficient vectors $\vec{a}_1, \ldots, \vec{a}_n$ of the variables $x_1, \ldots, x_n$ respectively.*

Example 13 *Consider the linear system of equations:*

$$\begin{array}{rrrcr} 2x_1 & -3x_2 & +10x_3 & = & -2 \\ x_1 & -2x_2 & +3x_3 & = & -2 \\ 2x_1 & -x_2 & +18x_3 & = & 2 \end{array}$$

which is not homogeneous. The associated vector equation is:

$$x_1 \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} -3 \\ -2 \\ -1 \end{pmatrix} + x_3 \begin{pmatrix} 10 \\ 3 \\ 18 \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \\ 2 \end{pmatrix}$$

And in matrix form, one could write this as:

$$\underbrace{\begin{bmatrix} 2 & -3 & 10 \\ 1 & -2 & 3 \\ 2 & -1 & 18 \end{bmatrix}}_{\text{coefficient matrix}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} -2 \\ -2 \\ 2 \end{bmatrix}}_{\text{constant vector}} \quad \leftrightarrow \quad \underbrace{\left(\begin{array}{rrr|r} 2 & -3 & 10 & -2 \\ 1 & -2 & 3 & -2 \\ 2 & -1 & 18 & 2 \end{array}\right)}_{\text{augmented matrix}}$$

whose non-unique equivalent echelon form is:

$$\begin{array}{r} \textit{Echelon form system of equations} \\ \\ \textit{Equalities} \end{array} \left\{ \begin{array}{r} x_1 - 2x_2 + 3x_3 = -2 \\ 2x_2 + 8x_3 = 4 \end{array} \right. \\ \qquad 0 = 0 \quad \Leftrightarrow \quad \left(\begin{array}{rrr|r} 1 & -2 & 3 & -2 \\ 0 & 2 & 8 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and whose unique equivalent reduced echelon form is:

$$\begin{array}{r} \textit{Echelon form system of equations} \\ \\ \textit{Equalities} \end{array} \left\{ \begin{array}{r} x_1 - 2x_2 + 3x_3 = -2 \\ x_2 + 4x_3 = 2 \end{array} \right. \\ \qquad 0 = 0 \quad \Leftrightarrow \quad \left(\begin{array}{rrr|r} 1 & 0 & 11 & 2 \\ 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The solution set is:

$$\left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 - 11x_3 \\ 2 - 4x_3 \\ x_3 \end{bmatrix} : x_3 \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 - 11s \\ 2 - 4s \\ s \end{bmatrix} : s \in \mathbb{R} \right\}$$

The general solution in vector form is:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -11 \\ -4 \\ 1 \end{bmatrix}$$

Remark 21 *By knowing the general solution and the list of free variables, one can reconstruct the reduced echelon system.*

Example 14 *Indeed, in the previous example, the general solution is $\vec{x} = \begin{bmatrix} 2 - 11s \\ 2 - 4s \\ s \end{bmatrix}$ and x_3 is the only free variable. Notice that this is re-affirmed by the fact that there is only one parameter, and the coefficient of the parameter in the* *third row is* 1*. Hence, replace* $\vec{x}$ *by* $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ *and* s *by* x_3 *to obtain*

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2-11x_3 \\ 2-4x_3 \\ x_3 \end{bmatrix}$$

Then using vector equality and rewriting in standard form, one obtains: $x_1 = 2-11x_3 \Rightarrow x_1+11x_3 = 2$, $x_2 = 2-4x_3 \Rightarrow x_2+4x_3 = 2$*, and* $x_3 = x_3$ *which is a tautology and it can be replaced by* $0=0$*. However, many systems shared the same reduced echelon form, and thus one cannot recover the original system of linear equations.*

Example 15 *Consider the vectors* $\vec{e}_1 = (1,0,0)^T$ *and* $\vec{e}_2 = (0,1,0)^T$ *in* $\mathbb{R}^3$*. Consider all linear transformations of these two vectors, which in general they are of the form* $a\vec{e}_1 + b\vec{e}_2 = (a,b,0)^T$*. Notice that* $\vec{e}_3 = (0,0,1)^T$ ***is not a linear combination of*** $\vec{e}_1$ ***and*** $\vec{e}_2$ *because regardless of what scalars* a *and* b *one chooses, one cannot achieve* $\vec{e}_3$*. In other words, the vector equation* $x_1\vec{e}_1 + x_2\vec{e}_2 = \vec{e}_3$ *or the equivalent system of equations with augmented matrix* $\left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$ *has no solution.*

3.1 The Span of a list of vectors.

The **Span** (over $\mathbb{R}$) of a list of vectors $\vec{u}_1, \ldots, \vec{u}_m$ in $\mathbb{R}^n$ is the set of all of their linear combinations, which are of the form $a_1\vec{u}_1 + \ldots + a_m\vec{u}_m$ where $a_1, \ldots, a_m \in \mathbb{R}$. The notation is:

$$Span(\vec{u}_1, \ldots, \vec{u}_m) = Span_{\mathbb{R}}(\vec{u}_1, \ldots, \vec{u}_m) = \{a_1\vec{u}_1 + \ldots + a_m\vec{u}_m / a_1, \ldots, a_m \in \mathbb{R}\}$$

Remark 22 *Geometrically, if one considers each vector* $\vec{u}_i$ *as a direction, then the span of the* $\vec{u}_i$*'s is all points in the euclidean space that one can reach by moving only in these directions.*

Remark 23 *In general the span must be defined over a field (a number set where multiplicative inverses exist) in order to be able to apply Jordan-Gauss elimination and obtain all leading coefficients to be* 1*. However, in this class we will always talk about the span over the real numbers unless otherwise stated.*

Remark 24 *Thus a system of linear equations is consistent if and only if the constant vector is in the span of the coefficient vectors of the variables.*

Example 16 *Suppose an ant is currently located at* $(0,0,0)^T \in \mathbb{R}^3$ *and can only do steps in the direction of the vectors* $\vec{u}_1 = (2,1,1)^T$ *and* $\vec{u}_2 = (1,2,3)^T$ *or their negatives. Can the ant reach the point* $(1,0,0)^T$*? First set up the vector equation and try to solve it using Gauss elimination:*

$$x_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \sim \left(\begin{array}{cc|c} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)$$

Therefore, the system is inconsistent and the ant cannot reach the point $(1,0,0)^T$*.*

Example 17 *The vectors* $\vec{u}_1 = (1,2)^T$ *and* $\vec{u}_2 = (3,1)^T$ *span all* $\mathbb{R}^2$*. Can you prove that? In other words, assume* $\vec{b} = (b_1,b_2)^T$*. If one wants to prove that all vectors of* $\mathbb{R}^2$ *are in the span of* $\vec{u}_1$ *and* $\vec{u}_2$*, one needs to show that the echelon form system associated to the vector equation* $x_1\vec{u}_1 + x_2\vec{u}_2 = \vec{b}$ *does not have false equality statements. Computationally, one just have to find the echelon equivalent to the coefficient matrix* $\begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$ *and show that it has no zero rows.*

Remark 25 *One needs at least* n *vectors in order to span* $\mathbb{R}^n$*. The simplest list of vectors spanning* $\mathbb{R}^n$ *is* $\{\vec{e}_1 = (1,0,\ldots,0)^T, \vec{e}_2 = (0,1,0,\ldots,0)^T, \ldots, \vec{e}_n = (0,0,\ldots,0,1)^T\}$*. Although this is not the only spanning set for* $\mathbb{R}^n$*, it requires more work to show that one cannot span* $\mathbb{R}^n$ *with strictly less than* n *vectors.*

From now on, given an Euclidean space $\mathbb{R}^n$, one would call the **standard basis** of $\mathbb{R}^n$ to $\{\vec{e}_i \in \mathbb{R}^n : i = 1,2,\ldots,n\}$ where $\vec{e}_i$ is the vector with zeros everywhere but a one in the i-th entry.

Theorem 3 *If $\vec{u} \in Span(\vec{u}_1, \ldots, \vec{u}_m)$, then $Span(\vec{u}, \vec{u}_1, \ldots, \vec{u}_m) = Span(\vec{u}_1, \ldots, \vec{u}_m)$. In other words, one cannot modify the span of a list by adding to the list a linear combination of these vectors.*

Proof: $\vec{u} = a_1\vec{v}_1 + \ldots + a_n\vec{v}_n$. Then

$$\begin{aligned} Span(\vec{u}, \vec{u}_1, \ldots, \vec{u}_m) &= \{x_0\vec{u} + x_1\vec{v}_1 + \ldots + x_n\vec{v}_n\} \\ &= \{x_0(a_1\vec{v}_1 + \ldots + a_n\vec{v}_n) + x_1\vec{v}_1 + \ldots + x_n\vec{v}_n\} \\ &= \{(a_1x_0 + x_1)\vec{v}_1 + \ldots + (a_nx_0 + x_n)\vec{v}_n\} \\ &= \{y_1\vec{v}_1 + \ldots + y_n\vec{v}_n / y_i = a_ix_0 + x_i\} \\ &= Span(\vec{v}_1, \ldots, \vec{v}_n) \end{aligned}$$

Remark 26 *The zero vector is a linear combination of any list of vectors by just choosing all coefficients equal to zero, i.e. the zero vector is in the span of any list of vectors.*

Theorem 4 *Take $S = \{\vec{u}_1, \ldots, \vec{u}_m\} \subseteq \mathbb{R}^n$. If $m < n$, then S cannot span $\mathbb{R}^n$, i.e. $Span(\vec{u}_1, \ldots, \vec{u}_m)$ is a proper subset of $\mathbb{R}^n$. Yet if $m \geq n$, then S might span $\mathbb{R}^n$, but one cannot conclude anything without additional information.*

Indeed, if $\mathbb{R}^n$ can be spanned by $m \leq n-1$ vectors, it can be spanned by $n-1$ vectors, by adding to the list linear combinations of the original vectors until one achieves $n-1$ vectors, assume then $\mathbb{R}^n$ is spanned by $\vec{u}_1, \ldots, \vec{u}_{n-1}$. Then one needs to find a vector $\vec{b} \in \mathbb{R}^n$ for which the equation $x_1\vec{u}_1 + \ldots + x_{n-1}\vec{u}_{n-1} = \vec{b}$ is inconsistent, thus showing that the assumption is incorrect. So, pick a general vector $\vec{b} = (b_1, \ldots, b_n)$ and apply Gauss Elimination. Since one started with $n-1$ variables, then the echelon form system has at most $n-1$ equations. Hence one has at least one equality statement, where 0 is equal to a linear combination of the entries of $\vec{b}$. Choose these entries so the equality statement is false. Hence, one has found a vector $\vec{b}$ which is not in the span of the vectors $\{\vec{u}_i / 1 \leq i \leq n-1\}$.

3.2 Multiplication of a Matrix times a vector

By adding additional parenthesis, one can understand a matrix being formed by adjoining column vectors, for example

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \left[\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} \right]$$

Assume $\vec{a}_1, \ldots, \vec{a}_m \in \mathbb{R}^n$. Let $A_{m \times n} = [\vec{a}_1, \ldots, \vec{a}_m]$ and $\vec{x}_{m \times 1} = (x_1, \ldots, x_m)^T$. Then the product of the matrix A times the vector $\vec{x}$ is the linear combination of the columns of A whose coefficients are the entries of $\vec{x}$

$$A\vec{x} = x_1\vec{a_1} + \ldots + x_m\vec{a}_m$$

Example 18

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 3 \end{pmatrix} + b \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} a + 2b \\ 3a + 4b \end{pmatrix}$$

Theorem 5 *Given $\vec{a}_1, \ldots, \vec{a}_m \in \mathbb{R}^n$, and let $A = [\vec{a}_1, \ldots, \vec{a}_m]$, then the following statements are equivalent (TFSAE): (1) $\vec{b} \in Span(\vec{a}_1, \ldots, \vec{a}_m)$. (2) The vector equation $x_1\vec{a}_1 + \ldots x_m\vec{a}_m = \vec{b}$ is consistent. (3) The linear system of equations with augmented matrix $[\vec{a}_1\, \vec{a}_2\, \ldots\, \vec{a}_m\, \vec{b}]$ is consistent. (4) The matrix system $A\vec{x} = \vec{b}$ is consistent, where the left hand side is the multiplication of the matrix A with the vector $\vec{x}$.*

3.3 Some sample questions with solutions

1. True or False: A homogeneous linear system must be consistent and have a unique solution. False because although it must be consistent, it may have infinitely many solutions. For example, take a system with more variables than equations. After applying Gauss Elimination, the number of equations may reduce but the number of variables always stays the same. Therefore, there are more variables than equations, and thus there are free variables. And hence it must have infinitely many solutions.

2. True or False: A linear system with more variables than equations can have a single solution. False. After applying Gauss Elimination, if false equality statements appear, then the system is inconsistent. Otherwise, the system is consistent, but since the number of equations may reduce but the number of variables stays the same, then one ends up with a consistent system with free variables, and thus infinitely many solutions. Since these are all the cases, the system cannot have a single solution.

3. Let's try to understand the span of a set of vectors by looking at chess: One bishop starts in the position: Row 8 - Column 3 or $(8,3)$, and it can only move diagonally. If our objective is to describe all positions that this bishop can reach, then it can clearly move in the directions $\vec{v}_1 = (1,1)$ and $\vec{v}_2 = (1,-1)$, or its negatives $(-1,-1)$ and $(-1,1)$. Each movement is an integer multiple of either of these two vectors. So the bishop can reach positions that can be written as $(8,3)+a(1,1)+b(1,-1)$ where $a,b \in \mathbb{Z}$. But at the same time, any sequence of movements must land the bishop in the board, so both coordinates must be between 1 and 8:

$$1 \leq 8+a+b \leq 8; 1 \leq 3+a-b \leq 8$$

This system of inequalities leads to $-6 \leq b \leq 1$ and $-4 \leq a \leq 2$. Then the possible positions of the bishop through a game are given by the set

$$\{(8+a+b, 3+a-b) : -4 \leq a \leq 2; -6 \leq b \leq 1; a, b \in \mathbb{Z}\} \subseteq (8,3) + Span_{\mathbb{Z}}((1,1),(1,-1))$$

Try a few elements in this spanning set and check that the bishop can reach those positions.

4. How can one prove that $Span_{\mathbb{R}}((1,0)^T,(1,1)^T) = \mathbb{R}^2$? One needs to show that any element $(a,b) \in \mathbb{R}^2$ is a linear combination of $(1,0)^T$ and $(0,1)^T$. Hence, one needs to show that the vector equation $x_1(1,0)^T + x_2(1,1)^T = (a,b)^T$ has a solution or equivalently that the system whose augmented matrix is $\begin{bmatrix} 1 & 1 & a \\ 0 & 1 & b \end{bmatrix}$ is consistent. The system is already in echelon form, and it does not have zero rows, hence it is always consistent regardless of the vector $(a,b)^T$.

   In general, one needs to apply Gauss elimination to the coefficient matrix of the system, and show that it does not contains zero rows. Otherwise, by selecting carefully the vector $\vec{b}$, one can form an inconsistent system and the vectors would not span $\mathbb{R}^n$.

   Moreover, not every couple of vectors span $\mathbb{R}^2$. Take for example the vectors $(1,0)^T$ and $(2,0)^T$. Any linear combination of them has second entry zero, and thus not all vectors of $\mathbb{R}^2$ are in their span.

3.4 Linear Independence

We saw that several lists can span the same subset of $\mathbb{R}^n$. Concretely, by adding redundant vectors or linear combinations of a list to the list itself, the span does not change. So one needs a way to identify redundant vectors on a list and find a smaller spanning list of the same subset of $\mathbb{R}^n$. This idea motivates the property of linear independence of a list of vectors.

A set of vectors $\{\vec{u}_1, \ldots, \vec{u}_m\}$ in $\mathbb{R}^n$ is **linearly independent** if the only solution to $x_1\vec{u}_1 + \ldots + x_m\vec{u}_m = \vec{0}$ is the trivial solution $x_1 = \ldots = x_m = 0$. Otherwise the set is called **linearly dependent.**

Remark 27 *In other words, a list of vectors is linearly independent if none of its vectors can be expressed as a linear combination of the other vectors. Moreover, if the vectors $\vec{v}_1, \ldots, \vec{v}_m$ are linearly independent in $\mathbb{R}^n$, and let $S = Span(\vec{v}_1, \ldots, \vec{v}_m)$, then there is not a sublist that spans S, i.e. the list of vectors is the smallest sub-list that spans the set S.*

Remark 28 *All of these is equivalent to say that the homogeneous vector equation $x_1\vec{u}_1 + \ldots + x_m\vec{u}_m = \vec{0}$ only has the trivial solution, or the linear system of equations with augmented matrix $[\vec{u}_1, \ldots, \vec{u}_n|\vec{b}]$ only has the trivial solution, or the matrix system $U\vec{x} = \vec{b}$ where $U = [\vec{u}_1, \ldots, \vec{u}_m]$ only has the trivial solution.*

Example 19 *Any set $\{\vec{u}_1 = \vec{0}, \vec{u}_2, \ldots, \vec{u}_m\}$ containing the zero vector is linearly dependent. Indeed, $\vec{0}$ can be written as a nontrivial linear combination of the vectors $1 \cdot \vec{0} + 0 \cdot \vec{u}_2 + \ldots + 0 \cdot \vec{u}_m = 0$, which is non trivial because 1 is a nonzero coefficient.*

Example 20 *Two vectors are linearly dependent if and only if one is a scalar multiple of the other.*

Theorem 6 *A set of m vectors in $\mathbb{R}^n$, where m is bigger than n, is linearly dependent.*

Proof: Indeed, in a previous example, we showed that a homogeneous linear system with more variables than equations must have infinitely many solutions.

Proposition 2 (Distributive property of Matrix multiplication) *Given a matrix $A_{n\times m}$ and vectors $\vec{x}, \vec{y} \in \mathbb{R}^m$, then $A(\vec{x} + \vec{y}) = A\vec{x} \pm A\vec{y}$.*

Theorem 7 (Solution Decomposition) *Assume $\vec{x}_p$ is a particular solution to the non-homogeneous system $A\vec{x} = \vec{b}$, and $\vec{x}_{gh}$ is the general solution to the corresponding homogeneous system $A\vec{x} = \vec{0}$ obtained by replacing the constant vector by $\vec{0}$. Then the general solution to the non-homogeneous system $\vec{x}_g$ satisfies $\vec{x}_g = \vec{x}_p + \vec{x}_{gh}$.*

Proof: First notice that $A(\vec{x}_p + \vec{x}_{gh}) = A\vec{x}_p + A\vec{x}_{gh} = \vec{b} + \vec{0} = \vec{b}$. Hence the sum $\vec{x}_p + \vec{x}_{gh}$ is contained in the general solution of the non-homogeneous system. However, any other particular solution $\vec{y}_p$ to the non-homogeneous system satisfies $A\vec{y}_p = \vec{b}$. Hence $A(\vec{x}_p - \vec{y}_p) = \vec{0}$. So $\vec{x}_p - vecy_p$ is a solution to the homogeneous system.

Example 21 *Consider the vectors $\vec{u}_1 = (1,0)^T, \vec{u}_2 = (0,1)^T, \vec{u}_3 = (2,3)^T \in \mathbb{R}^2$ and $\vec{b} = (1,1)^T$. If one solves the non-homogeneous system with augmented matrix $[u_1, u_2, u_3|b]$, one notices that it is already in reduced echelon form:*

$$\begin{pmatrix} 1 & 0 & 2 & | & 1 \\ 0 & 1 & 3 & | & 1 \end{pmatrix}$$

The solution set is $\{(1-2s, 1-3s, s)^T = (1,1,0)^T + (-2s, -3s, s)^T\}_{s\in\mathbb{R}}$. On the other hand, if one solves the homogeneous system $[u_1, u_2, u_3|0]$, the solution set is $\{(-2s, -3s, s)^T\}_{s\in\mathbb{R}}$. Notice that the general solution for the non-homogeneous system $\vec{x}_g = (1-2s, 1-3s, s)^T$ is equal to a particular solution to the non-homogeneous system $\vec{x}_p = (1,1,0)^T$ plus the general solution for the homogeneous system $\vec{x}_{gh} = (-2s, -3s, s)^T$.

Remark 29 *From this decomposition of the solution of a system, one concludes that a list of vectors is linearly independent, if any vector equation, not necessarily the one with zero constant vector, has at most one solution.*

Theorem 8 *Let $S = \{\vec{a}_1, \ldots, \vec{a}_n\} \subseteq \mathbb{R}^n$, and $A = [\vec{a}_1 \ \ldots \ \vec{a}_n]$. Notice that the number of vectors is the same as the number of equations. TFAE: (1) S spans $\mathbb{R}^n$. (2) S is linearly independent. (3) $A\vec{x} = \vec{b}$ has a unique solution for all $\vec{b} \in \mathbb{R}^n$.*

Remark 30 *In general linear independence and spanning $\mathbb{R}^n$ have no relation, i.e. there are list which are one but not the other, or both, or none. It is in the particular case, where the number of vectors matches the dimension of the euclidean space $\mathbb{R}^n$, where one implies the other, i.e. a list is either both of them or none of them.*

3.5 Vocabulary

Euclidean space, vector, vector equation, components or entries of a vector, addition of vectors, scalar multiplication of vectors, the zero vector, linear combination of vectors, span of a list of vectors, spanning set of $\mathbb{R}^n$, linear independence, linear dependence.

3.6 Worksheet: Linear System of Equations and Vectors Algebra

1. (Geometry Question) Find a cubic polynomial $f(x)$ (a) passing through the points $(-1,-1), (1,2), (2,1)$ and $(3,5)$. (b) passing through $(1,0)$ with derivative $+2$, and passing through $(2,3)$ with derivative -1.

2. (Geometry Question) Find the equation of a plane in $\mathbb{R}^3$ passing through the points $(1,2,3)$, $(-1,0,2)$ and $(2,3,5)$.

3. (Geometry Question) Consider the infinite system of linear equations in two variables given by $ax+by=0$ where (a,b) moves along the unit circle in the plane.
   - How many solutions does this system have?
   - How many linearly independent equations in the above system give you the same set of solutions? Write down two separate such a linear systems, in vector form.
   - What happens to the infinite linear system if you add to it the equation $0x + 0y = 0$?
   - What happens to the infinite linear system if one of the equations slightly perturbs to $ax + by = c$ where c is a small positive number?

4. For each of the systems (I) $\begin{array}{rcl} 3x_1 + 2x_2 & = & 4 \\ 2x_1 - 3x_2 & = & 1 \end{array}$ and (II) $\begin{array}{rcl} x_1 + 2x_2 - 3x_3 & = & 5 \\ 2x_1 - 3x_2 + 4x_3 & = & 10 \\ x_1 - x_2 - x_3 & = & 1 \end{array}$.

   (a) Write the matrix and the augmented matrix of the system. Find the dimensions of each of these matrices. Also write the system as a vector equation. Finally write the corresponding matrix system of the form $A\vec{x} = \vec{b}$. How many hyperplanes are describing this raw system?

(b) Find an Echelon form equivalent system and the Reduced Echelon form equivalent system. Also find its solution set.

(c) Classify the system as inconsistent, consistent with a unique solution, or consistent with infinitely many solutions. If the system is consistent, what is the minimum number of hyperplanes necessary to describe this system?

(d) List the leading variables, the free variables, the pivot positions and the leading positions of this system.

5. True or False: If $\{\vec{v}_1, \vec{v}_2\}, \{\vec{v}_1, \vec{v}_3\}, \{\vec{v}_2, \vec{v}_3\}$ are three linearly independent sets, then $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent. If false, write an example to justify it.

6. True or False: If $\vec{v}_1 \notin Span(\vec{v}_2, \vec{v}_3)$ and $\vec{v}_2 \notin Span(\vec{v}_1, \vec{v}_3)$, then $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent. If false, write an example to justify it.

4 Matrices and Linear Transformations

4.1 Matrix Algebra

Matrices are rectangular arrangements of numbers. However, as well as with vectors, studying the set of all matrices with a fix dimension (pair) would be useless unless one provides such set with more structure. Please review the definitions of a matrix, the dimension or dimensions of a matrix, diagonal matrices, and upper and lower triangular matrices.

One would name a matrix with a capital letter, and we usually write $A_{n\times m} = (a_{i,j})_{n\times m}$ to indicate that one would refer as $a_{i,j}$ to the entries of A.

$$A_{m\times n} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix}_{m\times n}$$

The additional structure of the set of matrices of a fix dimension $m \times n$ includes the definition of equality, addition, and scalar multiplication. Give two matrices $A_{m\times n} = (a_{i,j})$ and $B_{m\times n} = (b_{i,j})$ of the same dimension $m \times n$, one has:

1. The two matrices A and B are equal matrices if and only if the corresponding entries are equal, i.e. for all $i = 1, 2, \dots, m$ and for all $j = 1, 2, ldots, n$ one has $a_{i,j} = b_{i,j}$.

2. The sum or difference of the matrices $A \pm B$ is the matrix of dimension $m \times n$ obtained by adding or subtracting the corresponding entries, i.e. the entry at the (i, j) position is $a_{i,j} \pm b_{i,j}$ or $A \pm B = (a_{i,j} \pm b_{i,j})$.

   Remark 31 *One cannot add or subtract matrices of different dimensions.*

3. Given a real number or scalar $r \in \mathbb{R}$, the entries of the scalar multiple rA of A are obtained by distributing r over all entries of A, i.e. the entry at the (i, j) position is $ra_{i,j}$ or $rA = (ra_{i,j})$.

Proposition 3 *Given scalars $s, t \in \mathbb{R}$ and Matrices A, B of the same dimension $m \times n$, one has:*

- *Commutative addition of matrices: $A + B = B + A$*
- *Distribution of Scalar Multiplication over sum: $s(A + B) = sA + sB$ and $(s + t)A = sA + tA$.*
- *Associative addition of matrices: $(A+B)+C = A+(B+C)$. And associative scalar multiplication $(st)A = s(tA)$.*
- *There is a zero matrix, all whose entries are zero, which is the additive identity, i.e. $A + 0 = 0 + A = A$. Notice that 0 will represent a number, a vector, or a matrix depending on the context.*

However, there is an operation between matrices of different dimensions, which is **matrix multiplication**. Matrix multiplication is not commutative, which means that AB is not necessarily equal to BA. Also matrix multiplication does not occur among any two matrices. The product AB of $A_{m\times n}$ with $B_{n\times p}$ exists if and only if the second dimension n of the left matrix A is equal to the first dimension n of the second matrix B. The resulting matrix is of dimension $m \times p$, i.e. the first dimension of the first matrix times the second dimension of the second matrix.

If $A_{m\times n} = (a_{i,j})$, $B_{n\times p} = (b_{i,j})$ and $C_{m\times p} = AB = (c_{i,j})$, then the i, j-th entry of the product C is obtained by the formula:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{ir}b_{rj} = \sum_{k=1}^{r} a_{ik}b_{ki}$$

Alternative, one splits the second matrix B into column vectors, say $B = [\vec{b}_1, \ldots, \vec{b}_n]$, then $AB = [A\vec{b}_1, \ldots, A\vec{b}_n]$, where one uses the definition of the product of a matrix with a vector given above.

Proposition 4 *Given real numbers or scalars $r, s \in \mathbb{R}$, and matrices A, B, C (which one needs to assume the dimensions allow to perform the corresponding operation in each item):*

- *Associative multiplication: $A(BC) = (AB)C$.*
- *Distribution of multiplication over sum: $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$.*
- *Associative multiplication with scalar multiplicaion: $s(AB) = (sA)B = A(sB)$.*
- *For ever positive integer n, there is a multiplicative identity matrix I_n, which is a diagonal matrix with only ones in the diagonal. Thus, $AI = IA = A$.*

$$I_n = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

However, some matrix multiplication properties don't resemble the corresponding number properties:

- *Matrix multiplication is in general not commutative: $AB \neq BA$.*
- *The cancellation properties don't hold: (1) $AB = 0$ does not imply that $A = 0$ or $B = 0$, and (2) $AC = BC$ does not imply that $A = B$ or $C = 0$.*

An additional operation on matrices is Transposing, which is exchanging the roles of rows and columns, almost like rotating the matrix 90 degrees. The transpose of $A = (a_{i,j})$ is denoted $A^T = (a_{j,i})$. Thus,

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}_{n\times m} \Rightarrow A^T = \begin{pmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{pmatrix}_{n\times m}$$

Proposition 5 *Given real numbers or scalars $s, t \in \mathbb{R}$ and matrices A, B:*

- *The transpose of a sum or difference is the sum or difference of the transpose $(A \pm B)^T = A^T \pm B^T$.*
- *The transpose of a scalar multiple of a matrix is the scalar multiple of the transpose: $(sA)^T = sA^T$.*
- *The transpose of the product is the reverse product of the transposes: $(AB)^T = B^T A^T$.*

Example 22

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}^T_{3\times 1} \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}_{3\times 2} \begin{pmatrix} R \\ S \end{pmatrix}_{2\times 1} = \begin{pmatrix} X & Y & Z \end{pmatrix}_{1\times 3} \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}_{3\times 2} \begin{pmatrix} R \\ S \end{pmatrix}_{2\times 1}$$

By associativity of multiplication, one has choices:

$$= \begin{pmatrix} X & Y & Z \end{pmatrix}_{1\times 3} \begin{pmatrix} aR + bS \\ cR + dS \\ eR + fS \end{pmatrix}_{3\times 1} = X(aR + bS) + Y(cR + dS) + Z(eR + fS)$$

or alternatively

$$= \begin{pmatrix} aX + cY + eZ & bX + dY + fZ \end{pmatrix}_{1\times 2} \begin{pmatrix} R \\ S \end{pmatrix}_{2\times 1} = R(aX + cY + eZ) + S(bX + dY + fZ)$$

Example 23 ***Suggestion 1***: *Write I_3 and multiply it on the right by the matrix $A = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}_{3\times 2}$, i.e. find IA.*

Example 24 ***Suggestion 2:*** *Write I_2 and multiply it on the left by the matrix* $A = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}_{3\times 2}$*, i.e. find AI.*

Remark 32 *In general, finding the power A^n of a matrix A is not easy. One defines A^n to be the matrix obtained by multiplying A by itself n times. However, because of the definition of the product of matrices, if $A = (a_{i,j})$, then in general A^n is not equal to $(a^n_{i,j})$.*

Proposition 6 (Powers of Diagonal Matrices) *Let $D_{m\times m} = (d_{i,j})$ be a diagonal matrix, i.e. the only possibly nonzero entries are $d_{11}, d_{22}, \ldots, d_{mm}$. Then for any positive integer $k \in \mathbb{Z}^+$, one has $D^k = (d^k_{i,j})$, i.e.:*

$$D_n = \begin{pmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_{nn} \end{pmatrix}_{n\times n} \Rightarrow D_n^k = \begin{pmatrix} d^k_{11} & 0 & \ldots & 0 \\ 0 & d^k_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d^k_{nn} \end{pmatrix}_{n\times n}$$

Example 25 *The multiplication of a diagonal matrix and an upper triangular matrix is upper triangular:*

$$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}_{2\times 2} \begin{pmatrix} m & n \\ 0 & p \end{pmatrix}_{2\times 2} = \begin{pmatrix} am + 0\cdot 0 & an + 0\cdot p \\ 0\cdot m + d\cdot 0 & 0\cdot n + dp \end{pmatrix} = \begin{pmatrix} am & an \\ 0 & dp \end{pmatrix}_{2\times 2}$$

Example 26 *The multiplication of an upper triangular matrix and a diagonal matrix is upper triangular:*

$$\begin{pmatrix} m & n \\ 0 & p \end{pmatrix}_{2\times 2} \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}_{2\times 2} = \begin{pmatrix} ma + n\cdot 0 & m\cdot 0 + nd \\ 0\cdot a + p\cdot 0 & 0\cdot 0 + pd \end{pmatrix} = \begin{pmatrix} am & dn \\ 0 & pd \end{pmatrix}_{2\times 2}$$

Remark 33 *The product of two triangular matrices of the same type is another triangular matrix of the same type.*

4.2 Partitioned matrices

The process of partition a matrix can be understood as a subspliting a matrix into well fitting rectangular sub-arrangements. Notationally, this would mean to add extra parenthesis and to understand the sub-blocks or pieces as submatrices contained in the original matrix.

Example 27 *Given* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$*, one would understand by* $B = \begin{pmatrix} A & 0_{2\times 2} \\ 0_{2\times 2} & A^T \end{pmatrix}$ *to be* $B = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 4 \end{pmatrix}$.

Inversely, one could start with $B = \begin{pmatrix} 1 & 2 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 4 \end{pmatrix}$ *and then add the extra parenthesis such that*

$$B = \begin{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \end{pmatrix}$$

in order to realize that $B_{4\times 4}$ is conformed by four submatrices 2×2.

4.3 Review

A function is a rule assigning an element called the image in a set called the *codomain* to each element called the argument in another set call the *domain* of the function. Thus, for example $f : \mathbb{R} \to [0, \infty)$ where $f(x) = x^2$ is the function that takes real numbers $x \in \mathbb{R}$ and returns another number $f(x) = x^2 \in [0, \infty)$. The subset of all images of all possible arguments is a subset of the codomain called the **range**.

The **graph of a function** f is the set of all points of the form $(x, f(x))$ in the cross product of the domain times the codomain. For the example above, the graph of f is $\{(x, x^2) \in \mathbb{R} \times [0, \infty)\}$.

In the first two quarter of calculus, one studies functions of one variable or $\mathbb{R} \to \mathbb{R}$; later one studies functions of several variables or $\mathbb{R}^n \to \mathbb{R}$ and $\mathbb{R}^n \to \mathbb{R}^m$. For example the function $g : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x \sin(y)$, or the function $h : \mathbb{R}^2 \to \mathbb{R}^3$ given by $h(x, y) = (x + y, x - y, xy)$. In linear algebra, one is interested in these multivariables functions which are linear.

Remark 34 *Notice that in calculus, we used to write n-tuples as row vectors, but in here we will write them as column vectors.*

One verifies if a relation is a function using the **vertical test**: In the graph of f, if any vertical touches the graph at most once, then f is a function. For example $x^2 + y^2 = 1$ is not a function of either variable, but the union of two functions.

A function f is said to be **injective** if it assigns different values to different arguments, i.e. if $f(x) = f(y)$ implies $x = y$. A function f is said to be **surjective** if every element in the codomain is an image of an element in the domain, i.e. the range is not a proper subset of the codomain, but the range is actually equal to the codomain.

The composition of two functions $f : A \to B$ and $g : B \to C$ is defined $g \circ f : A \to C$ where $g \circ f(x) = g(f(x))$. Notice the reverse notation $g \circ f$ where one applies first f to the argument and then g to the result. This is better written as $A \xrightarrow{f} B \xrightarrow{g} C$. In general, **REMEMBER** that if $f \circ g$ is bijective, then g is injective and f is surjective.

Remark 35 *This is not linear algebra but you need to know this.*

4.4 Linear Transformations

A **Linear Transformation** is a linear function, without the constant terms, from one euclidean space to another euclidean space, i.e. $T : \mathbb{R}^m \to \mathbb{R}^n$. We will characterize linear transformations in several ways:

1. A linear transformation is a function $T : \mathbb{R}^m \to \mathbb{R}^n$ such that each component of the image is a linear function of the components of the argument.

$$T(\vec{x}) = A \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1m}x_m \\ a_{21}x_1 + \ldots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + \ldots + a_{nm}x_m \end{pmatrix} = \begin{pmatrix} a_{11} & \ldots & a_{m1} \\ \vdots & \ddots & \ldots \\ a_{n1} & \vdots & a_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

   Hence, each linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ corresponds to a matrix, which one denotes $A_T = (a_{i,j})$ or simply T too, called the **coefficient matrix** of the linear transformation or the **associated matrix** to it.

   Remark 36 *There is a bijection between matrices and linear transformations, i.e. any linear transformation determines a unique matrix, its coefficient matrix; and any matrix determines a unique linear transformation by the rule of multiplying by this matrix on the left. However, linear transformations and matrices enjoy different properties. Moreover, the coefficient matrix A of a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ has dimensions $n \times m$ (backward).*

   Remark 37 *Notice that each component of the image is linear in the components of the argument, but they do not have constant terms. If they had constant terms, it would be called an* ***affine transformation****, but we won't discuss affine transformations in this class.*

2. The abstract definition: A linear transformation is a function $T : \mathbb{R}^m \to \mathbb{R}^n$ satisfying two properties: (1) It commutes with addition, i.e. $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$, (2) It commutes with scalar multiplication, i.e. $T(r\vec{u}) = r\,T(\vec{u})$. Together they imply that $T(a\vec{x} \pm b\vec{y}) = aT(\vec{x}) \pm bT(\vec{y})$.

   Remark 38 *This is useful for proofs and it is a generalization of linearity in the lack of geometrical graphs. It represents the core of linearity.*

3. Linear transformations include rotations, projections, reflections, expansions and contractions, or composition of these transformations. However, linear transformations do not include translations.

Remark 39 *Recall that the standard basis of $\mathbb{R}^m$ is $\{\vec{e}_i \in \mathbb{R}^m : i = 1, 2, \ldots, m\}$. Notice that for a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, its coefficient matrix can be computed as $A_T = [T(\vec{e}_1), T(\vec{e}_2), \ldots, T(\vec{e}_m)]$.*

Example 28 *Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + 3x_2 \\ 5x_1 + 7x_2 \end{pmatrix}$. Then $T(\vec{e}_1) = T\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$ and $T(\vec{e}_2) = T\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$. Hence $A_T = \begin{pmatrix} 2 & 3 \\ 5 & 7 \end{pmatrix}$.*

The **range** or **image** of a linear transformation is its image as a function, or the set of all images of all possible arguments from the domain. It turns out to be that the range is the span of the column vectors of the coefficient matrix, i.e. if $T : \mathbb{R}^m \to \mathbb{R}^n$ has coefficient matrix $A = [\vec{a}_1, \ldots, \vec{a}_m]$, then $Range(T) = Span(\vec{a}_1, \ldots, \vec{a}_m)$. By the definition of surjectivity, one concludes that T is **surjective** or **onto** if and only if $Range(T) = Span(\vec{a}_1, \ldots, \vec{a}_m) = \mathbb{R}^n$. Since one needs at least n vectors to span $\mathbb{R}^n$, then in order for T to be surjective, one needs at least $m \geq n$, i.e. the dimension of the domain is at least as large as the dimension of the codomain. Yet, this is not enough as we will see later.

Proposition 7 *A linear transformation T with coefficient matrix A is surjective if and only if the span of the columns of A is the codomain of T, and this in turn happens if and only if any system $A\vec{x} = \vec{b}$ for any vector $\vec{b}$ in the codomain of T has at least one solution.*

A new element defined for linear transformations that is not defined for functions is the kernel of the transformation. The kernel $Ker(T)$ of a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ is the set of arguments $\vec{x} \in \mathbb{R}^m$ that map to the zero vector $\vec{0} \in \mathbb{R}^n$. By the definition of injectivity together with the abstract properties of linear transformations, one conclude that T is **injective** or **one-to-one** if $Ker(T) = \{\vec{0}\}$, i.e. the only vector mapping to zero, is the zero vector. In order for T to be injective, one needs at least $m \leq n$, i.e. the dimension of the codomain is at least as large as the dimension of the domain. Yet, this is not enough as we will see later.

Proposition 8 *A linear transformation T with coefficient matrix A is injective if and only if its kernel is trivial if and only if the columns of A are linearly independent, i.e. only contains the zero vector, and this in turn happens if and only if any system $A\vec{x} = \vec{b}$ for any vector $\vec{b}$ in the codomain of T has at most one solution.*

Example 29 (Summary) *There are different ways to define completely a linear transformation completely, so that one can determine all its elements.*

- *A linear transformation can be defined by its rule: If* $T(\vec{x}) = \begin{bmatrix} 2x_1 + 3x_3 \\ 2x_1 - 2x_2 - x_3 \\ x_2 \end{bmatrix}$. *Then clearly,* $T : \mathbb{R}^3 \to \mathbb{R}^3$*; the domain because there are three variables* x_i*'s, and the codomain because the rule vector has three components. Additionally, its coefficient matrix can be read from the coefficients of the variables involved:* $A = \begin{bmatrix} 2 & 0 & 3 \\ 2 & -2 & -1 \\ 0 & 1 & 0 \end{bmatrix}$. *Notice that the first column* $(2, 2, 0)^T$ *corresponds to the the coefficients of* x_1*, and so on.*

- *A linear transformation can be defined by describing the image of the standard basis:* $T(\vec{e}_1) = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, T(\vec{e}_2) = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix}, T(\vec{e}_3) = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}$. *Then its associated matrix is obtained by writing the vector images as columns of the matrix:* $A = \begin{bmatrix} 2 & 0 & 3 \\ 2 & -2 & -1 \\ 0 & 1 & 0 \end{bmatrix}$. *Notice that to compute images, we only need linearity:*

$$\vec{x} = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = 2\vec{e}_1 - 3\vec{e}_2 + 4\vec{e}_3 \Rightarrow T(\vec{x}) = 2T(\vec{e}_1) - 3T(\vec{e}_2) + 4T(\vec{e}_3) = 2\begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} + 4\begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 16 \\ 6 \\ -3 \end{bmatrix}$$

- *Later we will see that one can defined a linear transformation by declaring what it does to a any linearly independent spanning set of the domain, also called a basis of the domain: If* $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ *is a linearly independent spanning set of* $\mathbb{R}^3$*, then* $T(\vec{v}_1) = \vec{w}_1, T(\vec{v}_2) = \vec{w}_2, T(\vec{v}_3) = \vec{w}_3$ *implies that* $T[\vec{v}_1, \vec{v}_2, \vec{v}_3] = [\vec{w}_1, \vec{w}_2, \vec{w}_3] \Rightarrow A = [\vec{w}_1, \vec{w}_2, \vec{w}_3][\vec{v}_1, \vec{v}_2, \vec{v}_3]^{-1}$.

4.5 Rotations

Rotations of the plane are linear transformations of the form $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose coefficient matrix is of the form $A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$, it rotates every vector conterclockwise by an angle θ.

Example 30 *The linear transformation whose associated matrix is* $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ *is the rotation of the plane by* 90 *degrees.*

Example 31 *The matrices* $I_2, \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ *are four different solution to the matrix equation* $A^4 = I_2$ *because they correspond to rotations by* $0, 90, 180, 270$ *degrees and if one applies any of them four consecutive times, one obtains back the identity.*

Example 32 *A linear transformation* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *with coefficient matrix of the form* $A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$ *is a rotation of the space around the* z*-axis by an angle* θ *counterclockwise on the* xy*-plane.*

4.6 Inverses

Recall that one calls 1 the multiplicative identity of real numbers because $1 \cdot r = r \cdot 1 = r$. Similarly, we defined the identity matrix I_n to satisfy $I_m \cdot A_{m\times n} = A_{m\times n} \cdot I_n = A_{m\times n}$. Additionally, for non-zero real numbers r, one defines its reciprocal $1/r$, which is the multiplicative inverse $r \cdot (1/r) = 1$.

Similarly, we are left with the task of finding **the inverse A^{-1} of a given matrix** A such that $AA^{-1} = A^{-1}A = I$. By the dimension restrictions of multiplication, one finds that for A to be such an inverse A must be a square matrix, i.e. $A_{n\times n}$. Moreover, if A is invertible, then the inverse A^{-1} is unique.

Theorem 9 *The first algorithm to find the inverse of a square matrix* A *is:*

1. *Append the corresponding identity matrix:* $A \to [A_n|I_n]_{n\times 2n}$.
2. *Apply Jordan Gauss elimination to this extended matrix.*
3. *If the right half* $n \times n$ *matrix is the identity, then* A *is invertible and one can read* A^{-1} *from the reduced echelon matrix, which is* $[I_n|A^{-1}]$. *Otherwise* A *is not invertible.*

Example 33 *Assume* $ad - bc \neq 0$:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow [A|I_2] = \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \xrightarrow{JG} \left(\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right) \Rightarrow A^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Example 34 *Notice that*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}_{2\times 2} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}_{2\times 2} = \begin{pmatrix} ad + b(-c) & a(-b) + ba \\ cd + d(-c) & c(-b) + da \end{pmatrix}_{2\times 2} = \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}_{2\times 2}$$

. Notice that the columns of the matrix are linearly independent vectors, and thus they also span $\mathbb{R}^2$.

Example 35 *Try to find the inverse of the matrix* $A = \begin{pmatrix} 4 & 6 \\ 2 & 3 \end{pmatrix}_{2\times 2}$. *What goes wrong? Notice that the columns of the matrix are linearly dependent vectors, and thus they do not span* $\mathbb{R}^2$.

Similarly, for functions $f : A \to B$ where $y = f(x)$, one defines its composition inverse to be another function $f^{-1} : B \to A$ where $y = g(x)$ and $f(g(x))) = x$ and $g(f(x)) = x$. Moreover, a function has a an inverse if it is **bijective**, i.e. if it is both injective and surjective.

The **inverse** $T^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ **of a linear transformation** $T : \mathbb{R}^m \to \mathbb{R}^n$ satisfies $T^{-1} \circ T(\vec{x}) = \vec{x}$ and $T \circ T^{-1}(\vec{x}) = \vec{x}$, i.e. The composition of T and T^{-1} is the identity transformation on the domain whose associated matrix is the identity matrix, and similarly the composition of T^{-1} and T is the identity transformation on the codomain. From the bijective condition, one concludes that for T to be invertible, its domain must be equal to its codomain, i.e. $T : \mathbb{R}^n \to \mathbb{R}^n$. If the inverse exists, then it is unique, thus one speaks of the inverse and not an inverse.

Proposition 9 *A linear transformation* T *with coefficient matrix* A *is invertible if and only if (1) It is both injective and surjective, forcing* $m = n$, *yet it is enough to check that* $n = m$ *and that either it is injective or it is surjective. (2)* A *has linearly columns spanning the codomain of* T *forcing it to be a square matrix, yet it is enough to check that it is a square matrix with either linearly independent columns or columns that span the codomain. (3) Any system* $A\vec{x} = \vec{b}$, *for any* $\vec{b}$ *in the codomain, has exactly one solution.*

Proposition 10 *If A and B are invertible matrices, then (1) $(A^{-1})^{-1} = A$; (2) $(AB)^{-1} = B^{-1}A^{-1}$; (3) The cancelation property holds: $AC = AD \Rightarrow C = D$ and $CA = DA \Rightarrow C = D$.*

Example 36 *Find the inverse of the matrix $A = \begin{pmatrix} 1 & 3 \\ 4 & 5 \end{pmatrix}_{2\times 2}$. Step 1: Check that A is a square matrix: $n = 2$ because the dimensions is 2×2. Clearly, the columns of A are linearly independent and span $\mathbb{R}^2$, so one expects A to be invertible. Step 2: Write the partitioned matrix $[A|I_n]$ and apply Jordan Gauss elimination:*

$$\left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{array}\right)_{2\times 4} \sim \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & -7 & -4 & 1 \end{array}\right)_{2\times 4} \sim \left(\begin{array}{cc|cc} 1 & 3 & 1 & 0 \\ 0 & 1 & 4/7 & -1/7 \end{array}\right)_{2\times 4} \sim \left(\begin{array}{cc|cc} 1 & 0 & -5/7 & 3/7 \\ 0 & 1 & 4/7 & -1/7 \end{array}\right)_{2\times 4}$$

By (1) subtracting 4 times row 1 from row 2; (2) dividing row 3 by -7; (3) subtract 3 times row 2 from row 1. Step 3: Then notice one obtains obtain is of the form $[I_n|A^{-1}]$. Hence, A is invertible, and one can read its inverse from the right half of this matrix: $A^{-1} = \begin{pmatrix} -5/7 & 3/7 \\ 4/7 & -1/7 \end{pmatrix}_{2\times 2}$.

4.7 Worksheet: Matrix algebra and Linear Transformations

1. Given the following functions, identify whether they are injective, surjective, and/or bijective.
   (a) $f_1 : \mathbb{R} \to \mathbb{R}; f_1(x) = x^2$.
   (b) $f_2 : (-1, 1) \to \mathbb{R}; f_2(x) = x^2$.
   (c) $f_3 : \mathbb{R} \to [0, \infty); f_3(x) = x^2$
2. (Geometry Question) Suppose we are given the unit square A in the plane with corners $(0, 0)$, $(1, 0)$, $(1, 1)$ and $(0, 1)$. (a) Find a linear transformation T that sends A to the parallelogram B with corners $(0, 0)$, $(1, 2)$, $(2, 2)$ and $(1, 0)$.
3. (Geometry Question) How can you map the triangle $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$ to the plane so that its area is preserved and one of its corners is $(0, 0)$?
4. Given the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ with $A_T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$. Identify whether T is injective, surjective, bijective or none.
5. Consider two linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^3$ and $S : \mathbb{R}^3 \to \mathbb{R}^2$ with associated matrices A and B. Consider the composition of linear transformations $S \circ T : \mathbb{R}^2 \xrightarrow{T} \mathbb{R}^3 \xrightarrow{S} \mathbb{R}^2$.
   (a) What is the associated matrix for $S \circ T$ in terms of A and B.
   (b) If $S \circ T$ is bijective, what can we say about T: is it injective? is it surjective? is it bijective? what can we say about S: is it injective? is it surjective? is it bijective?
6. True or False: If it is false, write an example to show it.
   - The sum or difference of two diagonal matrices is a diagonal matrix.
   - The product of two diagonal matrices is a diagonal matrix.
   - The product of two lower triangular matrices is a lower triangular matrix; and the product of two upper triangular matrices is an upper triangular matrix.
   - An 3×3 triangular matrix A with 0 entries in the diagonal must satisfy $A^3 = 0$.
   - If the n-th power of a diagonal matrix D is zero, then $D = 0$.
7. Dimension Analysis: Assume A is $m \times n$ dimensional. Find the dimension of the other matrices or the conditions on m and n, so the operations are meaningful. For example, in order for $AB + A$ to have any meaning, we know that AB must have meaning so the first dimension of B is n (otherwise they could not multiply each other. So $B = B_{n\times p}$. But then $(AB)_{m\times p}$, and in order to be able to add it to A, they must have the same dimensions, so $m \times p = m \times n$ and hence $n = p$. Therefore, the dimension of B is $B_{n\times n}$.
   - $AB + B$.
   - $AC + DA$.
   - $AB + (BA)^T$.

4.8 Worksheet: Matrix algebra and Linear Transformations

For each of the following linear transformations with domain $\mathbb{R}^2$ and codomain $\mathbb{R}^2$, find the range and the coefficient matrix. Also determine whether it is one-to-one and onto.

1. A linear transformation such that $T\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}35\\-2\end{bmatrix}$ and $T\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}-7\\15\end{bmatrix}$.
2. A linear transformation such that $T\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}12x_1+13x_2\\-14x_2\end{bmatrix}$.
3. A linear transformation such that $T\begin{bmatrix}2\\3\end{bmatrix} = \begin{bmatrix}1\\-3\end{bmatrix}$ and $T\begin{bmatrix}-1\\2\end{bmatrix} = \begin{bmatrix}5\\-2\end{bmatrix}$.
4. A linear transformation T with $Range(T) = Span\left\{\begin{bmatrix}2\\3\end{bmatrix}\right\}$. Hint: there are infinitely many such transformations.
5. A linear transformation T with $Ker(T) = Span\left\{\begin{bmatrix}3\\-2\end{bmatrix}\right\}$. Hint: there are infinitely many such transformations.
6. A linear transformation that sends the unit square (the region or area, not just the perimeter) with vertices at $\vec{0}, \vec{e}_1, \vec{e}_2, \vec{e}_1+\vec{e}_2$ to the parallelogram (the region or area, not just the perimeter) with vertices at $\vec{0}$, $\begin{bmatrix}1\\-1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}$
7. A linear transformation that sends the unit square (the region or area, not just the perimeter) with vertices at $\vec{0}, \vec{e}_1, \vec{e}_2, \vec{e}_1+\vec{e}_2$ onto the segment with endpoints at $\vec{0}$ and $\vec{e}_1 - \vec{e}_2$.

5 Subspaces

Just like one has curves in the real plane $\mathbb{R}^2$, curves and surfaces in the real space $\mathbb{R}^3$, one needs to define linear subsets of the Euclidean space $\mathbb{R}^n$. When one can geometrically draw the plane and the space, one forms a notion of linear by studying lines and planes. But one is left with the task of generalizing the concept of being linear in higher dimensions, where one does not have these enlightning pictures.

Example 37 *A single equation $ax+by=c$ in $\mathbb{R}^2$ represents a line, which is a 1-dimensional object in the 2-dimensional Euclidean space.*

Example 38 *A single equation $ax+by+cz=d$ in $\mathbb{R}^3$ represents a plane, which is a 2-dimensional object in the 3-dimensional Euclidean space.*

Example 39 *A single equation represents in $a_1x_1+a_2x_2+a_3x_3+\ldots+a_nx_n=b$ in $\mathbb{R}^n$ represents a hyperplane, which is an $n-1$ dimensional object in the n-dimensional Euclidean space.*

Any other subset of an Euclidean Space with the linear properties is called a **Linear Subspace** or simply Subspace. Thus, they have additional structure than that of simple subsets, which is the linear properties.

A linear subspace S of the Euclidean space $\mathbb{R}^n$ is a subset satisfying one of the following three properties:

1. **Abstractly:** It contains the zero vector, i.e. $\vec{0}\in S$; It is closed under addition, i.e. for any two vectors $\vec{x},\vec{y}$ in S, the sum $\vec{x}+\vec{y}$ is also in S; finally, S must be closed under scalar-multiplication, i.e. for any vector $\vec{u}\in S$ and any scalar $r\in\mathbb{R}$, their product $r\vec{u}$ is also in S.

   Remark 40 *This is important for proofs, and it allows us to characterize linearity without recurring to a picture of the object.*

2. S is the solution set of a homogeneous linear system.

3. S is the span of some vectors in the Euclidean space containing it.

   Remark 41 *These later two definitions are used for computations.*

Example 40 *Determine if the set S of all vectors $(a,b,c)^T$ where $a+b+c=0$ is a subspace of $\mathbb{R}^3$.*

- *Using the abstract definition: Step 1: Notice that $\vec{0}_3\in S$ because $\vec{0}=(0,0,0)^T$ and $0+0+0=0$. Step 2: If $\vec{a},\vec{b}\in S$, then $a_1+a_2+a_3=0$ and $b_1+b_2+b_3=0$. Therefore, $(a_1+b_1)+(a_2+b_2)+(a_3+b_3)=0$ and thus $\vec{a}+\vec{b}=0$. Step 3: If $\vec{a}\in S$ and $r\in\mathbb{R}$, then $a_1+a_2+a_3=0$ and thus $ra_1+ra_2+ra_3=r(a_1+a_2+a_3)=0$. Hence $r\vec{a}\in S$. Therefore, S satisfies the three abstract properties of a linear subspace.*

- *Notice that S is the solution set of the single equation $x_1+x_2+x_3=0$, which can be considered a homogenous system of equations.*

- *Since $a+b+c=0$, then $c=-a-b$, so $S=Span((1,0,-1)^T,(0,1,-1)^T)$ because*

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}=\begin{pmatrix} a \\ b \\ -a-b \end{pmatrix}=a\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}+b\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

Another example comes from the kernel and the range of a linear transformation T. The kernel is a subspace of the domain, and it can be seen to be a subspace because it is the solution set to the system $T\vec{x}=\vec{0}$. The range is a subspace of the codomain, and it can be seen to be a subspace because it is the span of the columns of the coefficient matrix of T.

Example 41 *Consider the linear transformation $S:\mathbb{R}^3\to\mathbb{R}^3$ with matrix*

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

In other words $S(x,y,z)^T=(x+2y,2x-y,0)$.

- *The kernel is the set of all vectors* $(x,y,z)^T$ *in the domain* $\mathbb{R}^3$ *that are mapped to* $\vec{0}$, *i.e.:* $x+2y=0, 2x-y=0 \Rightarrow Ker(S)=\{(0,0,z)^T \in \mathbb{R}^3\} = Span(\vec{e}_3)$. *That means* $Ker(S)=\{\vec{x} : S(\vec{x})=\vec{0}\}$, *so it is also the span of some vectors.*

- *The range is the set of all vectors* $(a,b,0)$ *in the codomain* $\mathbb{R}^3$ *that are images of vectors in the domain, i.e.:* $a=x+2y, b=2x-y \Rightarrow S(x=\frac{a+2b}{5}, y=\frac{b-2a}{-5}, z)^T=(a,b,0)^T$. *Thus,* $Range(T)=Span(\vec{e}_1,\vec{e}_2)$.

Remark 42 *A nice reminder of these definitions and containments is*

$$\begin{array}{cccc} S: & \mathbb{R}^3 & \to & \mathbb{R}^3 \\ & Ker(S) & \to & \{0\} \\ & \mathbb{R}^3 & \to & Range(S)=Im(S) \end{array}$$

Example 42 (The Graph of a Transformation is a Linear Subspace) *Consider the linear transformation* $T:\mathbb{R}^m \to \mathbb{R}^n$. *Show that the graph of* T, *which is the set* $\Gamma(T)=\left\{\begin{pmatrix}\vec{x}\\ T\vec{x}\end{pmatrix} : \vec{x}\in\mathbb{R}^m\right\}$ *is a subspace of* $\mathbb{R}^{m+n}$.

1. *Notice that since* $T\vec{0}=\vec{0}$, *then* $\vec{0}=\begin{pmatrix}\vec{0}\\ T\vec{0}\end{pmatrix}\in\Gamma(T)$. *Secondly, by the linear properties of* T, *notice that* $\begin{pmatrix}\vec{x}\\ T\vec{x}\end{pmatrix}+\begin{pmatrix}\vec{y}\\ T\vec{y}\end{pmatrix}=\begin{pmatrix}\vec{x}+\vec{y}\\ T\vec{x}+T\vec{y}\end{pmatrix}=\begin{pmatrix}\vec{x}+\vec{y}\\ T(\vec{x}+\vec{y})\end{pmatrix}$, *hence* $\Gamma(T)$ *is closed under addition. Analogously, notice that* $r\begin{pmatrix}\vec{x}\\ T\vec{x}\end{pmatrix}=\begin{pmatrix}r\vec{x}\\ rT\vec{x}\end{pmatrix}=\begin{pmatrix}r\vec{x}\\ T(r\vec{x})\end{pmatrix}$. *So* $\Gamma(T)$ *is closed under scalar multiplication. Therefore,* $\Gamma(T)$ *is a linear subspace of* $\mathbb{R}^{m+n}$.

2. *Alternatively, notice that the elements of* $\Gamma(T)$ *are of the form* $\begin{pmatrix}\vec{x}\\ T\vec{x}\end{pmatrix}\in\mathbb{R}^{m+n}$ *where* $\vec{x}\in\mathbb{R}^m$ *whose standard basis is* $\{\vec{e}_1,\ldots,\vec{e}_m\}$. *Therefore,*

$$\Gamma(T)=Span\left\{\begin{pmatrix}\vec{e}_1\\ T\vec{e}_1\end{pmatrix},\ldots,\begin{pmatrix}\vec{e}_m\\ T\vec{e}_m\end{pmatrix}\right\}$$

   Since $\Gamma(T)$ *is the span of some vectors in* $\mathbb{R}^{m+n}$, *then* $\Gamma(T)$ *is a linear subspace of* $\mathbb{R}^{m+n}$.

3. *Alternatively, let* $A=[\vec{a}_1,\ldots,\vec{a}_m]$ *be the coefficient matrix for* T. *Notice that the elements of* $\Gamma(T)$ *are of the form* $\vec{y}=\begin{pmatrix}\vec{x}\\ T\vec{x}\end{pmatrix}\in\mathbb{R}^{m+n}$. *Let* $\vec{z}=T(\vec{x})=A\vec{x}\Rightarrow A\vec{x}-\vec{z}=\vec{0}$. *Therefore, the elements* $\vec{y}\in\Gamma(T)$ *satisfy the homogeneous system* $[A|-I_n]\vec{y}=\vec{0}$. *Hence,* $\Gamma(T)$ *is a linear subspace of* $\mathbb{R}^{m+n}$.

4. ***Remark 43*** *Notice the importance of the different characterizations of linearity and linear subspaces. Even though it would be impossible to try to picture the graph of a linear transformation* $T:\mathbb{R}^m\to\mathbb{R}^n$ *whenever* $m+n\geq 4$, *one is still able to determine that its graph is linear.*

5.1 Basis and Dimension

Notice that we have been talking about the dimension of a Euclidean space, saying that the dimension of $\mathbb{R}^n$ is n. However, we have not defined the dimension of a linear subspace of $\mathbb{R}^n$. When defining dimension, one needs to preserve the standard notions that one carries since high school. For example, a line is 1 dimensional regardless of whether the line is in the plane or in the space. Thus, we need a definition that agrees with our notion that $\mathbb{R}^n$ is n-dimensional and a line is 1-dimensional regardless of where it lives. For that, one uses the definition of a basis.

A **basis** (with plural bases) for $\mathbb{R}^n$ is an **ordered list**, which we will treat as an ordered set, of vectors in $\mathbb{R}^n$ which must be both linearly independent and span $\mathbb{R}^n$. Similarly, a basis for a linear subspace of $\mathbb{R}^n$ is a spanning set of S which is linearly independent. For any space or subspace, there are infinitely many basis. The remarkable aspect of all bases, of given a space or subspace S, is that they all have the same number of elements. Thus one defines the dimension of a space or subspace S, $dim(S)$, to be the cardinality of any given basis of S.

Remark 44 *Let* S *be a space or a subpace, and let* $\mathcal{B}=\{\vec{v}_1,\ldots,\vec{v}_m\}$ *be a basis for* S, *we will write* $S=\langle vecv_1,\ldots,\vec{v}_m\rangle$, *which will mean both* $S=Span(vecv_1,\ldots,\vec{v}_m)$ *but more than spanning* S, *the vectors* $vecv_1,\ldots,\vec{v}_m$ *are linearly independent.*

Example 43 *From the standard basis of $\mathbb{R}^n$, one finds that $dim(\mathbb{R}^n) = n$. Indeed, recall that a spaning set for $\mathbb{R}^n$ must have at least n elements, and a linearly independent set of vectors in $\mathbb{R}^n$ must have at most n elements. Thus, a basis must have n elements, and hence $dim(\mathbb{R}^n) = n$.*

Since a linear subspace S in $\mathbb{R}^n$ is the span of some vectors in $\mathbb{R}^n$, then it is natural to ask for a minimal spanning set. Thus, one can remove vectors from a spanning set of S which are linear combinations of the rest, until one obtains a minimal spanning set. This turns out to be a linearly independent spanning set, or a basis for S.

Remark 45 *Informally, a spanning set has enough vectors to fill out all directions of the space, but they can be too many. And a linearly independent set optimizes the number of directions requires to fill out the subspace. Together they guarantee that the basis has enough vectors, and optimally not too many.*

Example 44 *Consider $S = \langle \vec{e}_1, \vec{e}_2 \rangle \subseteq \mathbb{R}^3$. Although the vectors in S have 3 components, the dimension of S is not 3, since one already has a spanning set of less than 3 vectors. Since this spanning set is also linearly independent, then it is a basis, and thus $dim(S) = 2$, which is the cardinality of the spanning set. In calculus, one would call S to be the xy-plane of $\mathbb{R}^3$, and it clearly agrees that the dimension of a plane is 2, which is living in the 3-dimensional space $\mathbb{R}^3$.*

Example 45 *The exception to this rule is the subspace containing the vector zero by itself, i.e. $\{\vec{0}\} \subseteq \mathbb{R}^n$. It is a linear subspace, whose dimension is 0 and whose unique basis is the empty set, i.e. $\emptyset$.*

Example 46 *Since all subspaces contain $\vec{0}$, then $\mathbb{R}^n$ has (1) one subspace of dimension 0, say $\{\vec{0}\}$. (2) Infinitely many subspaces of dimension 1, which are all lines passing through the origin. Each one has a basis containing a single non-zero vector. (3) Actually, infinitely many subspaces of dimension i for any $1 \leq i \leq n-1$. (4) And one n-dimensional subspace, which is itself, i.e. $\mathbb{R}^n$ is a subspace of $\mathbb{R}^n$.*

Theorem 10 (Basis of a subspace) *Let S be a subspace of $\mathbb{R}^n$.*

- *If S is described as the span of some vectors $\vec{v}_1, \dots, \vec{v}_m$. Then apply Gauss elimination to the matrix $[\vec{v}_1, \dots, \vec{v}_m]$ to identify the indexes of the pivot columns, and select the vectors $\vec{v}_i$'s with these sub-indexes. They form a basis for S. Alternatively, apply Gauss-elimination or Jordan-Gauss elimination to the matrix $[\vec{v}_1, \dots, \vec{v}_m]^T$. Transpose the echelon form matrix obtained. The non-zero columns of the resulting matrix form a basis for S.*

- *If S is described as the solution to a homogeneous system $A\vec{x} = \vec{0}$. Then solve the system and write the general solution $\vec{x}_g$, which can be only the trivial solution, and hence $S = \{\vec{0}\}$ whose basis is $\emptyset$. Or $\vec{x}_g$ depends on some parameters. Hence, the set of coefficient vectors of these parameters form a basis for S.*

Remark 46 *Implicitely, any subspace depends not only in the equations that define it, but also on the Euclidean space where it lies. For example, relabel the variables, so x is the only coordinate of $\mathbb{R}$, x and y are the coordinates of $\mathbb{R}^2$, and x, y and z are the coordinates of $\mathbb{R}^3$. Then in $\mathbb{R}$, $x = 5$ is a point; in $\mathbb{R}^2$, $x = 5$ is a vertical line; in $\mathbb{R}^3$, $x = 5$ is a plane parallel to yz axis.*

Example 47 *Let $S = Span(\vec{u}_1 = (2,3,1)^T, \vec{u}_2 = (1,3,-1)^T, \vec{u}_3 = (5,9,1)^T)$. Find a basis for S and determine if the vectors $\vec{w}_1 = (1,2,0)^T$ and $\vec{w}_2 = (1,2,1)^T$ are in the subspace S.*

Solution: Let's use the three methods described in the theorem:

Method 1: Consider the matrix formed by the vectors in S, say $A = \begin{pmatrix} 2 & 1 & 5 \\ 3 & 3 & 9 \\ 1 & -1 & 1 \end{pmatrix}$. Apply Gauss-elimination to obtain its echelon form $A_E = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 3 & 3 \\ 0 & 0 & 0 \end{pmatrix}$. Notice that the pivot columns are the 1^{st} and 2^{nd} columns. Read the corresponding columns in A not in A_E, and they are a basis for S, i.e.

$$S = \left\langle \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \right\rangle$$

Method 2: Consider the matrix formed by the transpose vectors in S, i.e. write these vectors as the rows of B, say $B = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 3 & -1 \\ 5 & 9 & 1 \end{pmatrix}$. Apply Gauss-elimination to obtain its echelon form $B_E = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{pmatrix}$. Now, the transpose of the nonzero row vectors in B_E are a basis for S, i.e.

$$S = \left\langle \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -3 \end{bmatrix} \right\rangle$$

Method 3: Consider the matrix formed by the transpose vectors in S, i.e. write these vectors as the rows of B, say $B = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 3 & -1 \\ 5 & 9 & 1 \end{pmatrix}$. Apply Jordan-Gauss-elimination to obtain its reduced echelon form $B_{RE} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$. Now, the transpose of the nonzero row vectors in B_E are a basis for S, i.e.

$$S = \left\langle \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\rangle$$

Now, one can notice that the first method and the second method are equally complex. The first method gives a subset of the spanning set as a basis, yet this vectors usually have complex entries. The second method gives vectors which usually have simpler entries. Yet although the third method is the most complex to perform due to the backward phase of Jordan-Gauss elimination, it gives the simplest entries vectors. Now, if one wants find whether some vectors are in S, the first and second method are not optimal, because one needs to write the vectors $\vec{w}_i$'s as linear combinations of these complex entries basis vectors. Yet the third method is optimal because the systems to solve are the easiest, i.e. in this example one needs to solve the systems whose augmented matrices are

$$\left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & -1 & 0 \end{array}\right)$$

which has solution $(2, -1)^T$ and means that $\vec{w}_1$ is indeed in S, and

$$\left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 2 & -1 & 1 \end{array}\right)$$

which has no solution and means that $\vec{w}_2$ is not in S.

Remark 47 *Naively, one may try to form a basis for a subspace S by picking a nonzero vector at a time, which is not in the span of the previously selected vectors. However, without previously knowing the dimension of S, one wouldn't know when to terminate this selection.*

Proposition 11 *Consider the subspaces $S_1 \subseteq S_2 \subseteq \mathbb{R}^m$, notice that S_1 is contained in S_2. Then $dim(S_1) \leq dim(S_2) \leq m$. Moreover, maintaining the assumption that $S_1 \subseteq S_2$, the sub-spaces are equal if and only if the dimensions are equal, i.e. $dim(S_1) = dim(S_2)$.*

Theorem 11 (Completing a basis) *Given a linearly independent set $S = \{\vec{v}_1, \ldots, \vec{v}_m\} \subseteq \mathbb{R}^n$, consider the matrix $A = ([\vec{v}_1, \ldots, \vec{v}_m]|I_n)$. Apply Gauss elimination to obtain its echelon form A_E in order to find the indices of the pivot columns. Add the columns vectors in A which correspond to he pivot columns of A_E from a basis for $\mathbb{R}^n$ that includes our original set.*

5.2 The Row, Column and Null Spaces of a Matrix

There are three important subspaces related to a given matrix $A_{m\times n} = [veca_1, \ldots, \vec{a}_n]$, which are defined as spanning sets, and our job will be to derive algorithms to find basis for these subspaces, i.e. minimal spanning sets which are linearly independent.

- The **Row Space** $Row(A)$ of a matrix $A_{m\times n}$ is the sub-space of $\mathbb{R}^n$ spanned by the transpose of the m rows of A.
- The **Column Space** $Col(A)$ of a matrix $A_{m\times n}$ is the sub-space of $\mathbb{R}^m$ spanned by the n columns of A.
- The **Null Space** $Null(A)$ of a matrix $A_{n\times m}$ is the set of solutions to the system $A\vec{x} = \vec{0}$. The dimension of the null space is denoted $Nullity(A)$.

One uses Gauss elimination or Jordan-Gauss elimination to find the basis for these subspaces.

Example 48 *Let* $A_{3\times 4} = \begin{pmatrix} 1 & -2 & 7 & 5 \\ -2 & -1 & -9 & -7 \\ 1 & 13 & -8 & -4 \end{pmatrix}$. *In order to find bases for* $row(A)$ *and* $col(A)$,

1. *Find the Echelon form* A_E *of the Matrix* A *using Gauss-Elimination:*

$$A = \begin{pmatrix} 1 & -2 & 7 & 5 \\ -2 & -1 & -9 & -7 \\ 1 & 13 & -8 & -4 \end{pmatrix} \sim A_E = \begin{pmatrix} 1 & -2 & 7 & 5 \\ 0 & -5 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

   The echelon form A_E *has two non-zero rows, whose leading entries are in the 1st and 2nd columns respectively.*

2. *The non-zero rows of the echelon form* A_E *are a basis for* $row(A)$.

$$row(A) = \left\langle \begin{pmatrix} 1 \\ -2 \\ 7 \\ 5 \end{pmatrix}, \begin{pmatrix} 0 \\ -5 \\ 5 \\ 3 \end{pmatrix} \right\rangle$$

3. *The columns in* A *corresponding to the pivot columns of the echelon form* A_E *are a basis for* $col(A)$.

$$col(A) = \left\langle \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \\ 13 \end{pmatrix} \right\rangle$$

Example 49 *Continuing with the previous example, in order to find a basis for* $Null(A)$, *append* $\vec{0}$ *to* A_E *and find the solution of the system whose augmented matrix is* $[A_E|\vec{0}]$

$$[B|0] = \left(\begin{array}{cccc|c} 1 & -2 & 7 & 5 & 0 \\ 0 & -5 & 5 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Then $x_3 = r$ *and* $x_4 = s$ *are free variables and* $x_2 = r + \frac{3s}{5}$ *and* $x_1 = -5r - 19s/5$. *Therefore, the solution set is:*

$$\left\{ \begin{pmatrix} -5r - \frac{19}{5}s \\ r + \frac{3}{5}s \\ r \\ s \end{pmatrix} = \begin{pmatrix} -5 \\ 1 \\ 1 \\ 0 \end{pmatrix} r + \begin{pmatrix} -\frac{19}{5} \\ \frac{3}{5} \\ 0 \\ 1 \end{pmatrix} s : r, s \in \mathbb{R} \right\} = Span \left\{ \begin{pmatrix} -5 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{19}{5} \\ \frac{3}{5} \\ 0 \\ 1 \end{pmatrix} \right\}$$

Hence $Null(A) = \left\langle \begin{pmatrix} -5 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{19}{5} \\ \frac{3}{5} \\ 0 \\ 1 \end{pmatrix} \right\rangle$, *and* $nullity(A) = 2$.

Remark 48 *In the previous examples, notice that both* $row(A)$ *and* $col(A)$ *have the same dimension* 2, *which is the number of nonzero rows of the echelon matrix* A_E. *This is always the case, the dimension of both the column and row spaces are the same, which is called* ***rank*** *of the matrix* A.

Remark 49 *In the previous examples, although the dimension of* $Null(A)$ *is* 2, *which agrees with the* $rank(A)$, *this is not always the case. The nullity of a matrix* A *is equal to the number of free variables of the equivalent echelon matrix* A_E.

Proposition 12 *It is always the case that the* $rank(A) + nullity(A)$ *is equal to the number of columns of* A. *Indeed,* $rank(A)$ *is the number of pivot columns, and* $nullity(A)$ *is the number of free variables, which is the number of non-pivot columns.*

5.3 Change of Basis Matrices

So far, we have studied vectors in $\mathbb{R}^n$ written in terms of the standard basis $\langle \vec{e}_1, \dots, \vec{e}_n \rangle$. For example in $\mathbb{R}^2$, the vector $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2$, i.e. the components of $\vec{x}$ are the coefficients of the standard basis vectors when $\vec{x}$ is expressed as a linear combination of them.

Consider any basis for $\mathbb{R}^n = \langle \vec{v}_1, \dots, \vec{v}_n \rangle$, call the basis as a set $\mathcal{B} = \{\vec{v}_1, \dots, \vec{v}_n\}$. Take any vector $\vec{x} \in \mathbb{R}^n$, so $\vec{x}$ is a linear combination of the vectors $\vec{v}_i$'s. Assume $\vec{x} = \tilde{x}_1\vec{v}_1 + \ldots + \tilde{x}_n\vec{v}_n$. Then one defines the vector $\vec{x}$ in terms of the basis $\mathcal{B}$ denoted by $\vec{x}_{\mathcal{B}}$ to be the vector whose components are the coefficients $\tilde{x}_i$ of the vectors $\vec{v}_i$'s when $\vec{x}$ is expressed as a linear combination of them.

$$\vec{x}_B = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{bmatrix}_B$$

Example 50 *Call $SB = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ to the standard basis of $\mathbb{R}^3$ and $\mathcal{B} = \{(1,1,1)^T, (1,1,0)^T, (1,0,0)^T\}$ to this other basis of R^3. Then,*

$$\vec{x} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} = a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = c\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (b-c)\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + (a-b)\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \vec{x}_B = \begin{pmatrix} c \\ b-c \\ a-b \end{pmatrix}_B$$

Remark 50 *Recall that a basis is an ordered list of vectors, thus changing the order of its elements produces a different basis.*

The process of changing one vector $\vec{x}$ in terms of the standard basis to the vector $\vec{x}_{\mathcal{B}}$ in terms of another basis $\mathcal{B}$ is a linear process, i.e. there is a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n_{\mathcal{B}}$ such that for any vector $\vec{x} \in \mathbb{R}^n$, one has $T(\vec{x}) = \vec{x}_{\mathcal{B}}$. By the definition of linear transformations, there is a matrix M such that for all vectors, it satisfies $M\vec{x} = \vec{x}_{\mathcal{B}}$.

More general, given two bases $\mathcal{B}_1$ and $\mathcal{B}_2$ for $\mathbb{R}^n$, there is a matrix $M_{\mathcal{B}_1 \to \mathcal{B}_2}$ called **the change of basis matrix from basis $\mathcal{B}_1$ to $\mathcal{B}_2$** such that

$$M_{\mathcal{B}_1 \to \mathcal{B}_2}\vec{x}_{\mathcal{B}_1} = \vec{x}_{\mathcal{B}_2}$$

.

Proposition 13 *The change of basis matrices satisfy:*

- ***Reflexivity**: The change of basis matrix from one base $\mathcal{B}$ to the same base $\mathcal{B}$ is the identity matrix, i.e. $M_{\mathcal{B} \to \mathcal{B}} = I_n$.*
- ***Transitivity** The product of the change of basis matrices from basis $\mathcal{B}$ to basis $\mathcal{C}$ and that from basis $\mathcal{A}$ to basis $\mathcal{B}$ is equal to the change of basis matrix from $\mathcal{A}$ to $\mathcal{C}$, i.e. $M_{\mathcal{B} \to \mathcal{C}} M_{\mathcal{A} \to \mathcal{B}} = M_{\mathcal{A} \to \mathcal{C}}$.*
- ***Invertibility** The change of basis matrix from basis 1 to basis 2 is the inverse of the change of basis matrix from basis 2 to basis 1, i.e. $M_{\mathcal{B} \to \mathcal{A}} = M^{-1}_{\mathcal{A} \to \mathcal{B}}$.*

Remark 51 *Invertibility can be derived from the other two properties, and yet it describes a major property which is that all change of basis matrices are invertible. Under the right interpretation, all invertible matrices are change of basis matrices from the standard basis to another basis, or from another basis to the standard basis.*

Theorem 12 *Assume $\mathbb{R}^n = \langle \underbrace{\vec{v}_1, \dots, \vec{v}_n}_{\mathcal{B}_1} \rangle = \langle \underbrace{\vec{w}_1, \dots, \vec{w}_n}_{\mathcal{B}_2} \rangle$. Then*

$$M_{\mathcal{B}_1 \to \mathcal{B}_2} = \underbrace{[\vec{w}_1, \dots, \vec{w}_n]^{-1}}_{M_{SB \to \mathcal{B}_2}} \underbrace{[\vec{v}_1, \dots, \vec{v}_n]}_{M_{\mathcal{B}_1 \to SB}}$$

In other words, $M_{\mathcal{B}_1 \to \mathcal{B}_2} = [(\vec{v}_1)_{\mathcal{B}_2}, \dots, (\vec{v}_n)_{\mathcal{B}_2}]$.

Example 51 *Let $\mathcal{B}_1 = \{(1,1,1)^T, (1,1,0)^T, (1,0,0)^T\}$, which is a basis for $\mathbb{R}^3$, and let SB be the standard basis. Then*

$$M_{\mathcal{B}_1 \to SB} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

By invertibility,

$$M_{SB\to\mathcal{B}_1} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{bmatrix}$$

Let $\mathcal{B}_2 = \{(7,3,-2)^T, (0,-5,9)^T, (2,-4,4)^T\}$, *which is another basis of* $\mathbb{R}^3$. *Then*

$$M_{B_2\to SB} = \begin{pmatrix} 7 & 0 & 2 \\ 3 & -5 & -4 \\ -2 & 9 & 4 \end{pmatrix}$$

And by invertibility again,

$$M_{SB\to B_2} = M_{B_2\to SB}^{-1} = \frac{1}{146}\begin{pmatrix} 16 & 18 & 10 \\ -4 & 32 & 34 \\ 17 & -63 & -35 \end{pmatrix}$$

Finally, by the theorem of change of basis matrices,

$$M_{\mathcal{B}_1\to\mathcal{B}_2} = M_{SB\to\mathcal{B}_2} M_{\mathcal{B}_1\to SB} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \frac{1}{146}\begin{bmatrix} 16 & 18 & 10 \\ -4 & 32 & 34 \\ 17 & -63 & -35 \end{bmatrix}$$

Now, we need the hashtag *#BringBackGAUSS* from the organization *http://this.is.a.joke.com*, we found that

Theorem 13 *Assume* $\mathbb{R}^n = \langle\underbrace{\vec{v}_1, \ldots, \vec{v}_n}_{\mathcal{B}_1}\rangle = \langle\underbrace{\vec{w}_1, \ldots, \vec{w}_n}_{\mathcal{B}_2}\rangle$. *Then one can find* $M_{\mathcal{B}_1\to\mathcal{B}_2}$ *by applying Jordan-Gauss elimination to* $M_{B_2\to SB}|M_{B_1\to SB}]$, *which transforms it into* $[I|M_{B_1\to B_2}]$.

Example 52 *Following the example above,*

$$\left(\begin{array}{ccc|ccc} 7 & 0 & 2 & 1 & 1 & 1 \\ 3 & -5 & -4 & 1 & 1 & 0 \\ -2 & 9 & 4 & 1 & 0 & 0 \end{array}\right) \xrightarrow{Jordan-Gauss} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{22}{73} & \frac{17}{73} & \frac{8}{73} \\ 0 & 1 & 0 & \frac{31}{73} & \frac{14}{73} & -\frac{2}{73} \\ 0 & 0 & 1 & -\frac{81}{146} & -\frac{23}{73} & \frac{17}{146} \end{array}\right) \Rightarrow M_{B_1\to B_2} = \frac{1}{73}\begin{bmatrix} 22 & 17 & 8 \\ 31 & 14 & -2 \\ -81 & -23 & 17 \end{bmatrix}$$

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 7 & 0 & 2 \\ 1 & 1 & 0 & 3 & -5 & -4 \\ 1 & 0 & 0 & -2 & 9 & 4 \end{array}\right) \xrightarrow{Jordan-Gauss} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 9 & 4 \\ 0 & 1 & 0 & 5 & -14 & -8 \\ 0 & 0 & 1 & 4 & 5 & 6 \end{array}\right) \Rightarrow M_{B_2\to B_1} = \begin{bmatrix} -2 & 0 & 4 \\ 5 & -14 & -8 \\ 4 & 5 & 6 \end{bmatrix}$$

5.4 Change of Basis of Subspaces

So far we found how to change a vector $\vec{x} \in \mathbb{R}^n$ to its representation in other bases of $\mathbb{R}^n$. Now, we want to do the same for vectors in a subspace S of $\mathbb{R}^n$ and bases of S. Assume $S = \langle\underbrace{\vec{v}_1, \ldots, \vec{v}_m}_{\mathcal{B}]_\infty}\rangle$ where $dim(S) = m < n$. Any vector $\vec{x}$ in S is a linear combination of the vectors in $\mathcal{B}_1$ because this is a basis for S, i.e. a spanning set for S. Hence, $\vec{x} = \tilde{x}_1\vec{v}_1 + \ldots + \tilde{x}_m\vec{v}_m$. So we define the vector $\vec{x}_{\mathcal{B}_1}$ as

$$\vec{x}_{B_1} = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_m \end{bmatrix}_{B_1}$$

Remark 52 *Notice that* $\vec{x} \in \mathbb{R}^n$ *is a vector with* n *components or entries. However, when one represents* $\vec{x}$ *in terms of a basis* $\mathcal{B}_1$ *of the subspace* S *of dimension* $dim(S) = m < n$, $\vec{x}_{\mathcal{B}_1}$ *only has* m *components. Therefore, the change of basis matrix* $M_{SB\to\mathcal{B}_1}$ *from* $\vec{x}$ *to* $\vec{x}_{\mathcal{B}_1}$ *has dimensions* $m \times n$. *Indeed,* $M_{\mathcal{B}_1\to SB} = [\vec{v}_1, \ldots, \vec{v}_m]_{n\times m}$.

Now, given two bases $S = \langle\underbrace{\vec{v}_1, \ldots, \vec{v}_m}_{\mathcal{B}_1}\rangle = \langle\underbrace{\vec{w}_1, \ldots, \vec{w}_m}_{\mathcal{B}_2}\rangle$. We want to find the change of basis matrix $M_{\mathcal{B}_1\to\mathcal{B}_2}$ such that for any vector $\vec{x} \in S$, it satisfies

$$M_{\mathcal{B}_1\to\mathcal{B}_2}\vec{x}_{\mathcal{B}_1} = \vec{x}_{\mathcal{B}_2}$$

Remark 53 *The matrix to be obtained is an* $m \times m$ *matrix and it only changes representations of vectors in* S. *So if one applies it to a vector in* $\mathbb{R}^n$ *but not in* S, *it would be meaningless because such vector does not have a representation in any of these bases.*

Example 53 *Le $S \subseteq \mathbb{R}^3$ be a subspace with bases $B_1 = \{(1,-5,8)^T, (3,-8,3)^T\}$ and $B_2 = \{(1,-3,2)^T, (-1,2,1)^T\}$. Find the change of basis matrix $M_{B_1 \to B_2}$. Naively, one would express the vectors in B_1 as linear combinations of B_2:*

$$\begin{array}{ccccccc} (1,-5,8)^T & = & 3 & (1,-3,2)^T & + & 2 & (-1,2,1)^T \\ (3,-8,3)^T & = & 2 & (1,-3,2)^T & + & (-1) & (-1,2,1)^T \end{array} \Rightarrow M_{B_1 \to B_2} = \begin{bmatrix} 3 & 2 \\ 2 & -1 \end{bmatrix}$$

However, this method requires to solve a system for each vector in B_1 in order to write it as a linear combination of the vectors in B_2.

Theorem 14 *Let $\mathcal{B}_1 = \{\vec{v}_1, \ldots, \vec{v}_d\}$ and $\mathcal{B}_2 = \{\vec{w}_1, \ldots, \vec{w}_d\}$ be two basis for a subspace S of $\mathbb{R}^n$. Define the matrices $B_1 = [\vec{v}_1, \ldots, \vec{v}_d]$ and $B_2 = [\vec{w}_1, \ldots, \vec{w}_d]$, and append B_1 to B_2 to obtain $[B_2|B_1]$. The reduced echelon form of this last matrix is of the form*

$$\begin{pmatrix} I_d & M_{B_1 \to B_2} \\ 0_{n-d \times d} & 0_{n-d \times n-d} \end{pmatrix}$$

Example 54 *Continuing the previous example,*

$$[B_2|B_1] = \left(\begin{array}{rr|rr} 1 & -1 & 1 & 3 \\ -3 & 2 & -5 & -8 \\ 2 & 1 & 8 & 3 \end{array}\right) \xrightarrow{Jordan\ Gauss} \left(\begin{array}{rr|rr} 1 & 0 & 3 & 2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow M_{B_1 \to B_2} = \begin{pmatrix} 3 & 2 \\ 2 & -1 \end{pmatrix}$$

5.5 Worksheet - Subspaces

1. Find a basis for the column space, for the row space and for the null space of the following matrix.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 1 & 3 & 5 & 7 \\ 1 & -1 & 1 & -1 \end{pmatrix}$$

2. Expand the set $\{(-1,0,1,0)^T, (0,-1,0,1)^T\}$ to be a basis of the subspace $S = \{(a,b,c,d) \in \mathbb{R}^4 : a+b+c+d = 0\}$.

3. Find an example of four vectors $u_1, u_2, u_3, u_4 \in \mathbb{R}^2$ such that if we remove any two of them, the remaining 2 are a basis.

4. Find the simplest basis for $S = Span((1,2,3,4)^T, (1,2,1,2)^T, (-1,2,1,-2)^T, (1,1,1,1)^T, (5,6,7,8)^T)$. What is the dimension of S?

5. Given a linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ and its coefficient matrix $A = A_T$.
   (a) Find the dimensions of A.
   (b) Could T be injective?
   (c) If T is surjective, then what are $rank(A)$ and $nullity(A)$?

6. Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ described by $T_{2\times 3}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 4x - 3y + z \\ 2y - 5z \end{pmatrix}$. And consider the linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ with matrix $B = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$.
   - Find the matrix A that represents T.
   - Is T injective? Is T surjective? Is T invertible?
   - Find $Ker(T)$ and $Null(A)$.
   - Find $Range(T)$ and $Col(A)$.
   - Is S injective? Is S surjective? Is S invertible? If yes, can you find S^{-1}? Do it in two ways described in lecture!
   - What is the matrix C associated to $S^{-1}T$?
   - Is $S^{-1}T$ injective? Is $S^{-1}T$ surjective? Is $S^{-1}T$ invertible?

7. True and False

   - Given two equivalent matrices (one can be transform to the other by elementary operations) $A_{m\times n}$ and $B_{m\times n}$, both of them without zero rows; if the rows of A are linearly dependent, then so are the rows of B.
   - If S_1 and S_2 are subspaces of $\mathbb{R}^n$ of the same dimension, then $S_1 = S_2$.
   - The union (as sets) of two sub-spaces of $\mathbb{R}^n$ is always a subspace of $\mathbb{R}^n$.
   - The intersection (as sets) of two sub-spaces of $\mathbb{R}^n$ is always a subspace of $\mathbb{R}^n$.
   - Given a vector $b_{n\times 1} \neq \vec{0}$ and a matrix $A_{n\times m}$. The set of solutions to the system $Ax = b$ is not a subspace of $\mathbb{R}^m$.
   - Let S_1 and S_2 be sub-spaces of $\mathbb{R}^n$, and define S to be the set of all vectors of the form $s_1 + s_2$ where s_1 is in S_1 and s_2 is in S_2. Then S is a subspace of $\mathbb{R}^n$.

6 The Unifying Theorem

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation with associated or coefficient matrix $A = [\vec{a_1}, \ldots, \vec{a}_m]_{n\times m}$, where the column vectors $\vec{a}_1, \ldots, \vec{a}_m \in \mathbb{R}^n$.

1. In terms of the dimensions of the matrix: $Rank(A) + Nullity(A) = m$. In terms of the linear transformation $dim(Range(T)) + dim(Ker(T)) = dim(Domain(T)) = m$.

2. (1) $Nullity(A) = dim(Ker(T)) \leq m$, $Rank(A) = dim(Range(T)) \leq \min(m, n)$. Moreover, $Ker(T) = E_0$ (the eigenspace for $\lambda = 0$) if it exists, or $Ker(T) = \{\vec{0}\}$ if it does not.

3. **Injective Equivalences** or One-To-One:

   - $T(\vec{x}) = T(\vec{y}) \Rightarrow \vec{x} = \vec{y}$.
   - $Null(A) = Ker(T) = \{\vec{0}\}$ or $Nullity(A) = dim(Ker(T)) = 0$.
   - $\{\vec{a}_1, \ldots, \vec{a}_m\}$ is linearly independent.
   - All systems $A\vec{x} = \vec{b}$ (for any $\vec{b}$) have at most one solution.
   - 0 is not an eigenvector of A.
   - T has a surjective left inverse $T_L^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ such that $T_L^{-1} \circ T = I_m : \mathbb{R}^m \to \mathbb{R}^m$ where I_m is the identity transformation. Also A has a left inverse A_L^{-1} such that $A_L^{-1}A = I_m$.

   How to test: Check that $m \leq n$, otherwise it is not injective. Then apply Gauss Elimination to A and check that the number of nonzero rows is equal to the number of columns.

4. **Surjective Equivalences** or Onto:

   - $Range(T) = Codomain(T)$ or $dim(Range(T)) = dim(Codomain(T)) = n$.
   - $Col(A) = Range(T) = Span(\vec{a}_1, \ldots, \vec{a}_n) = Codomain(T) = \mathbb{R}^n$ or $Rank(A) = dim(Range(T)) = n$.
   - All systems $A\vec{x} = \vec{b}$ (for any $\vec{b}$) have at least one solution.
   - T has an injective right inverse $T_R^{-1} : \mathbb{R}^n \to \mathbb{R}^m$ such that $T \circ T_R^{-1} = I_n : \mathbb{R}^n \to \mathbb{R}^n$ where I_n is the identity transformation. Also A has a right inverse A_R^{-1} such that $AA_R^{-1} = I_n$.

   How to test: Check first that $m \geq n$, otherwise it is not surjective. Apply Gauss elimination to A (without any constant vector,) and T is surjective if and only if there are no zero rows.

5. **Bijective Equivalence**:

   - Enough to prove injective/one-to-one and $m = n$.
   - Enough to prove surjective/onto and $m = n$.
   - All systems $A\vec{x} = \vec{b}$ (for any $\vec{b}$) have exactly one solution.
   - A is invertible or $\det(A) \neq 0$.

6.1 Examples

Example 55 *Let $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \subseteq \mathbb{R}^3$ be a linearly independent set and let $A_{3\times 3}$ be invertible. Show that $\{A\vec{v}_1, A\vec{v}_2, A\vec{v}_3\}$ is also linearly independent.*

Answer: Indeed, by the number of vectors and the equivalence of bijectivity, this set is a basis for $\mathbb{R}^3$, and thus $\det([\vec{v}_1, \vec{v}_2, \vec{v}_3]) \neq 0$. Then notice that $\det([A\vec{v}_1, A\vec{v}_2, A\vec{v}_3]) = \det(A)\det([\vec{v}_1, \vec{v}_2, \vec{v}_3]) \neq 0$, and thus $\{A\vec{v}_1, A\vec{v}_2, A\vec{v}_3\}$ is also a basis of $\mathbb{R}^3$, and in particular it is linearly independent. For an easier example, one knows that $\{\vec{v}_1, \vec{v}_1 + \vec{v}_2, \vec{v}_1 + \vec{v}_2 + \vec{v}_3\}$ is linearly independent because

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

Example 56 *Let A be the following 2×2 matrix. Let $T : \mathbb{R}^2 \to \mathbb{R}$ be the function whose rule is $T(\vec{x}) = \vec{x}^T A\vec{x}$. Is T a linear transformation?*

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

Solution: Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Then

$$T(\vec{x}) = [x_1, x_2] \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + 5x_1x_2 + 4x_2^2$$

Hence, T is not a linear transformation because the image is not a linear function but a quadratic function. Moreover, the rule for all linear transformations is $A\vec{x}$ and T does not have such rule, it has an extra left factor $\vec{x}^T$.

Example 57 *Let P be the plane in $\mathbb{R}^3$ passing through the points $\vec{v}_1 = (0, 0, 0)^T, \vec{v}_2 = (1, 1, 1)^T$ and $\vec{v}_3 = (3, -2, 1)^T$. Find the scalar $a \in \mathbb{R}$ such that the one parameter solution set for the intersection of the plane P with the xy-plane is $\{(as, s, 0)^T : s \in \mathbb{R}\}$.*

Solution: One knows that a plane satisfies an equation of the form $ax + by + cz = d$. Let's assume $a \neq 0$, so by dividing by a, P satisfies an equation of the form $x + by + cz = d$. By plugging the given points of P, one knows that (1) $d = 0$, (2) $1 + b + c = 0$, and (3) $3 - 2b + c = 0$. Then $b = 2/3$ and $c = -5/3$. Hence, $P : x + 2y/3 - 5z/3 = 0$ or $3x + 2y - 5z = 0$. On the other hand, the equation of the xy-plane is $z = 0$. The intersection of these planes is the solution of the system whose augmented matrix is $\left(\begin{array}{ccc|c} 3 & 2 & -5 & 0 \\ 0 & 0 & 1|0 & \end{array}\right) \sim \left(\begin{array}{ccc|c} 3 & 2 & 0 & 0 \\ 0 & 0 & 1|0 & \end{array}\right)$, which is already in echelon form. Then x and z are leading variables, and y is a free variable. Notice that the solution set is $\{(-2/3y, y, 0) : y \in \mathbb{R}\}$. However, one can relabel y to be a parameter s, and by comparison, $a = -2/3$.

7 Determinants

7.1 Definitions and Overviews

The determinant of a square matrix is a number associated to it. Non-square matrices don't have a determinant. So you can consider the determinant as a function from the set of all square matrices to the real numbers. Notice however, that if a matrix has complex entries, then its determinant is a complex number, but we won't be dealing with this in this class.

Example 58 (Determinants of 1-dim matrices) $\det(a) = a$, *i.e. the determinant is the number obtained by just removing the parenthesis.*

Example 59 (Determinants of 2-dim matrices) $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$, *i.e. the determinant is the product of the main diagonal minus the product of the other diagonal.*

Example 60 (Determinants for 3-dim matrices) $\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - afh - bdi - ceg$. *An easy way to remember this is to extend the matrix by copying the first and second columns on the right, then the determinant* *is the sum of the three products of the up-left-to-down-right diagonals entries minus the sum of the three products of the up-right-to-down-left diagonals entries.*

$$\left(\begin{array}{ccc|cc} a & b & c & a & b \\ d & e & f & d & e \\ g & h & i & g & h \end{array}\right)$$

Remark 54 *The formal definition of the determinant involves the symmetric group of n-objects S_n or the group of n-permutations. Each permutation $\sigma \in S_n$ has a sign, positive or negative. Then*

$$\det(a_{ij})_{n\times n} = \sum_{\sigma\in S_n} Sign(\sigma)\prod_{i=1}^{n} a_{i\sigma(i)} = \sum_{\sigma\in S_n} Sign(\sigma)\prod_{i=j}^{n} a_{\sigma(j)j}$$

In other words, every term of the resulting polynomial on the entries is the signed product of n entries, all in different columns and different rows, where n is the dimension of the square matrix given.

Remark 55 *The determinant of an n-dimensional matrix is the sum of $n!$ terms, and each term is the signed product of n entries. Thus, writing a formula for the determinant of an $n \geq 4$ dimensional matrix will be exhausting. Indeed, for $n = 4$, the formula would have $4! = 24$ terms each one being the multiple of 4 entries of the matrix; for $n = 5$, the formula would have $5! = 120$ terms each one being the multiple of 5 entries of the matrix; etc. Thus, rather than writing long formulas, one uses a recursive formula to compute determinants.*

Given a matrix A, let $A^{(i,j)}$ be the matrix obtained by deleting the i-th row and the j-th column. Then the i,j-th **Minor** of the matrix A is the determinant of $A^{(i,j)}$, i.e. $M_{ij}(A) = \det(A^{(i,j)})$. Thus the minor is a number and not a submatrix. Then the i,j-th **Cofactor** of the matrix A is the signed minor, $C_{ij}(A) = (-1)^{i+j}\det(A^{(i,j)})$. Thus again, the cofactor is a number and not a submatrix.

Remark 56 *There is one minor and one cofactor for each entry in a matrix. Indeed, one tends to identify minors and cofactors not with the i,j-position but with the i,j-th entry of the matrix.*

In order to compute the determinant of a matrix, one would need to compute the cofactors corresponding to all positions in a row, or to all positions in a column. The **Cofactor expansion of the Determinant** on the i-th row is:

$$\begin{array}{rclcc} \det(A_{n\times n}) & = & a_{i1}C_{i,1}(A) & +\ldots+ & a_{i,n}C_{in}(A) \\ & = & (-1)^{i+1}a_{i,1}\det(A^{i,1}) & +\ldots+ & (-1)^{i+n}a_{i,n}\det(A^{i,n}) \end{array}$$

The Cofactor expansion of the Determinant on the j-th column is:

$$\begin{array}{rclcc} \det(A_{n\times n}) & = & a_{1,j}C_{1,j}(A) & +\ldots+ & a_{n,j}C_{n,j}(A) \\ & = & (-1)^{1+j}a_{1,j}\det(A^{1,j}) & +\ldots+ & (-1)^{n+j}a_{n,j}\det(A^{n,j}) \end{array}$$

Remark 57 *Since each term depends on both the cofactors and the entries of the matrix, then one should try to identify the row or column with the largest number of zero entries, thus one knows that the corresponding terms are zero without even computing the determinant of the submatrix factor. Hence, one would have to compute the smallest number of determinant of submatrices.*

Example 61 *The cofactor expansion of the determinant on the 3rd column is:*

$$\det\begin{pmatrix} 1 & 2 & 3 & 4 \\ -1 & 0 & 0 & 5 \\ 7 & 7 & 1 & 1 \\ 1 & -2 & 0 & 3 \end{pmatrix} = (-1)^{1+3}3\det\begin{pmatrix} -1 & 0 & 5 \\ 7 & 7 & 1 \\ 1 & -2 & 3 \end{pmatrix} + (-1)^{3+3}1\det\begin{pmatrix} 1 & 2 & 4 \\ -1 & 0 & 5 \\ 1 & -2 & 3 \end{pmatrix} = ...$$

Alternatively, the cofactor expansion of the determinant on the 2nd row is:

$$\det\begin{pmatrix} 1 & 2 & 3 & 4 \\ -1 & 0 & 0 & 5 \\ 7 & 7 & 1 & 1 \\ 1 & -2 & 0 & 3 \end{pmatrix} = (-1)^{2+1}(-1)\det\begin{pmatrix} 2 & 3 & 4 \\ 7 & 1 & 1 \\ -2 & 0 & 3 \end{pmatrix} + (-1)^{2+4}(5)\det\begin{pmatrix} 1 & 2 & 3 \\ 7 & 7 & 1 \\ 1 & -2 & 0 \end{pmatrix} = ...$$

If you finish these computations, they must give you the same value of the determinant.

Proposition 14

- *The determinant of the identity is always 1, i.e. $\det(I_n) = 1$.*

- *The determinant of a diagonal or triangular matrix is the product of the diagonal entries.*
- *The determinant of the transpose of a matrix is the same as the determinant of the matrix, i.e.* $\det(A^T) = \det(A)$.
- *The determinant of a matrix with a zero-row or a zero-column is zero.*
- *The determinant of a matrix with two equal rows or two equal columns is zero.*
- *The product formula is* $\det(AB) = \det(A)\det(B)$.

Proposition 15 *A matrix is invertible or non-singular if and only if the determinant is nonzero. Hence, if the determinant is zero, the matrix is not invertible, i.e. it is singular. Moreover, if the matrix A is invertible, then* $\det(A^{-1}) = \frac{1}{\det(A)}$. *However, the determinant only says whether A is invertible or not, but it does not says what the inverse is.*

Although one would hope that equivalent matrices have the same determinant, this is not the case. Row elementary operations change the determinant of a matrix, but luckily, they change the determinant in a tractable way. **Exchanging** rows changes the determinant by a factor -1, i.e. if B is obtained from A by exchanging two rows, then $\det(B) = -\det(A)$ or $\det(A) = -\det(B)$.

Example 62

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = -\det\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$$

Multiplying a row by a scalar c changes the determinant by the same factor c, i.e. if B is obtained from A by multiplying one row by c, then $\det(B) = c\det(A)$ or $\det(A) = \frac{1}{c}\det(B)$.

Example 63

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \frac{1}{5}\det\begin{pmatrix} 5 & 10 \\ 3 & 4 \end{pmatrix}$$

Adding a multiple of a row to another row does not change the determinant, i.e. if B is obtained from A by adding a constant multiple of one row to another, then $\det(B) = \det(A)$.

Example 64

$$\det\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \det\begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}$$

Thus, if one keeps track of these factors, one can compute the determinant of a matrix from the determinant of its echelon form. Thus, since the factors appearing while performing row elementary operations are nonzero, two equivalent matrices are either both invertible or both non-invertible.

Example 65 *In order to compute the following determinant, one can use the cofactor expansion:*

$$\det\begin{pmatrix} 0 & a & 0 & 0 \\ -a & 0 & b & 0 \\ 0 & -b & 0 & c \\ 0 & 0 & -c & 0 \end{pmatrix} =_{row\,1} (-1)^{1+2} a \det\begin{pmatrix} -a & b & 0 \\ 0 & 0 & c \\ 0 & -c & 0 \end{pmatrix} =_{col\,1} (-a)(-1)^{1+1}(-a)\det\begin{pmatrix} 0 & c \\ -c & 0 \end{pmatrix} = a^2c^2$$

On the other hand, one can move the first row to last, which entails to exchange row 1 and row 2, then row 2 and row 3, then row 3 and row 4, so it changes the determinant by three factors -1. *Then add* a/b *times row 2 to row 4, which does not change the determinant. Hence,*

$$\det\begin{pmatrix} 0 & a & 0 & 0 \\ -a & 0 & b & 0 \\ 0 & -b & 0 & c \\ 0 & 0 & -c & 0 \end{pmatrix} = -\det\begin{pmatrix} -a & 0 & b & 0 \\ 0 & -b & 0 & c \\ 0 & 0 & -c & 0 \\ 0 & 0 & 0 & \frac{ac}{b} \end{pmatrix} = -(-a)(-b)(-c)(ac/b) = a^2c^2$$

because the echelon form of a square matrix is always upper triangular, and also the determinant of a triangular matrix is the product of its diagonal entries.

7.2 The Use of the determinant for finding Inverse Matrices

Recall that we already know how to compute the inverse of a matrix. Indeed, say one wants to find whether A is invertible and what the inverse A^{-1} is, then one applies Jordan-Gauss elimination to $[A|I]$ to obtain $[I|A^{-1}]$ if A is invertible, or a different left half than the identity if A is non-invertible. However, if one makes a mistake in the middle of the computation, several entries will be wrong. Fortunately, one may use determinants and coffactors to compute the inverse of a matrix.

First, the **Cofactor Matrix** $C(A)_{n\times n} = (C_{i,j}(A))$ of a matrix $A_{n\times n}$ is the matrix whose i,j-th entry is the i,j-th coffactor of A. Then the **Adjoint Matrix** $Adj(A) = C(A)^T$ is the transpose of the cofactor matrix of A.

Theorem 15 *For any square matrix $A_{n\times n}$, one has $Adj(A)\cdot A = \det(A)I_n$. Thus, one may compute the determinant of A to check if A is invertible. After that, if $\det(A) \neq 0$, then A is invertible and by the uniqueness property of the inverse, $A^{-1} = \frac{1}{\det(A)}Adj(A)$.*

Example 66 *Let* $A = \begin{pmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & 8 & -9 \end{pmatrix} \Rightarrow \det(A) = -282$. *Then one computes the nine cofactors corresponding to the nine entries of A.*

- *The $1,1$-Cofactor of A is* $C_{1,1}(A) = (-1)^{1+1}\det\begin{bmatrix} 5 & 6 \\ 8 & -9 \end{bmatrix} = -93.$
- *The $1,2$-Cofactor of A is* $C_{1,2}(A) = (-1)^{1+2}\det\begin{bmatrix} -4 & 6 \\ 7 & -9 \end{bmatrix} = 6.$
- *The $1,3$-Cofactor of A is* $C_{1,3}(A) = (-1)^{1+3}\det\begin{bmatrix} -4 & 5 \\ 7 & 8 \end{bmatrix} = -67.$
- *The $2,1$-Cofactor of A is* $C_{2,1}(A) = (-1)^{2+1}\det\begin{bmatrix} 2 & 3 \\ 8 & -9 \end{bmatrix} = 42.$
- *The $2,2$-Cofactor of A is* $C_{2,2}(A) = (-1)^{2+2}\det\begin{bmatrix} 1 & 3 \\ 7 & -9 \end{bmatrix} = -30.$
- *The $2,3$-Cofactor of A is* $C_{2,3}(A) = (-1)^{2+3}\det\begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix} = 6.$
- *The $3,1$-Cofactor of A is* $C_{3,1}(A) = (-1)^{3+1}\det\begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix} = -3.$
- *The $3,2$-Cofactor of A is* $C_{3,2}(A) = (-1)^{3+2}\det\begin{bmatrix} 1 & 3 \\ -4 & 6 \end{bmatrix} = -18.$
- *The $3,3$-Cofactor of A is* $C_{3,3}(A) = (-1)^{3+3}\det\begin{bmatrix} 1 & 2 \\ -4 & 5 \end{bmatrix} = 13.$

Then, the ***Cofactor Matrix*** *of A is:*

$$C(A) = \begin{bmatrix} C_{1,1}(A) & C_{1,2}(A) & C_{1,3}(A) \\ C_{2,1}(A) & C_{2,2}(A) & C_{2,3}(A) \\ C_{3,1}(A) & C_{3,2}(A) & C_{3,3}(A) \end{bmatrix} = \begin{bmatrix} -93 & 6 & -67 \\ 42 & -30 & 6 \\ -3 & -18 & 13 \end{bmatrix}$$

Then, the ***Adjoint Matrix*** *is the transpose of the cofactor matrix:*

$$Adj(A) = C(A)^T = \begin{pmatrix} -93 & 42 & -3 \\ 6 & -30 & -18 \\ -67 & 6 & 13 \end{pmatrix}$$

Since $\det(A) = -282 \neq 0$, *one finds that:*

$$A^{-1} = \frac{1}{-282}\begin{pmatrix} -93 & 42 & -3 \\ 6 & -30 & -18 \\ -67 & 6 & 13 \end{pmatrix} = \begin{pmatrix} \frac{31}{94} & -\frac{7}{47} & \frac{1}{94} \\ -\frac{1}{47} & \frac{5}{47} & \frac{3}{47} \\ \frac{67}{282} & -\frac{1}{47} & -\frac{13}{282} \end{pmatrix}$$

Remark 58 *For square matrices of dimension 2, one has an easy formula to find both the determinants and the inverses. For square matrices of dimension 3, one has an easy to memorize formula to find the determinant, and it is not that bad to compute the adjoint and thus the inverse of the matrix. However, for larger dimensions, we now have two methods to compute the inverses: Either applying Jordan-Gauss elimination to $[A|I_n]$ hoping to obtain $[I_n|A^{-1}]$. Or find the determinant and the adjoint matrix to A, and $A^{-1} = \frac{1}{\det(A)} Adj(A)$.*

The first method is usually faster, but one error leads to the error of several entries. The second method is usually slower unless one can compute products really fast, but it computes the entries of the Adjoint matrix independently of each other. In this class, one may choose any method.

7.3 Some Special Partitioned Matrices

Theorem 16 *Assume $A_{2n\times 2n} = \begin{pmatrix} B_{n\times n} & C_{n\times n} \\ D_{n\times n} & E_{n\times n} \end{pmatrix}$. If any of the matrices B, C, D, E is the zero matrix, then* $\det(A) = \det(B)\det(E) - \det(C)\det(D)$. *However, if none of the matrices B, C, D, E is the zero matrix, one cannot conclude whether this is true or not.*

Theorem 17 *Let $A_{2n\times 2n} = \begin{pmatrix} B_{n\times n} & C_{n\times n} \\ D_{n\times n} & E_{n\times n} \end{pmatrix}$.*

1. *If $D = 0_{n\times n}$, and B and E are invertible, then A is invertible and $A^{-1} = \begin{pmatrix} B^{-1} & -B^{-1}CE^{-1} \\ 0 & E^{-1} \end{pmatrix}$.*

2. *If $E = 0_{n\times n}$, and C and D are invertible, then A is invertible and $A^{-1} = \begin{pmatrix} 0 & D^{-1} \\ C^{-1} & -C^{-1}BD^{-1} \end{pmatrix}$.*

Example 67 *Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 0 & 0 & 9 & 10 \\ 0 & 0 & 11 & 12 \end{bmatrix}$. Find its determinant and its inverse. In order to find its determinant,*

- *One could use cofactor expansion on the first column for example:*

$$\det(A) = 1\det\begin{bmatrix} 6 & 7 & 8 \\ 0 & 9 & 10 \\ 0 & 11 & 12 \end{bmatrix} - 5\det\begin{bmatrix} 2 & 3 & 4 \\ 0 & 9 & 10 \\ 0 & 11 & 12 \end{bmatrix} = 8$$

- *One could apply Gauss-elimination to reduce it to an upper triangular matrix, which can be done only using adding constant multiples of a row to another row, and thus the determinant is the same. Hence,*

$$A \sim \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -8 \\ 0 & 0 & 9 & 10 \\ 0 & 0 & 0 & -2/9 \end{pmatrix} \Rightarrow \det(A) = (1)(-4)(9)(-2/9) = 8$$

- *One could use the fact that it is an upper block-triangular matrix, and thus*

$$\det\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 0 & 0 & 9 & 10 \\ 0 & 0 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 9 & 10 \\ 11 & 12 \end{pmatrix} = (-4)(-2) = 8$$

In order to compute its inverse, one could use Jordan-Gauss elimination in $[A|I_4]$, or compute the cofactor and adjoint matrices to A now that one knows that A is invertible since its determinant is not zero. However, we want to exemplify the last theorems. Since A is a block upper triangular matrix, then notice that $\begin{pmatrix} 1 & 2 \\ 5 & 6 \end{pmatrix}^{-1} = \frac{-1}{4}\begin{pmatrix} 6 & -2 \\ -5 & 1 \end{pmatrix}$ and $\begin{pmatrix} 9 & 10 \\ 11 & 12 \end{pmatrix}^{-1} = \frac{-1}{2}\begin{pmatrix} 12 & -10 \\ -11 & 9 \end{pmatrix}$. Finally, applying the last theorem, one obtains:

$$A^{-1} = \begin{pmatrix} -3/2 & 1/2 & 5 & -4 \\ 5/4 & -1/4 & -9/2 & 7/2 \\ 0 & 0 & -6 & 5 \\ 0 & 0 & 11/2 & -9/2 \end{pmatrix}$$

Remark 59 *Notice that the matrix A has all integer entries. Yet the inverse has rational entries. But the common denominator of the entries in the inverse is 8, which is the absolute value of the determinant of A. This comes from the factor $\frac{1}{\det(A)}$ in the formula $A = \frac{1}{\det(A)} Adj(A)$.*

8 Eigen-stuff

Recall that linear transformations include rotations, dilations-contractions, reflections and projections, and composition of these transformations. Under a dilation-contractions and reflections, some vectors mapped to a scalar multiple of itself. Under rotations, for example, if a vector belongs to the axis of rotation, it maps to itself. Hence, given a matrix, which corresponds to a unique linear transformation, one would be interested in vectors that map to a scalar multiple of themselves, i.e. multiplying by the matrix does not change their direction. However, notice that $A\vec{0} = \vec{0}$ regardless of the matrix A, i.e. this solution is trivial and does not contribute any advance.

Example 68

$$\begin{pmatrix} 3 & 5 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \end{pmatrix} = 7 \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

So given a matrix A, one wants to find all nonzero vectors, called eigenvectors $\vec{x} \neq \vec{0}$, such that $A\vec{x} = \lambda\vec{x}$, where one calls λ the corresponding eigenvalue: $A \underbrace{\vec{x}}_{\text{Eigenvector}} = \overbrace{\lambda}^{\text{Eigenvalue}} \underbrace{\vec{x}}_{\text{Eigenvector}}$.

Remark 60 *Notice that the eigenvectors are nonzero, i.e. $\vec{x} \neq \vec{0}$, but the eigenvalues can be zero.*

Since one wants to solve $A\vec{x} = \lambda\vec{x}$, first one rewrites it as $A\vec{x} - \lambda\vec{x} = \vec{0}$ or equivalently $(A - \lambda I)\vec{x} = \vec{0}$, and hence $\vec{x} \in Null(A - \lambda I)$ for some λ. However, since $\vec{x} \neq \vec{0}$, one wants this null space to be non-trivial, i.e. $dim(Null(A - \lambda I)) > 0$. By the unifying theorem, this means $A - \lambda I$ is non-invertible or $\det(A - \lambda I) = 0$. Hence, the eigenvalues of A are the roots of the characteristic polynomial $\det(A - \lambda I)$ of the matrix. This polynomial has degree n where n is the dimension of the matrix; also the leading coefficient is $(-1)^n$, and the constant term is $\det(A)$.

Notice that one needs to solve a polynomial equation in order to find the eigenvalues of a square matrix. Recall that the solution to $ax + b = 0$ is $x = -b/a$, and the solution to $ax^2 + bx + c = 0$ is $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$. However, one would need other tools to solve larger degree polynomial equations. One useful tool to solve them is the **Rational Root Theorem**, which states that if $P(x) = a_n x^n + \ldots + a_1 x + a_0$ is a polynomial with integral coefficients, i.e. all coefficients $a_0, a_1, \ldots, a_n$ are integers, then all the rational roots of this polynomial have divisor of a_0 as the numerator and a divisor of a_n as the denominator, i.e. all of them are of the form $\pm n/d$ where n is a divisor of a_0 (the constant term) and d is a divisor of a_n (the leading coefficient). Recall that the leading coefficient of the characteristic polynomial is $(-1)^n$ and the constant term is $\det(A)$. Hence, the rational eigenvalues of a matrix with integer entries are integers dividing the determinant of the matrix. Moreover, a polynomial with real coefficients decompose into the product of some linear polynomial factors and some irreducible quadratic polynomials whose roots are complex.

Example 69 *Let*

$$A = \begin{pmatrix} 4 & 4 & -2 \\ 1 & 4 & -1 \\ 3 & 6 & -1 \end{pmatrix}$$

Then its characteristic polynomial is

$$\det(A - xI) = \begin{pmatrix} 4 - x & 4 & -2 \\ 1 & 4 - x & -1 \\ 3 & 6 & -1 - x \end{pmatrix} = -x^3 + 8x - 21x + 18 = -(x - 3)(x - 2)^2$$

Then the eigenvalues are $\lambda = 2, 3$. Notice that 2 is a repeated root of this polynomial. One says that the root $\lambda = 2$ has algebraic multiplicity 2, which can be seen by the factor $(x - 2)^2$; and the root $\lambda = 3$ has algebraic multiplicity 1. In general, the multiplicity of a root λ is the exponent of the factor $(x - \lambda)$ in the decomposition of the characteristic polynomial.

Now, notice that all vectors $\vec{x}$ in $Null(A - \lambda I)$ satisfy the equation of $A\vec{x} = \lambda\vec{x}$. One wants all the vectors in $Null(A - \lambda I)$ but the zero vector, which by definition is not an eigenvector. In other words, the set of eigenvectors with a common eigenvalue λ are almost a subspace, such set would only be lacking the zero vector $\vec{0}$. Let $E_\lambda = Null(A - \lambda I)$, which is called the **eigenspace associated to** λ. Hence, in order to find the eigenspace associated to λ, one must solve the sytem $(A - \lambda I)\vec{x} = \vec{0}$, whose augmented matrix is $[A - \lambda I | \vec{0}]$.

Example 70 *Continuing with the example above, recall that A has eigenvalues 2 and 3. Let's find first the eigenvectors associated to $\lambda = 2$. One needs to solve the system whose augmented matrix is*

$$[A-2I|0] = \left(\begin{array}{ccc|c} 2 & 4 & -2 & 0 \\ 1 & 2 & -1 & 0 \\ 3 & 6 & -3 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow E_2 = \left\langle \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\rangle_{\mathbb{R}}$$

Now, in order to find the eigenvectors associated to $\lambda = 3$, one needs to solve the system whose augmented matrix is

$$[A-3I|0] = \left(\begin{array}{ccc|c} 1 & 4 & -2 & 0 \\ 1 & 1 & -1 & 0 \\ 3 & 6 & -4 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 3 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow E_3 = \left\langle \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} \right\rangle_{\mathbb{R}}$$

Remark 61 *If one makes a mistake and computes wrongly the eigenvalues, when trying to compute the eigenspaces for a value of λ which is not an eigenvalue, one would obtain that $Nullity(A - \lambda I) = 0$. Hence, one can go back and check the characteristic equation and the roots found.*

Theorem 18 *Let $A_{n\times n}$ be a square matrix such that the roots of its characteristic polynomial are $\lambda_1, \ldots, \lambda_m$ with corresponding algebraic multiplicities $d_1, \ldots, d_m$. Then for all $i = 1, 2, \ldots, m$, one has $1 \le dim(E_{\lambda_i}) \le d_i$.*

Remark 62 *The previous theorem is valid even if one works in the complex numbers. By the fundamental theorem of algebra, all the roots of the characteristic polynomial are complex numbers, which include the real numbers. Hence, in the notation of the previous theorem, $d_1 + \ldots + d_m = n$. So it is special when a matrix with real entries has only real eigenvalues and the dimensions of the eigenspaces add up to n.*

Lemma 1 *Let $A_{n\times n}$ be a square matrix such that the characteristic polynomial has roots λ_i and multiplicities d_i. Assume that $E_{\lambda_i} = \langle \vec{v}_{i,1}, \ldots, \vec{v}_{i,e_i} \rangle$. By the previous theorem, one knows that $e_i \le d_i$. Hence, the union of all bases of all the eigenspaces is a linearly independent set in $\mathbb{R}^n$.*

One way to prove the theorem is as follows:

- Step 1: Take a basis for $E_\lambda = \langle \vec{v}_1, \ldots, \vec{v}_d \rangle$ and expand it to a basis for $\mathbb{R}^n = \langle \vec{v}_1, \ldots, \vec{v}_n \rangle$. And define $M = [\vec{v}_1, \ldots, \vec{v}_n]$.
- Step 2: Notice that
$$\det(M^{-1}AM - xI) = \det(M^{-1}(A - xI)M) = \det(A - xI)$$
- Step 3: Notice that $AM = [A\vec{v}_1, \ldots, A\vec{v}_d, \ldots, A\vec{v}_n] = [\lambda\vec{v}_1, \ldots, \lambda\vec{v}_d, w_{d+1}, \ldots, w_n]$. Thus, $M^{-1}AM$ has an upper-left corner submatrix equal to λI_d, and only 0 entries below that.
- Step 4: $\det(A - xI) = \det(M^{-1}AM - xI) = (x - \lambda)^d Q(x)$

8.1 Computation Strategies

Theorem 19 *Given a square matrix $A_{n\times n}$ with an eigen pair $(\lambda, \vec{x})$ consisting of an eigenvalue and a corresponding eigenvector, i.e. $A\vec{x} = \lambda\vec{x}$. Then the matrix cA has an eigen pair $(c\lambda, \vec{x})$. Indeed,*

$$(cA)\vec{x} = c(A\vec{x}) = c(\lambda\vec{x}) = (c\lambda)\vec{x}$$

Example 71 *Determine the eigenvectors and the eigenspaces for the matrix $A = \begin{bmatrix} 1/7 & 3/7 \\ 1/7 & -1/7 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}$. One can decide to proceed to compute its characteristic polynomial, its roots, and the eigenspaces. Rather than doing that, let's apply the previous theorem. Let $B = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}$. Since B has only integer coefficients, then it is easier to algebraically manipulate B than A. Then $ChaR(B) = \det(B - xI_2) = x^2 - 4$ and thus the eigenvalues of B are $\lambda = \pm 2$. Moreover, the eigenspaces for B are $E_2(B) = \left\langle \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\rangle$ and $E_{-2}(B) = \left\langle \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\rangle$. Now, one applies the theorem and finds out that the eigenvalues of $A = \frac{1}{7}B$ are $\lambda = \pm 2/7$ and the eigenspaces for A are $E_{2/7}(A) = \left\langle \begin{bmatrix} 3 \\ 1 \end{bmatrix} \right\rangle$ and $E_{-2/7}(A) = \left\langle \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\rangle$.*

Theorem 20 *Given a square matrix $A_{n\times n}$ with an eigenpair $(\lambda, \vec{x})$.*

- *Then for all* $d \in \mathbb{Z}^+$, *the matrix* A^d *has an eigenpair of the form* $(\lambda^n, \vec{x})$.
- *If* $p(x) = a_n x^n + \ldots + a_0$ *is a polynomial, then the matrix* $p(A)$ *has an eigenpair of the form* $(p(\lambda), \vec{x})$.
- *If* A *is also invertible and* $p(x) = a_{-m}x^{-m} + \ldots + a_0 + \ldots a_n x^n$ *is a generalization of a polynomial that allows negative exponents, then* $p(A)$ *has an eigenpair of the form* $(p(\lambda), \vec{x})$.

Example 72 *Let* $A_{3\times 3}$ *be an invertible matrix with eigenvalues* $\lambda = 1, 2, 3$. *Then* A^2 *has eigenvalues* $1^2, 2^2, 3^2$; A^3 *has eigenvalues* $1^3, 2^3, 3^3$; *etc. Additionally, in order to find the eigenvalues of* $A^2 - 7A + 2I_3$, *one uses the auxiliary polynomial* $p(\lambda) = \lambda^2 - 7\lambda + 2$, *so its eigenvalues are* $p(1) = -4$, $p(2) = -8$, *and* $p(3) = -10$. *Moreover, since* A *has at most* 3 *eigenvalues and one knows that none of them is zero, then* A *is invertible, so in order to find the eigenvalues of* $A^{-2} + 2I_3 + 3A$, *one uses the auxiliary function* $p(\lambda) = \lambda^{-2} + 2 + 3\lambda$, *so its eigenvalues are* $p(1) = 6$, $p(2) = 33/4$ *and* $p(3) = 100/9$.

Example 73 *Let* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *be a rotation by* 90 *degrees around some linear subspace of dimension* 1 *whose coefficient matrix is*

$$A_T = \begin{pmatrix} -\frac{17}{42} & \frac{32}{21} & -\frac{23}{42} \\ -\frac{17}{42} & \frac{11}{21} & \frac{19}{42} \\ \frac{31}{42} & -\frac{4}{21} & \frac{37}{42} \end{pmatrix}$$

Find a base for the axis of rotation as a linear subspace of $\mathbb{R}^3$. *and extend this to a base for* $\mathbb{R}^3$. *Solution: If* T *is a rotation, then any vector in the axis of rotation must be unchanged by this transformation. Hence one needs to find the eigenspace associated to* $\lambda = 1$ *for the coefficient matrix* A_T, *i.e.* E_1. *Again this is too complicated due to the fraction in its entries. Let* $B = 42A = \begin{pmatrix} -17 & 64 & -23 \\ -17 & 22 & 19 \\ 31 & -8 & 37 \end{pmatrix}$. *Then* $Char(B) = -x^3 + 42x^2 - 176x + 74088$. *One is expecting* 1 *to be an eigenvalue of* A, *so* 42 *must be an eigenvalue of* B. *Thus* $(x - 42)$ *must be a factor of* $Char(B)$. *Indeed, through long division or factoring, one can verify that it is* $Char(B) = -(x-42)(x^2+1764)$. *Hence,* $\lambda_B = 42$ *is the only real eigenvalue for* B. *Now, by the bounds of the dimension of the eigenspaces, one knows that* $E_{42}(B) = Null(B - 42I)$ *is* 1*-dimensional. A long computation leads to* $E_{42}(B) = Null(B - 42I) = \left\langle \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\rangle$. *By the previous theorem,* $E_1(A) = E_{42}(B)$. *Therefore, the rotation is about the span of the vector* $(1, 2, 3)^T$. *Clearly, an extension to a basis of the whole Euclidean space leads to* $\mathbb{R}^3 = \langle \vec{e}_1, \vec{e}_2, (1, 2, 3)^T \rangle$.

8.2 Diagonalization

Recall that the dimension of the eigenspace E_λ associated to λ satisfies $1 \leq dim(E_\lambda) \leq mult(\lambda)$ where $mult(\lambda)$ is the algebraic multiplicity of λ or the multiplicity as a root of the characteristic polynomial.

The following statements are equivalent: Given a matrix $A_{n\times n}$: (1) The sum of the dimensions of all the eigenspaces is equal to n. (2) There is a basis for $\mathbb{R}^n$ formed by combining bases of all the eigenspaces, i.e. there is a basis for $\mathbb{R}^n$ consisting of eigenvectors of A associated to the different eigenvalules. (3) The dimension of each eigenspace E_λ is equal (not less than) to $mult(\lambda)$.

If any of these statements occurs, then A is said to be **diagonalizable**, and it has a diagonal decomposition, i.e. there exist a diagonal matrix $D_{n\times n}$ and an invertible matrix P such that $A = PDP^{-1}$. In order to find these matrices D and P, assume that $\mathbb{R}^n = \langle \vec{v}_1, \ldots, \vec{v}_n \rangle$ is formed by eigenvectors $\vec{v}_i$, additionally assume that the eigenvalue associated to $\vec{v}_i$ is λ_i. Then the list $\lambda_1, \ldots, \lambda_n$ may contained repeated eigenvalues according to its multiplicities. Then D is the diagonal matrix whose diagonal entries are $\lambda_1, \ldots, \lambda_n$; and the matrix $P = [\vec{v}_1, \ldots \vec{v}_n]$.

Example 74 *If* $A = \begin{bmatrix} 3 & 6 & 5 \\ 3 & 2 & 3 \\ -5 & -6 & -7 \end{bmatrix}$, *then its characteristic polynomial is* $Char(A) = \det(A - xI_3) = -(x+2)^2(x-2)$ *and its eigenspaces are* $E_{-2} = \left\langle \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\rangle$ *and* $E_2 = \left\langle \begin{bmatrix} -3 \\ -2 \\ 3 \end{bmatrix} \right\rangle$. *Notice that* $dim(E_{-2}) = 1 < 2 = multiplicity(-2)$, *so this matrix is NOT diagonalizable.*

Example 75 *If* $A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}$, *then* $Char(A) = x^2 - 4 = (x - 2)(x + 2)$, $E_2 = \langle (3, 1)^T \rangle$ *and* $E_{-2} = \langle (1, -1)^T \rangle$. *Since the dimensions of the eigenspaces are both* 1, *and none of the eigenvalues is zero, then* A *is invertible and* *diagonalizable. In order to find its diagonal decomposition, let* $B = \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix}$ *be the matrix whose columns are the basis vectors of the eigenspaces, and* $D = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$ *is the diagonal matrix whose diagonal entries are the corresponding eigenvalues.*

Remark 63 *The diagonal decomposition* $A = PDP^{-1}$ *is not unique, but it depends on an arbitrary order of the eigenvalues and the basis of* $\mathbb{R}^n$ *consisting of eigenvectors of* A*.*

It is important to verify in the previous example that:

$$A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix}^{-1} = BDB^{-1}$$

Similarly,

$$A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}^{-1} = BDB^{-1}$$

8.3 Examples

Example 76 *Let* $T : \mathbb{R}^{n+1} \to \mathbb{R}^n$ *be a surjective linear transformation with coefficient matrix* A*. If* 0 *is an eigenvalue, then find the dimension of the corresponding eigenspace, i.e.* $dim(E_0)$*.*

Solution: By the Rank-Nullity theorem, one knows that $dim(Ker(T)) + dim(Range(T)) = n + 1$. However, one also knows that $E_0 = Ker(T)$ and $Range(T) = \mathbb{R}^n$ because T is surjective. Therefore, $dim(E_0)+n = n+1 \Rightarrow dim(E_0) = 1$.

Example 77 *Let* $A_{6\times 6}$ *be a diagonalizable matrix with eigenvalues* $1, 2, 3$ *and multiplicities* $3, 2, 1$ *respectively. Find* $\det(A)$*.*

Solution: From polynomials, one knows that the constant term of the polynomial is the product of its roots. Since A is diagonalizable, one knows that the roots of its characteristic polynomial are $1, 1, 1, 2, 2, 3$ counting repetitions. Additionally, one knows that the constant term of its characteristic polynomial is its determinant. Hence $\det(A) = 1 \cdot 1 \cdot 1 \cdot 2 \cdot 2 \cdot 3 = 12$.

Example 78 *Assume* $A_{2\times 2}$ *has two different eigenvalues* λ_1, λ_2 *with corresponding eigenvectors* $\vec{v}_1, \vec{v}_2$*. Show that* $\det[\vec{v}_1, \vec{v}_2] \neq 0$*.*

Solution: Since they are eigenvectors, one has $A\vec{v}_1 = \lambda_1\vec{v}_1$ and $A\vec{v}_2 = \lambda_2\vec{v}_2$, and they are nonzero vectors, i.e. $\vec{v}_1 \neq \vec{0}$ and $\vec{v}_2 \neq \vec{0}$. Let $B = [\vec{v}_1, \vec{v}_2]$. Then B is invertible if and only if $\vec{v}_1$ and $\vec{v}_2$ are linearly independent vectors. If they were linearly dependent, then there is a scalar $c \in \mathbb{R}$ with $\vec{v}_1 = c\vec{v}_2$. However, this would imply that $\lambda_2\vec{v}_2 = A\vec{v}_2 = Ac\vec{v}_1 = cA\vec{v}_1 = c\lambda_1\vec{v}_1 = \lambda_1 c\vec{v}_1 = \lambda_1\vec{v}_2$, which is a contradiction because $\lambda_1 \neq \lambda_2$ and $\vec{v}_2 \neq \vec{0}$.

Example 79 *Let* $A_{2\times 2}$ *have eigenvalues* ± 1*. Determine if* $A^{2019} = A$*.*

Solution: Since A has different eigenvectors, A is diagonalizable. Assume the diagonal decomposition of A is PDP^{-1} where $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Notice that $A^n = PD^nP^{-1}$ and $D^k = I_2$ if k is even and $D^k = D$ if k is odd. Therefore, $A^{2019} = PD^{2019}P^{-1} = PDP^{-1} = A$.

Example 80 *Let* $A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \end{bmatrix}$*. Find* A^{20}*.*

Solution: Since A is not a diagonal matrix, then finding such a large power of A by direct computation would be an enourmous task. Notice however that $Char(A) = (x - 2)(x + 2)$ and A has two different eigenvalues ± 2. Hence A is diagonalizable. A direct computation would show that $E_2 = \langle (3, 1)^T \rangle$ and $E_{-2} = \langle (1, -1)^T \rangle$. Therefore,

$$A^{20} = \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2^{20} & 0 \\ 0 & (-2)^{20} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 2^{20} & 0 \\ 0 & 2^{20} \end{bmatrix} = 2^{20} I_2$$

Theorem 21 *If* $A_{n\times n}$ *has characteristic polynomial* $Char(A) = a_n x^n + \ldots + a_0$*, then* $a_n A^n + \ldots + a_0 I_n = 0_{n\times n}$

Example 81 *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$ *whose characteristic polynomial is* $x^3 - 15x^2 + 78x + 96$*. Find* $A^3 - 15A^2 + 78A$*.*

Solution: From the previous theorem, one has $A^3 - 15A^2 + 78A = -96I_3$.

Example 82 *Let* $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$ *whose characteristic polynomial is* $x^3 - 15x^2 + 78x + 96$*. Find scalars* a, b, c *such that* $A^5 = aA^2 + bA + cI_3$*.*

Solution: From the previous theorem, one knows that $A^3 = 15A^2 - 78A - 96I$. Therefore,

$$A^4 = A \cdot A^3 = A(15A^2 - 78A - 96I) = 15A^3 - 78A^2 - 96A = 15(15A^2 - 78A - 96I) - 78A^2 - 96A = 147A^2 - 486A - 480I$$

And

$$A^5 = A \cdot A^4 = A(147A^2 - 486A - 480I) = 147A^3 - 486A^2 - 480A = 1719A^2 - 11946A - 14112$$

Therefore, $a = 1719, b = -1194, c = 14112$.

Example 83 *Consider the hyperplane* H *in* $\mathbb{R}^4$ *determined by the equation* $ax_1 + bx_2 + cx_3 + dx_4 = 0$ *where none of the coefficients* a, b, c, d *are zero. Find a basis for this hyperplane* H*.*

Solution: Since $a \neq 0$, then x_1 is the only leading variable and x_2, x_3, x_4 are free variables. Therefore the general solution to this equation has the form $(-\frac{b}{a}x_2 - \frac{c}{a}x_3 - \frac{d}{a}x_4, x_2, x_3, x_4)$ where x_2, x_3, x_4 are free to have any real value. Therefore,

$$H = \left\langle \begin{bmatrix} -\frac{b}{a} \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{c}{a} \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{d}{a} \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\rangle$$

Example 84 *Let* $\vec{v}_1, \vec{v}_2, \vec{v}_3$ *be linearly independent vectors in* $\mathbb{R}^4$*. Let* $\vec{w}_1, \vec{w}_2$ *two other linearly independent vectors in* $\mathbb{R}^4$ *such that both* $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1\}$ *and* $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1\}$ *are bases for* $\mathbb{R}^4$*. Let* $\vec{x} = (x_1, x_2, x_3, x_4)^T$ *be the unique solution to the system whose augmented matrix is* $[\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1|\vec{w}_2]$*. Find the unique solution* $\vec{y}$*, in terms of* $\vec{x}$*, of the system whose augmented matrix is* $[\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1|\vec{w}_2]$*.*

Solution: The vector equation corresponding to the augmented matrix $[\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1|\vec{w}_2]$ is

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 + x_4\vec{w}_1 = \vec{w}_2$$

Which solving for $\vec{w}_1$ translates to

$$-\frac{x_1}{x_4}\vec{v}_1 - \frac{x_2}{x_4}\vec{v}_2 - \frac{x_3}{x_4}\vec{v}_3 + \frac{1}{x_4}\vec{w}_2 = \vec{w}_1$$

Therefore,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \Rightarrow \vec{y} = \begin{bmatrix} -x_1/x_4 \\ -x_2/x_4 \\ -x_3/x_4 \\ 1/x_4 \end{bmatrix}$$

Example 85 *Let* $a, b, c, d \in \mathbb{R}$*. Find the following determinant:*

$$\det \begin{bmatrix} 1 & a & a^2 & a^3 \\ 1 & b & b^2 & b^3 \\ 1 & c & c^2 & c^3 \\ 1 & d & d^2 & d^3 \end{bmatrix}$$

Solution: One could proceed as regularly and compute the determinant either by Gauss elimination and keeping track of the row elementary operations performed, or by cofactor expansions. We would however mixed the later with some basic algebra. By cofactor expansion on the first row, the determinant is equal to:

$$\det \begin{bmatrix} b & b^2 & b^3 \\ c & c^2 & c^3 \\ d & d^2 & d^3 \end{bmatrix} - a \det \begin{bmatrix} 1 & b^2 & b^3 \\ 1 & c^2 & c^3 \\ 1 & d^2 & d^3 \end{bmatrix} + a^2 \det \begin{bmatrix} 1 & b & b^3 \\ 1 & c & c^3 \\ 1 & d & d^3 \end{bmatrix} - a^3 \det \begin{bmatrix} 1 & b & b^2 \\ 1 & c & c^2 \\ 1 & d & d^2 \end{bmatrix}$$

Hence, the determinant is a cubic polynomial on a. By algebra, one knows that a cubic polynomial satisfies:

$$c_3a^3 + c_2a^2 + c_1a + c_0 = c_3(a - r_1)(a - r_2)(a - r_3)$$

where notice the role of the leading coefficient, and the roots r_i's'. Our leading coefficient is $-\det \begin{bmatrix} 1 & b & b^2 \\ 1 & c & c^2 \\ 1 & d & d^2 \end{bmatrix}$, which could be analized as a quadratic polynomial on b, and analogously to the following reasoning, one discovers that the leading coefficient is $-(d-c)(b-c)(b-d) = (b-c)(b-d)(c-d)$. Indeed, the roots of the polynomial on a above are b, c, d because if $a = b$, c, or d, then the matrix would have two identical roots, and thus the determinant would be zero. Hence,

$$\det \begin{bmatrix} 1 & a & a^2 & a^3 \\ 1 & b & b^2 & b^3 \\ 1 & c & c^2 & c^3 \\ 1 & d & d^2 & d^3 \end{bmatrix} = (a-b)(a-c)(a-d)(b-c)(b-d)(c-d)$$

Example 86 *Let $a, b \in \mathbb{R}$. Compute a formula for the determinant of a square matrix $A_{m\times m}$ which has repeated entries $a+b$ right above the diagonal, and b in all other entries.*

Solution: Let $A_m = A_{m\times m}$ for short. Since $\det(A) = \det(A^T)$, one can assume that A has entries $a+b$ right below the diagonal and b everywhere else. For example

$$A_3 = \begin{bmatrix} b & b & b \\ a+b & b & b \\ b & a+b & b \end{bmatrix}$$

By subtracting the first row from all other rows, one has:

$$\det(A_3) = \det \begin{bmatrix} b & b & b \\ a & 0 & 0 \\ 0 & a & 0 \end{bmatrix}$$

And by cofactor expansion on the last column, one obtains:

$$\det(A_m) = (-1)^{m+1} b \det(aI_{m-1}) = ba^{m-1}$$

Example 87 *Let $a, b \in \mathbb{R}$. Compute a formula for the determinant of a square matrix $A_{m\times m}$ which has repeated entries $a+b$ in the diagonal, and b in all other entries.*

Solution: If $a = 0$, then the matrix has all columns equal, and thus $\det(A) = 0$. If $b = 0$, then $A_m = aI_m$ and hence $\det(A_m) = a^m$. For all other cases, assume $a \neq 0$ and $b \neq 0$.

For $m = 1$, the matrix is $A = [a+b]$ with determinant $a+b$. For $m = 2$, $\det(A) = \det \begin{bmatrix} a+b & b \\ b & a+b \end{bmatrix}$. Therefore, the determinant is $(a+b)^2 - b^2 = a^2 + 2ab = a(a+2b)$. For $m = 3$, the matrix A would be equal to:

$$A = \begin{bmatrix} a+b & b & b \\ b & a+b & b \\ b & b & a+b \end{bmatrix}$$

whose determinant is $a^3 + 3a^2b = a^2(a+3b)$. A careful notice of these results shows that a good candidate is $\det(A_{m\times m}) = a^{m-1}(a+mb)$. Now, see if one can use these results and generalize. Use row elementary operations and subtract the first row from all other rows. For example,

$$\det(A_m) = \det \left(\begin{array}{c|c} a+b & (b, \dots, b)_{1\times m-1} \\ \hline \begin{bmatrix} -a \\ \vdots \\ -a \end{bmatrix}_{m-1\times 1} & aI_{m-1} \end{array} \right)$$

Since, $\det(A) = \det(A^T)$, transpose and add each row to the first one,

$$\det(A_m) = \det \left(\begin{array}{c|c} a+mb & (0,\dots,0)_{1\times m-1} \\ \hline \begin{bmatrix} b \\ \vdots \\ b \end{bmatrix}_{m-1\times 1} & aI_{m-1} \end{array} \right)$$

By cofactor expansion on the first row, one has

$$\det(A_m) = (a+mb)\det(aI_{m-1}) = a^{m-1}(a+mb)$$

8.4 Worksheet

For each of the following matrices A, answer all the following questions:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & -8 & 9 \end{bmatrix}$$

1. Find the dimensions of the matrix and call them m and n, i.e. $A_{m\times n}$
2. Find $row(A), col(A), null(A)$. Also find $rank(A), nullity(A)$.
3. Find the determinant of $B = AA^T$, determine if B is invertible. Also find the cofactor matrix, the adjugate matrix, and the inverse of B whenever it is possible.
4. Find the eigenvalues and bases for the eigenspaces of B.
5. Determine if A has a left inverse matrix and find it if possible. Also determine if A has a right inverse matrix and find it if possible.
6. Let $T : \mathbb{R}^p \to \mathbb{R}^q$ be the associated linear transformation, i.e. $T\vec{x} = A\vec{x}$. Find the dimensions of the domain p and codomain q.
7. Find bases for $Ker(T)$ and $Range(T)$, and find $dim(Ker(T))$ and $dim(Range(T))$.
8. Determine if T is injective, surjective, and/or bijective.
9. Determine if T is invertible and find the coefficient matrix of its inverse if possible.
10. Determine if T has a left inverse and/or a right inverse transformation, and find their associated matrices whenever it is possible.

9 Derived Subspaces

9.1 The Intersection of Two Subspaces

Sometimes it is important to understand the relation of two subspaces in the same Euclidean domain. Recall that when one is contained in the other, i.e. $S \subseteq T \subseteq \mathbb{R}^n$, one knows that $dim(S) \le dim(T)$, and $S = T$ if and only if $dim(S) = dim(T)$. However, if there is no such relation, one might be interested in the in the largest common subspace, i.e.: the largest subspace U such that $U \subseteq S$ and $U \subseteq T$. THis is known as the intersection subspace, i.e. $S \cap T$.

Recall that geometrically the solution set to a system of linear equations can be understood as the intersection of the hyperplanes defined by all of these equations. Similarly, one has the following result.

Theorem 22 *Let S and T be subspaces defined by homogeneous systems with coefficient matrices are A and B respectively. Then the intersection $S \cap T$ is given by the homogeneous system with coefficient matrix* $\left(\dfrac{A}{B}\right)$.

However, when the subspaces are defined by their bases, it implies the following result.

Theorem 23 *Let $S = \langle \vec{v}_1, \ldots, \vec{v}_m \rangle$ and $T = \langle \vec{w}_1, \ldots, \vec{w}_n \rangle$ be two subspaces of $\mathbb{R}^N$. Then an algorithm to find the intersection $S \cap T$ is:*

1. *Find a basis for $Null[\vec{v}_1, \ldots, \vec{v}_m, \vec{w}_1, \ldots, \vec{w}_m]$, and denote it by $\langle \vec{n}_1, \ldots, \vec{n}_q \rangle$.*
2. *Replace each of the vectors $\vec{x}_i \in \mathbb{R}^{m+n}$ by a smaller vector $\vec{y}_i \in \mathbb{R}^m$ consisting of the first entries of $\vec{x}_i$.*
3. *Then $S \cap T$ is the span of all the vectors of the form $[\vec{v}_1, \ldots, \vec{v}_m]\vec{y}_i$.*

Yet, a more sophisticated results is the following:

Theorem 24 (Zassenhaus' algorithm) *Let $S = \langle \vec{v}_1, \ldots, \vec{v}_m \rangle$ and $T = \langle \vec{w}_1, \ldots, \vec{w}_n \rangle$ be two subspaces of $\mathbb{R}^N$. Write the following matrix and apply Jordan-Gauss elimination*

$$\left(\begin{array}{c|c} \vec{v}_1^T & \vec{v}_1^T \\ \vdots & \vdots \\ \vec{v}_m^T & \vec{v}_m^T \\ \hline \vec{w}_1^T & \vec{0}_N^T \\ \vdots & \vdots \\ \vec{w}_n^T & \vec{0}_N^T \end{array}\right)$$

to obtain a matrix of the form $\begin{pmatrix} A_{d\times N} & B_{d\times n} \\ 0_{N-d\times N} & C_{N-d\times N} \end{pmatrix}$. *Then the transpose of the rows of the matrix A are a basis of $S \cap T$. Moreover, the dimension of $S \cap T$ is d.*

Example 88 *Let $U = \langle (1,-1,0,1)^T, (0,0,1,-1)^T \rangle$ and $W = \langle (5,0,-3,3)^T, (0,5,-3,-2)^T \rangle$. Find a basis for $U \cap W$.*

Solution: Notice first that $dim(U) = dim(W) = 2$ and they live in $\mathbb{R}^4$. Consider the following matrix

$$A = \left(\begin{array}{rrrr|rrrr} 1 & -1 & 0 & 1 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\ \hline 5 & 0 & -3 & 3 & 0 & 0 & 0 & 0 \\ 0 & 5 & -3 & -2 & 0 & 0 & 0 & 0 \end{array}\right)$$

Applying Gauss elimination to obtain

$$A_E = \left(\begin{array}{rrrr|rrrr} 1 & 4 & 0 & -4 & 1 & -1 & 3 & -2 \\ 0 & 5 & 0 & -5 & 0 & 0 & 3 & -3 \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\ \hline 0 & 0 & 0 & 0 & 5 & -5 & 0 & 5 \end{array}\right)$$

which indicates that

$$S \cap T = \left\langle \begin{bmatrix} 1 \\ 4 \\ 0 \\ -4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ 0 \\ -5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\rangle$$

Or alternative, apply Jordan-Gauss elimination to obtain

$$A_{RE} = \left(\begin{array}{rrrr|rrrr} 1 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} & -\frac{3}{5} \\ 0 & 1 & 0 & -1 & 0 & 0 & \frac{3}{5} & -\frac{3}{5} \\ 0 & 0 & 1 & -1 & 0 & 0 & 1 & -1 \\ \hline 0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right)$$

which indicates that

$$S \cap T = \left\langle \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \right\rangle = \langle \vec{e}_1, \vec{e}_2 - \vec{e}_4, \vec{e}_3 - \vec{e}_4 \rangle$$

9.2 The Sum of Two Subspaces

Regretably the union of two subspaces S and T of $\mathbb{R}^N$ is in general not a subspace. A simple example is obtained by considering $S = Span(\vec{e}_1)$ and $T = Span(\vec{e}_2)$ in $\mathbb{R}^2$. The union of them would be the two axis by themselves. However, the union is not closed under addition in this example because $\vec{e}_1 + \vec{e}_2$ is outside the axis.

Just like one considers the intersection to be the largest common subspace, the sum is the smallest common space containing both subspaces. Formally, one defines

$$S + T = \{\vec{s} + \vec{t} \in \mathbb{R}^N : \vec{s} \in S, \vec{t} \in T\}$$

Theorem 25 *Let $S = Span(\vec{v}_1, \ldots, \vec{v}_m)$ and $T = Span(\vec{w}_1, \ldots, \vec{w}_n)$ be two subspaces of $\mathbb{R}^N$. Then*

$$S + T = Span(\vec{v}_1, \ldots, \vec{v}_m, \vec{w}_1, \ldots, \vec{w}_n)$$

Remark 64 *Notice that even if $S = \langle \vec{v}_1, \ldots, \vec{v}_m \rangle$ and $T = \langle \vec{w}_1, \ldots, \vec{w}_m \rangle$, one does not necessarily have*

$$S + T = \langle \vec{v}_1, \ldots, \vec{v}_m, \vec{w}_1, \ldots, \vec{w}_n \rangle$$

Indeed, even if one starts with two linearly independent spanning sets, one may end up with a linearly dependent spanning set for the sum.

Example 89 *Let $S = \langle \vec{e}_1, \vec{e}_2 \rangle$ and $T = \langle \vec{e}_2, \vec{e}_3 \rangle$ be two subspaces of $\mathbb{R}^4$. Notice that the sum is $S + T = \langle \vec{e}_1, \vec{e}_2, \vec{e}_3 \rangle$.*

10 Some Geometric Linear Transformations

10.1 Orthogonal Projection

Although visually almost anyone understands what two orthogonal or perpendicular vectors are, the generalization to higher dimensions must be well defined.

The **dot product** of two real vectors in the same Euclidean space $\mathbb{R}^n$ is defined to be: (1) From the matrix perspective, $\vec{v} \cdot \vec{w} := \vec{v}^T \cdot \vec{w}$, where the dot in the left hand side is called dot product, and the dot in the right hand side is the standard matrix multiplication. (2) Alternatively, one might look at the coordinates of the two vectors $\vec{v} = (v_1, \ldots, v_n)^T$ and $\vec{w} = (w_1, \ldots, w_n)^T$, then

$$\vec{v} \cdot \vec{w} = \sum_{i=1}^{n} v_i \cdot w_i = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$$

Proposition 16 *Let $a \in \mathbb{R}$ and $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$. Then the following properties hold:*

- *Positiveness: $\vec{x} \cdot \vec{x} \geq 0$. Moreover, $\vec{x} \cdot \vec{x} = 0$ if and only if $\vec{x} = 0$.*
- *Commutative: $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$. Commutative with scalar multiplication: $a(\vec{x} \cdot \vec{y}) = (a\vec{x}) \cdot \vec{y} = \vec{x} \cdot (a\vec{y})$.*
- *Distributive over sum/difference: $\vec{x} \cdot (\vec{y} \pm \vec{z}) = \vec{x} \cdot \vec{y} \pm \vec{x} \cdot \vec{z}$. And $(\vec{x} \pm \vec{y}) \cdot \vec{z} = \vec{x} \cdot \vec{z} + \vec{y} \cdot \vec{z}$.*

If one defines $|\vec{v}|^2 = \vec{v} \cdot \vec{v}$, then one has $\vec{v} \cdot \vec{w} = |\vec{v}|\,|\vec{w}| \cos(\theta)$ where θ is the angle formed by the two vectors in the plane that contains them both.

Remark 65 *All of these properties only hold in $\mathbb{R}^n$, and changes must be made in $\mathbb{C}^n$. In particular, one wants to keep positiveness, which in the real euclidean space, it holds because $\vec{x} \cdot \vec{x} = \sum_{i=1}^{n} x_i^2$ and all squares are non-negative numbers, and so is their sum.*

For completeness, one defines the **dot product** of two complex vectors $\vec{x}, \vec{y} \in \mathbb{C}^n$ by $\vec{x} \cdot \vec{y} = \vec{x}^T \cdot \overline{\vec{y}} = \sum_{i=1}^{n} x_i \overline{y_i}$.

Proposition 17 *Let $a \in \mathbb{C}$ and $\vec{x}, \vec{y}, \vec{z} \in \mathbb{C}^n$. Then the following properties hold:*

- *Positiveness: $\vec{x} \cdot \vec{x} \geq 0$. Moreover, $\vec{x} \cdot \vec{x} = 0$ if and only if $\vec{x} = 0$.*
- *Commutative: $\vec{x} \cdot \vec{y} = \overline{\vec{y} \cdot \vec{x}}$. Commutative with scalar multiplication: $a(\vec{x} \cdot \vec{y}) = (a\vec{x}) \cdot \vec{y} = \vec{x} \cdot (\overline{a}\vec{y})$.*
- *Distributive over sum/difference: $\vec{x} \cdot (\vec{y} \pm \vec{z}) = \vec{x} \cdot \vec{y} \pm \vec{x} \cdot \vec{z}$. And $(\vec{x} \pm \vec{y}) \cdot \vec{z} = \vec{x} \cdot \vec{z} + \vec{y} \cdot \vec{z}$.*

  One defines two vectors $\vec{v}$ and $\vec{w}$ to be orthogonal if and only if $\vec{v} \cdot \vec{w} = 0$.

10.2 Orthogonal Projections

Geometrically, the orthogonal projection of $(x_1, x_2)^T$ to the x-axis is $(x_1, 0)^T$. A great example of these are building blueprints, which forget about the third dimension and illustrate the projection of the walls of the building into a 2-dimensional plane.

One defines the **orthogonal projection of a vector $\vec{v}$ onto a vector $\vec{w}$** as:

$$Proj_{\vec{w}}\vec{v} = \left(\frac{\vec{v}\cdot\vec{w}}{\vec{w}\cdot\vec{w}}\right)\vec{w}$$

The same can be said about the projection of a vector into a linear subspace S of dimension 1, because $S = \langle\vec{w}\rangle$, and hence $Proj_S\vec{v} = Proj_{\vec{w}}\vec{v}$.

Theorem 26 (Gram-Schmidt) *Given a list of linearly independent vectors*

$$\vec{v}_1, \dots, \vec{v}_m$$

in $\mathbb{R}^n$, one can produce a list of linearly independent orthogonal vectors $\vec{w}_1, \dots, \vec{w}_m$ with the same span, i.e.

$$Span(\vec{v}_1, \dots, \vec{v}_m) = Span(\vec{w}_1, \dots, \vec{w}_m)$$

as follows:

- *Let $\vec{w}_1 = \vec{v}_1$.*
- *For $k = 2, 3, \dots, m$: Let*

$$\vec{w}_k = \vec{v}_k - \sum_{i=1}^{k-1} Proj_{\vec{w}_i}\vec{v}_k$$

Remark 66 *One may normalize at the end of the process or along the process by replacing the vectors $\vec{w}_i$ by $\frac{\vec{w}_i}{|\vec{w}|}$.*

Remark 67 *Just like one was able to extend a linearly independent set S in $\mathbb{R}^n$ to a basis of $\mathbb{R}^n$ containing S, one might extend a linearly independent pair-wise orthogonal set S in $\mathbb{R}^n$ to a basis of $\mathbb{R}^n$ containing S by extending the basis first, and applying Gram-Schmidt process to the resulting basis.*

Now, one is able to generalize projections and one defines the **orthogonal projection of a vector $\vec{v}$ onto a subspace S with orthogonal basis $\{\vec{w}_1, \dots, \vec{w}_m\}$ in $\mathbb{R}^n$** to be

$$Proj_S(\vec{v}) = \sum_{i=1}^{m} Proj_{\vec{w}_i}\vec{v}$$

Remark 68 *Although this process does not work when applied to a non-orthogonal basis of S, one might combine Gram-Schmidt process to obtain an orthogonal basis and this definition to obtain the projection onto a subspace given by any basis.*

Theorem 27 *Let $S = \langle\vec{v}_1, \dots, \vec{v}_m\rangle$ be a subspace $\mathbb{R}^n$ where the given basis vectors are pairwise orthogonal and normal, i.e. $\vec{v}_i \cdot \vec{v}_j = \delta_{ij}$ (equal to 0 if $i = j$ and equal to 1 if $i \neq j$). Then the function $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\vec{x}) = Proj_S(\vec{x})$ is a linear transformation whose coefficient matrix is*

$$[\vec{v}_1, \dots, \vec{v}_m]_{n\times m}[\vec{v}_1, \dots, \vec{v}_m]^T_{m\times n}$$

Moreover, if one extends the given basis to a basis of $\mathbb{R}^n = \langle\vec{v}_1, \dots, \vec{v}_n\rangle$ whose vectors are orthonormal, i.e. orthogonal and normal, then T has eigenvalues 1 with multiplicity m and 0 with multiplicity $n - m$ whose eigenspaces have basis:

$$E_1 = \langle\vec{v}_1, \dots, \vec{v}_m\rangle = S$$

$$E_0 = \langle\vec{v}_{m+1}, \dots, \vec{v}_n\rangle$$

Remark 69 *Given a subspace S of $\mathbb{R}^n$ with orthonormal basis $\mathcal{B}$, if one extends $\mathcal{B}$ to an orthonormal basis $\mathcal{B}'$ of $\mathbb{R}^n$, then the span of the added vectors $\mathcal{B}' - \mathcal{B}$ span another subspace $S^\perp$ called the orthogonal complement of S in $\mathbb{R}^n$. This is denoted $S \bigoplus S^\perp = \mathbb{R}^n$, which means that every vector $\vec{v} \in \mathbb{R}^n$ can be written in a unique way as the sum of a vector $Proj_S(\vec{v})$ in S and a vector $Proj_{S^\perp}(\vec{v})$ in $S^\perp$.*

10.3 Rotations

Another important linear transformation is to rotate the Euclidean space.

Example 90 *A rotation of the plane $\mathbb{R}^2$ counterclockwise by an angle θ is a linear transformation $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ whose coefficient matrix is*

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

It is characterized because it does not have real eigenvalues, unless $\theta = 0, \pi$, in which case, the eigenvalue is 1 or -1 with multiplicity 2 respectively. Indeed, its characteristic polynomial $Char(R_\theta) = 1 - 2\cos(\theta)x + x^2$ has discriminant $4(\cos(\theta)^2 - 1)$.

This example motivates a generalization of rotation of any Euclidean domain $\mathbb{R}^n$ where $n > 2$. Since the definition of an angle only makes sense in the plane containing the rays defining the angle, a rotation of $\mathbb{R}^n$ is a linear transformation with Eigenvalue 1 and $dim(E_1) \geq n-2$. If one finds an orthonormal basis of $E_1 = \langle \vec{v}_1, \ldots, \vec{v}_{n-2} \rangle$ and extends it to an orthonormal basis of $\mathbb{R}^n = \langle \vec{v}_1, \ldots, \vec{v}_n \rangle$, then the matrix of this linear transformation would be:

$$[\vec{v}_1, \ldots, \vec{v}_n] \begin{bmatrix} I_{n-2} & 0_{n-2\times 2} \\ 0_{2\times n-2} & R_\theta \end{bmatrix} [\vec{v}_1, \ldots, \vec{v}_n]$$

where R_θ is the rotation of the plane defined above.

Example 91 *Find the rotation matrix that sends $\vec{e}_1$ to $\vec{v} = \frac{1}{13}(3, 4, 12)^T$.*

Solution: Notice that $\|\vec{e}_1\| = \sqrt{1^2 + 0^2 + 0^2} = 1$ and $\|\vec{v}\| = \sqrt{3^2 + 4^2 + 12^2}/13 = 1$. At this moment, since the vectors are in $\mathbb{R}^3$, one knows that the rotation happens in a plane perpendicular to a single vector. One way to find that vector is:

Alternative 1: Using the cross-product, which sadly does not generalize to larger dimensions but it involves determinants and vectors. Indeed, one finds that

```
R.<i,j,k> = QQ[]
A = Matrix(R,3,[i,j,k,1,0,0,3/13,4/13,12/13])
```

$$\vec{e}_1 \times \vec{v} = \det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ 1 & 0 & 0 \\ 3/\sqrt{13} & 4/\sqrt{13} & 12/\sqrt{13} \end{bmatrix} = -\frac{12}{13}j + \frac{4}{13}k = \frac{1}{13}(0, -12, 4)^T$$

Alternative 2: One orthogonalizes the two given vectors by Gram-Schmidt, so $\vec{v}$ get replaced by $\vec{v}' = \vec{v} - \frac{\vec{v}\cdot\vec{e}_1}{\vec{e}_1\cdot\vec{e}_1}\vec{e}_1 = (0, 4/13, 12/13)^T$. And then find a vector that is orthogonal to them, i.e. $\vec{w} \cdot \vec{e}_1 = \vec{0}$, which forces $\vec{w} = (0, w_2, w_3)^T$ and $\vec{w} \cdot \vec{v}' = \vec{0}$, which forces $w_2 + 3w_3 = 0$.

Therefore, an ortho-normal basis for $\mathbb{R}^3$ that includes these vectors is $\mathcal{B} = \{\vec{e}_1, (0, 1, 3)^T/\sqrt{10}, (0, -3, 1)^T/\sqrt{10}\}$. Notice that $\vec{e}_1, \vec{v} \in \langle \vec{e}_1, (0, 1, 3)^T/\sqrt{10} \rangle$. Call $\mathbb{B} = \{(0, 1, 3)^T/\sqrt{10}, (0, -3, 1)^T/\sqrt{10}\}$. Moreover, $(\vec{e}_1)_{\mathbb{B}} = (1, 0)^T_{\mathbb{B}}$ and $\vec{v}_{\mathbb{B}} = (3/13, 4\sqrt{10}/13)^T_{\mathbb{B}}$.

So in order to rotate $\vec{e}_1$ to $\vec{v}$, one needs to find an angle such that:

$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{\mathbb{B}} = \begin{bmatrix} \frac{3}{13} \\ \frac{4\sqrt{10}}{13} \end{bmatrix}_{\mathbb{B}}$$

Hence $\cos(\theta) = 3/13$ and $\sin(\theta) = \frac{4\sqrt{10}}{13}$. Then $\theta = 1.3379281485368142$

```
P = Matrix(3, [3/13, -4*sqrt(10)/13, 0, 4*sqrt(10)/13, 3/13, 0, 0, 0, 1])
C = Matrix(3, [1,0,0,0,1/sqrt(10), 3/sqrt(10), 0,-3/sqrt(10), 1/sqrt(10)]).transpose()
R = C*P * C.inverse()
```

Therefore, the rotation matrix is given by:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{10}\sqrt{10} & -\frac{3}{10}\sqrt{10} \\ 0 & \frac{3}{10}\sqrt{10} & \frac{1}{10}\sqrt{10} \end{pmatrix} \begin{pmatrix} \frac{3}{13} & -\frac{4}{13}\sqrt{10} & 0 \\ \frac{4}{13}\sqrt{10} & \frac{3}{13} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{10}\sqrt{10} & \frac{3}{10}\sqrt{10} \\ 0 & -\frac{3}{10}\sqrt{10} & \frac{1}{10}\sqrt{10} \end{pmatrix} = \begin{pmatrix} \frac{3}{13} & -\frac{4}{13} & -\frac{12}{13} \\ \frac{4}{13} & \frac{12}{13} & -\frac{3}{13} \\ \frac{12}{13} & -\frac{3}{13} & \frac{4}{13} \end{pmatrix}$$

In order to verify our result, notice that the rotation matrix has characteristic polynomial given by:

$$x^3 - \frac{19}{13}x^2 + \frac{19}{13}x - 1 = (x-1)\left(x^2 - \frac{6}{13}x + 1\right)$$

which has one real eigenvalue 1 and two complex eigenvalues. Moreover, notice that $E_1 = \langle(0,-3,1)^T\rangle$

Example 92 *Find the rotation matrix that sends $\vec{v}_1 \in \mathbb{R}^n$ to $\vec{v}_2 \in \mathbb{R}^n$ assuming that the two vectors are linearly independent.*

Solution: Since this is a rotation, then $|\vec{v}_1| = |\vec{v}_2|$. So one may assume that $|\vec{v}_1| = 1$ or replace both vectors $\vec{v}_i$ by $\vec{v}_i/|\vec{v}_i|$. Then, one may extend this linearly independent set to a basis for $\mathbb{R}^n = \langle\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\rangle$ and apply Gram-Schmidt to obtain an orthonormal basis $\mathbb{R}^n = \langle\vec{w}_1, \ldots, \vec{w}_n\rangle$. Write $\vec{v}_2 = y_1\vec{v}_1 + y_2\vec{v}_2$. Solve the system $R_\theta\vec{e}_1 = \vec{y}$ in order to find θ. Then the matrix of this rotation is:

$$[\vec{w}_1, \ldots, \vec{w}_n]\begin{bmatrix} R_\theta & 0_{2\times n-2} \\ 0_{n-2\times 2} & I_{n-2}\end{bmatrix}[\vec{w}_1, \ldots, \vec{w}_n]^{-1}$$

.

Example 93 *Let $S = \langle\vec{v}_1, \ldots, \vec{v}_m\rangle$ be a proper subspace of $\mathbb{R}^n$, i.e. $n > m$, where the given basis $\mathcal{A} = \{\vec{v}_1, \ldots, \vec{v}_m\}$ is orthonormal. Let $\mathcal{B} = \{\vec{w}_1, \ldots, \vec{w}_m\}$ be any other basis for S. Let $\mathcal{C} = \{\vec{u}_1, \ldots, \vec{u}_n\}$ be any basis for $\mathbb{R}^n$. Find the coefficient matrix of the linear transformation T that maps any vector $\vec{x}_\mathcal{C} \in \mathbb{R}^n_\mathcal{C}$ to $Proj_S(\vec{x}))_\mathcal{B} \in S_\mathcal{B}$.*

Solution: Notice that one possible set of steps that the vector $\vec{x}_\mathcal{C}$ undergoes to get to $Proj_S(\vec{x}))_\mathcal{C}$ is

$$\vec{x}_\mathcal{C} \mapsto \vec{x} \mapsto Proj_S(\vec{x})_\mathcal{A} \mapsto Proj_S(\vec{x})_\mathcal{B}$$

Hence the coefficient matrix of T can be obtained by:

$$M_{\mathcal{A}\to\mathcal{B}}[\vec{v}_i]^T_{m\times n}[\vec{u}_i]_{n\times n}$$

where one may find $M_{\mathcal{A}\to\mathcal{B}}$ by the process described in the change of basis matrix section.

10.4 Reflections

A basic example of an orthogonal reflection across a subspace is to reflect the vector $\vec{x} = (x_1, x_2)^T \in \mathbb{R}^2$ across the x-axis to obtain $Ref_{\vec{e}_1}\vec{x} = (x_1, -x_2)^T$. In this basic example, one finds that $\vec{x} \mapsto Ref_{vece_1}\vec{x}$ is a linear transformation with eigenvalues $1, -1$ both of multiplicity 1. Moreover, if one denotes $S = Span(\vec{e}_1)$, then one finds that the orthogonal complement is the y-axis, i.e. $S^\perp = Span(\vec{e}_2)$. Then

$$\vec{x} = Proj_S(\vec{x}) + Proj_{S^\perp}(\vec{x}) \Rightarrow Ref_S(\vec{x}) = Proj_S(\vec{x}) - Proj_{S^\perp}(\vec{x})$$

.

Theorem 28 (Definition-Theorem) *Let $S = \langle\underbrace{\vec{v}_1, \ldots, \vec{v}_m}_{\mathcal{B}}\rangle$ be a subspace of $\mathbb{R}^n$ given by an orthonormal basis $\mathcal{B}$. Assume $\mathbb{R}^n = \langle\underbrace{\vec{v}_1, \ldots, \vec{v}_n}_{\mathbb{B}}\rangle$ is an extension of this basis $\mathcal{B}$ to an ortho-normal basis $\mathbb{B}$ of $\mathbb{R}^n$. Hence $S^\perp = \langle\vec{v}_{m+1}, \ldots, \vec{v}_n\rangle$. Then the reflection of a vector $\vec{x}$ across S where*

$$\vec{x}_\mathbb{B} = x_1\vec{v}_1 + \ldots + x_m\vec{v}_m + x_{m+1}\vec{v}_{m+1} + \ldots + x_n\vec{v}_n$$

is given by

$$\vec{x}_\mathbb{B} = x_1\vec{v}_1 + \ldots + x_m\vec{v}_m + x_{m+1}\vec{v}_{m+1} + \ldots + x_n\vec{v}_n$$

Hence, the function $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\vec{x}) = Ref_S(\vec{x})$ is a linear transformation whose coefficient matrix is given by

$$[\vec{v}_1, \ldots, \vec{v}_n]Diag(\underbrace{1, \ldots, 1}_{m\ times}, \underbrace{-1, \ldots, -1}_{n-m\ times})[\vec{v}_1, \ldots, \vec{v}_n]^{-1}$$

Where $Diag(1, \ldots, 1, -1, \ldots, -1)$ is the $n \times n$ diagonal matrix with the first m-diagonal entries equal to 1, and the rest of the diagonal entries equal to -1. Moreover, its eigenvalues are 1 with multiplicity m and -1 with multiplicity $n-m$, and whose eigenspaces are $E_1 = S$ and $E_{-1} = S^\perp$.

11 Project 1: Generalizing the Dot Product

In Multivariable Calculus, we learned that two real vectors $\vec{v}$ and $\vec{w}$ of the same dimension are orthogonal if and only if their dot product is zero, i.e. $\vec{v} \cdot \vec{w} = 0$, where:

$$\vec{v} \cdot \vec{w} = v_1 w_1 + \ldots + v_n w_n = \sum_{i=1}^{n} v_i w_i = \vec{v}^T \cdot \vec{w}$$

Among the many interseting properties of the dot product, one has:

Proposition 18 *Given vectors $\vec{a}, \vec{b}, \vec{c} \in \mathbb{R}^n$ and a scalar $r \in \mathbb{R}$, one has:*

- *Positiveness $\vec{a} \cdot \vec{a} \geq 0$. Moreover, $\vec{a} \cdot \vec{a} = 0$ if and only if $\vec{a} = \vec{0}$.*
- *Commutativity or Symmetry: $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$.*
- *Distributive over sum: $\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$.*
- *Associative with scalar multiplication: $(r\vec{a}) \cdot \vec{b} = r(\vec{a} \cdot \vec{b}) = \vec{a} \cdot (r\vec{b})$.*

These properties guarantee that one can define a distance function or distance concept called **Norm**. Let $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$ define the norm of a vector. Notice that one requires positiveness in order to be able to compute the square root. Then the norm enjoys the following properties:

Proposition 19 *Let's define the norm of a vector like above, then*

- *The norm is non-negative, i.e. $\|\vec{v}\| \geq 0$ for any vector $\vec{v} \in \mathbb{R}^n$*
- *The norm is homogeneous, i.e. $\|r\vec{v}\| = |r| \cdot \|\vec{v}\|$ for any vector $\vec{v} \in \mathbb{R}^n$ and a scalar $r \in \mathbb{R}$, where $|r|$ denotes the absolute value.*
- *The norm satisfies the triangle inequality, i.e.: $\|\vec{a} + \vec{b}\| \leq \|\vec{a}\| + \|\vec{b}\|$.*
- *Geometric interpretation of the dot product: $\vec{a} \cdot \vec{b} = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \cos(\theta)$ where θ is the angle formed by the two vectors $\vec{a}$ and $\vec{b}$.*
- *Geometric interpretation of the norm: One may define the distance between two vectos by $d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\|$, and this agrees with the Euclidean distance given by the generalization of Pythagoras' theorem to n-dimensions.*

Example 94 *Derive the law of cosines for triangles using the definition and properties of the dot product and the norm.*

Solution: Let ABC be a triangle whose length of its sides are $a = |BC|, b = |AC|, c = |AB|$. Therefore,

$$\begin{aligned}
c^2 &= \|\vec{AB}\|^2 \\
&= \vec{AB} \cdot \vec{AB} \\
&= (\vec{CB} - \vec{CA}) \cdot (\vec{CB} - \vec{CA}) \\
&= \vec{CB} \cdot \vec{CB} - \vec{CB} \cdot \vec{CA} + \vec{CA} \cdot \vec{CB} + \vec{CA} \cdot \vec{CA} \\
&= \|\vec{CB}\|^2 - 2\vec{CB} \cdot \vec{CA} + \|\vec{CA}\|^2 \\
&= a^2 - 2\|\vec{CB}\|\|\vec{CA}\| \cos(\theta) + b^2 \\
&= a^2 + b^2 - 2ab\cos(\theta)
\end{aligned}$$

However, if one keeps this definition of the dot product for the complex Euclidean space $\mathbb{C}^n$, one does not enjoy all of these properties.

Example 95 *Let $\vec{a}, \vec{b} \in \mathbb{C}^n$. Define $\vec{a} \bullet \vec{b} := \sum_{i=1}^{n} a_i b_i$.*

- *Show with an example that $\vec{a} \cdot \vec{a}$ could be a negative number.*
- *Show with an example that $\vec{a} \cdot \vec{a}$ could be a complex number, and thus not comparable to 0.*

Solution:

- Let $\vec{a} = \begin{bmatrix} i \\ i \end{bmatrix}$. Then $\vec{a} \bullet \vec{a} = i \cdot i + i \cdot i = -2 \leq 0$.
- Let $\vec{a} = \begin{bmatrix} 1+i \\ 1+i \end{bmatrix}$. Then $\vec{a} \bullet \vec{a} = (1+i)^2 + (1+i)^2 = 4i$, which is not even a real number and thus it cannot be compared to 0, i.e. experessions like $4i > 0$ or $4i < 0$ are meaningless.

Define the dot product for complex vectors by:

$$\vec{v} \cdot \vec{w} = v_1\overline{w_1} + \ldots + v_n\overline{w_n} = \sum_{i=1}^{n} v_i\overline{w_i}$$

Then one has the following properties:

Proposition 20 *Given vectors $\vec{a}, \vec{b}, \vec{c} \in \mathbb{C}^n$ and a scalar $z \in \mathbb{R}$, one has:*

- *Positiveness $\vec{a} \cdot \vec{a} \geq 0$. Moreover, $\vec{a} \cdot \vec{a} = 0$ if and only if $\vec{a} = \vec{0}$.*
- *Conjugate Symmetry: $\vec{a} \cdot \vec{b} = \overline{\vec{b} \cdot \vec{a}}$.*
- *Distributive over sum: $\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$.*
- *Conjugate Associative with scalar multiplication: $(z\vec{a}) \cdot \vec{b} = z(\vec{a} \cdot \vec{b}) = \vec{a} \cdot (\overline{z}\vec{b})$.*

Remark 70 *Additionally, one could define a complex norm in the same manner, and it would have the same properties. However, one must be careful when computing absolute values, because the absolute value of a complex number $z = a+bi$ where $a, b \in \mathbb{R}$ is $|z| = \sqrt{z \cdot \overline{z}} = \sqrt{a^2 + b^2}$.*

11.1 Hermitian Forms

One possible generalization of the complex inner product are Hermitian forms. Let $\Phi : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ be a sesquilinear form given by:

$$\Phi(\vec{v}, \vec{w}) = \overline{\vec{v}}^T \cdot A\vec{w}$$

for some positive definite matrix $A_{n\times n}$.

A positive definite matrix $A_{n\times n}$ is a $\mathbb{C}$-matrix characterized by either of the following:

- For all vectors $\overline{\vec{x}}^T A\vec{x} \geq 0$.
- A is symmetric and all its eigenvalues are strictly positive real numbers.
- The determinant of all upper-left square submatrices are positive.

Example 96 *Given a square real matrix $A_{3\times 3} = (a_{ij})$ such that $a_{11} > 0$, $\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} > 0$ and $\det(A) > 0$. Show that A is positive definite matrix.*

Solution: This is just the third condition on the theorem.

11.2 Inner Products

An inner product in a space V is a function $\langle \bullet, \bullet \rangle : V \times V \to \mathbb{C}$ satisfying three properties:

- Positiveness: $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.
- Conjugate symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$.
- Linear in the first entry: $\langle a_1x_1 + a_2x_2, y \rangle = a_1\langle x_1, y \rangle + a_2\langle x_2, y \rangle$.

Example 97 *A popular example of inner products is the Hilbert Space of continuous functions in a compact interval. Let $f, g : [a, b] \to \mathbb{C}$ be two continuos functions on $[a, b]$, which it is denoted by $f, g \in C[a, b]$. Then their Hilbert product is given by:*

$$\langle f, g \rangle := \int_a^b f(z)\overline{g(z)}dz$$

Remark 71 *The condition of continuity could be relaxed to require finitely many removable or jump discontinuities, and one could apply this inner product to study periodic functions. This lead to Fourier expansions of such functions.*

12 Reviews

12.1 Review 307

1. **First Degree Ordinary Differential Equation - ODE**: $\frac{dy}{dt} = f(t,y)$. **Existence and Uniqueness** of solution to Initial Condition Problem ICP requires $f, \partial f/\partial y$ continuous on an interval containing $t_0 \in I$

   - **Separable**: $y' = M(t)N(y) \Rightarrow \int \frac{1}{N} dy = \int M dt + c$.
   - **Exact**: $M + Ny' = 0$ where $M_y = N_t$. Then $y = \int M dt + \int \left(N - \frac{\partial}{\partial y} \int M dt \right) dy + c$.

2. **Linear ODE**: $\frac{dy}{dt} + p(t)y = g(t)$. **Existence and Uniqueness** of solution requires p, g continuous on $t_0 \in I$.

   - **Integrating Factor** for Linear: μ such that $\mu' = \mu p \Rightarrow y = \frac{1}{\mu} \left(\int_{t_0}^{t} \mu g dt + c \right)$.

3. **Second Degree ODE**: $\frac{d^2y}{dt^2} = f(t, y, y')$

4. **Second Order Linear ODE**: $y'' + p(t)y' + q(t)y = g(t)$. Homogeneous if $g(t) = 0$. **Existence and Uniqueness** requires p, q, g continuous on $t_0 \in I$. The general solution is the sum of the general solution of the homogeneous counterpart plus a particular solution $y_g = h_{hg} + y_p$.

   - **Constant Coefficients Homogeneous**: $ay'' + by' + c = 0$. Solve $am^2 + bm + c = 0$.

     - Real Solutions $r_1 \neq r_2$: $y = Ae^{r_1 t} + Be^{r_2 t}$.
     - Real Solutions equal $r_1 = r_2 = r$: $y = Ate^{rt} + Be^{rt}$.
     - Complex Conjugate Solutions $\alpha \pm \beta i$: $y = e^{\alpha t}[A\cos(\beta t) + B\sin(\beta t)]$.
   - Important Theorems:
     - **Superposition** for Homogeneous: Any linear combination of solutions is a solution.
     - **Wronskian**: $W(y_1, y_2) = y_1 y_2' - y_1' y_2 = \det \begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix}$. **Abel's Theorem**: $W = ce^{-\int p\,dt}$ for some c.
     - **Fundamental Set of Solutions** or *basis of solutions*: $y_g = c_1 y_1 + c_2 y_2$. Either $W(y_1, y_2)(t_0) \neq 0$ for some $t_0 \in I$. In particular $W(y_1, y_2)(t_0) = I_2$.
   - **Reduction of Order** for homogeneous: Knowing one solution y_1, assume $y_2 = vy_1$, then $y_1 v'' + (2y_1' + py_1)v' = 0$. Solve for v' and then integrate.
   - **Undetermined Coefficients** for Constant coefficients and $g(t)$ of specific kinds or types:
     - $g(t) = P_n(t) \Rightarrow y_p(t) = U_n(t)$. Except if 1, or 1 and t are already solutions of the homogeneous. Then multiply by t or t^2 respectively.
     - $g(t) = e^{\alpha t} P_n(t) \Rightarrow y_p = e^{\alpha t} U_n(t)$. Except if $e^{\alpha t}$, or $e^{\alpha t}$ and $te^{\alpha t}$ are already solutions of the homogeneous. Then multiply by t or t^2 respectively.
     - $g(t) = e^{\alpha t} P_n(t)\cos(\beta t)$ or $e^{\alpha t} P_n(t)\sin(t)$ or a linear combination $\Rightarrow y_p = e^{\alpha t} U_n(t)\cos(\beta t) + e^{\alpha t} V_n(t)\cos(\beta t)$. Except if $e^{\alpha t}\cos(\beta t)$ and $e^{\alpha t}\sin(\beta t)$ are already solutions of the homogeneous. Then multiply by t.
   - **Variation of Parameters**: $y_{gh} = c_1 y_1 + c_2 y_2$. Assume $y_p = u_1 y_1 + u_2 y_2$ with $u_1' y_1 + u_2' y_2 = 0$, and then

     $$y_p = -y_1 \int \frac{y_2 g}{W} dt + y_2 \int \frac{y_1 g}{W} dt$$

   - **Euler 2nd Order Equation**: $t^2 \frac{d^2y}{dt^2} + \alpha t \frac{dy}{dt} + \beta y = 0$ for $t > 0$. Replace $t = e^x$ or $x = \ln(t)$ and the equation transforms $\frac{d^2y}{dx^2} + (\alpha - 1)\frac{dy}{dx} + \beta y = 0$. Then the solution is $y(x) = y(\ln(t))$.
   - **Laplace Transform**:
     - $L(f')(s) = sL(f)(s) - f(0)$.
     - Tail Bounded previous derivatives: $L(f^{(n)})(s) = s^n L(f)(s) - \sum_{i=1}^{n-1} s^i f^{(n-i)}(0)$.
     - $L(u_c(t)f(t-c)) = e^{-cs}L(f)(s)$. Moreover if $f(t) = L^{-1}(F)$, then $L^{-1}(e^{-cs}F(s))(t) = u_c(t)f(t-c)$.
     - $L(e^{ct}f(t)) = L(f)(s-c)$ for $s > a + c$. Moreover if $f(t) = L^{-1}(F)$, then $L^{-1}(F(s-c)) = e^{ct}f(t)$.
     - Table 6.2.1 on page 321.

12.2 Review 308

1. **Equivalence of Systems**:
   (1) Linear System of Equations: $\forall i : \sum_j a_{ij}x_j = b_j$.
   (2) Vector Equation: $\sum_j \vec{a}_j x_j = \vec{b}_j$.
   (3) Matrix System Equation $A\vec{x} = \vec{b}$.

2. **Gauss and Jordan Gauss Elimination**: Using Row Elementary Operations to obtain either a stair-shaped System, or a stair-shaped system with leading coefficients 1 and pivot columns full of zeros expect by a one.

3. All Systems have either 0, 1, or infinitely many solutions. The general solution of a non-homogeneous system is equal to a particular solution plus the general solution of the homogeneous counterpart: $y_g = y_{gh} + y_p$.

4. **Minors**: $\det(A_{ij})$ where A_{ij} is the submatrix obtained by deleting the i-th row and j-th column. **Cofactor** $C_{ij}(A) = (-1)^{ij}\det(A_{ij})$ or signed minor. **The Cofactor Matrix**: the matrix whose ij-th entry is the ij-th cofactor $C_{ij}(A)$. The **Adjugate Matrix**: the transpose of the cofactor matrix.

5. **The Determinant**:
$$\det(A) = \sum_{\sigma \in S_n} sgn(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}$$
$$\det(A) = \sum_{j=1}^{n} a_{ij}C_{ij}(A)$$

6. Inverse of a Matrix:
   (1) $[A|I]$, apply Jordan Gauss and obtain $[I|A^{-1}]$.
   (2) $Adj(A) \times A = \det(A)I_n$.

7. **Linear Independence**: $\sum_i x_i\vec{a}_i = \vec{0}$ has a unique solution $x_1 = x_2 = \ldots = 0$.

8. **Span of a list**: $Span(\vec{a}_1, \vec{a}_2, \ldots)$ is the set of all linear combinations (finite sums of constant multiples) of the vectors in the list.

9. **Eigenvalues, Eigenvectors, and Eigenspaces**:
   (1) Roots of Characteristic Polynomial $\det(A - xI) = 0$ are eigenvalues.
   (2) For each eigenvalue, $E_\lambda = Null(A - \lambda I)$.
$$1 \le dim(E_\lambda) \le mult(\lambda)$$
   $dim(E_\lambda)$ is called the **geometric multiplicity**, and $mult(\lambda)$ is called the algebraic multiplicity.

10. **Big Theorem**: For a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ and the coefficient or associated matrix $A = [\vec{a}_j]_j$ where $\vec{a}_j \in \mathbb{R}^n$. That means $T(\vec{x}) = A\vec{x}$ where $A = (a_{ij})$ has by j-th column the vector $\vec{a}_j$. That means $T(\vec{0}_m) = \vec{0}_n$ and $T(a\vec{x} + b\vec{y}) = aT(\vec{x}) + bT(\vec{y})$.
    - $Rank(A)+Nullity(A) = dim(Ker(T))+dim(Range(T)) = m$. Therefore, $Nullity(A) = dim(Ker(T)) \le m$ and $Rank(A) = dim(Range(T)) \le \min(m, n)$.
    - **Injective** or one-to-one: Requires $m \le n$ and equivalent to any of the following: $E_0 = Null(A) = Ker(T) = \{\vec{0}\}$; $Nullity(A) = dim(Ker(T)) = 0$; $\{\vec{a}_i : i = 1, \ldots, m\}$ is LI; all systems $A\vec{x} = \vec{b}$ have at most one solution; 0 is not an eigenvalue. Computational Test: Jordan Gauss on A and the number of equations is equal to the number of variables.
    - **Surjective** or onto: Requires $n \le m$ and equivalent to any of the following: $Col(A) = Range(T) = Span(\vec{a}_i)_i = \mathbb{R}^n$; $Rank(A) = dim(Range(T)) = n$; all systems $A\vec{x} = \vec{b}$ have at least one solution. Computational Test: Jordan Gauss on A and the number of pivot columns is equal to n.
    - **Bijective**: Requires $m = n$ and equivalent to any of the following: Injective; Surjective; A is invertible or non-singular, $\det(A) \neq 0$, $\langle \vec{a}_i \rangle_i$ is a basis for $\mathbb{R}^n$.

11. **Affine Transformation**: $T : \mathbb{R}^m \to \mathbb{R}^n$ with $T(\vec{x}) = A\vec{x} + \vec{b}$. The image is not a subspace unless $\vec{b} = \vec{0}$.

12. The dot product $\vec{x} \cdot \vec{y} = \vec{x}^T \times \vec{y}$ (product as matrices) and $\vec{x} \perp \vec{y} \Leftrightarrow \vec{x} \cdot \vec{y} = 0$.

13. Gram-Schmidt: Given $\vec{x}_1, \dots, \vec{x}_n \in \mathbb{R}^m$. Then let $\vec{s}_1 = \vec{x}_1$ and $\forall k > 1$:

$$\vec{s}_k = \vec{x}_k - \sum_{i=1}^{k-1} Proj_{\vec{s}_i}(\vec{x}_k) = \vec{x}_k - \sum_{i=1}^{k-1} \frac{\vec{s}_i \cdot \vec{x}_k}{\|\vec{s}_i\|^2} \vec{s}_i$$

It returns a list of pairwise orthogonal vectors with $\forall 1 \le d \le m : Span(\vec{x}_1, \dots, \vec{x}_d) = Span(\vec{s}_1, \dots, \vec{s}_d)$.

12.3 Formulas for 309

1. **General Solution to the Homogeneous** System $\vec{x}' = A\vec{x}$:

   - Two different real eigenvalues and corresponding eigenvectors:

     $$\vec{x}_{gh} = C_1 \vec{v}_1 e^{\lambda_1 t} + C_2 \vec{v}_2 e^{\lambda_2 t}$$

     Classification of $(0,0)$: if $\lambda_1, \lambda_2 < 0$: asymptotically stable node; if $\lambda_1, \lambda_2 > 0$: unstable node; $\lambda_1 > 0 > \lambda_2$; (unstable) saddle point. If $\lambda_1 = \lambda_2$ but $\vec{v}_1 \neq \vec{v}_2$, then Star.

   - Two complex conjugate eigenvectors $\lambda = \alpha \pm \beta i$ and $\vec{v} = \vec{a} + i\vec{b}$.

     $$\vec{x}_{gh} = C_1 e^{\alpha t} \left[\vec{a}\cos(\beta t) - \vec{b}\sin(\beta t) \right] + C_2 e^{\alpha t} \left[\vec{a}\sin(\beta t) + \vec{b}\cos(\beta t) \right]$$

     Classification of $(0,0)$: if $a = 0$, then center; if $a > 0$, then unstable spiral; if $a < 0$, then asymptotically stable spiral.

   - Single Eigenvalue - Single eigenvector: Find generalized eigenvector $\vec{w}$ from the eigenvector $\vec{v}$ by solving $(A - \lambda I)\vec{w} = \vec{v}$.

     $$\vec{x}_{gh} = C_1 \vec{v} e^{\lambda t} + C_2 \left(\vec{v} t e^{\lambda t} + \vec{w} e^{\lambda t} \right)$$

     Classification of $(0,0)$: if $\lambda > 0$, unstable improper node, if $\lambda < 0$, asymptotically stable improper node.

2. **Fundamental Matrix** for Homogeneous: $\Phi = \left(\vec{x}^{(1)}, \vec{x}^{(2)} \right)$ where the columns are two independent solutions.

3. **General Solution to the Special Homogeneous** system $t\vec{x}' = A\vec{x}$ and $t > 0$:

   - Two different real eigenvalues and corresponding eigenvectors:

     $$\vec{x}_{gh} = C_1 \vec{v}_1 t^{\lambda_1} + C_2 \vec{v}_2 t^{\lambda_2}$$

   - Two complex conjugate eigenvectors $\lambda = \alpha \pm \beta i$ and $\vec{v} = \vec{a} + i\vec{b}$.

     $$\vec{x}_{gh} = C_1 t^{\alpha} \left[\vec{a}\cos(\beta \ln(t)) - \vec{b}\sin(\beta \ln(t)) \right] + C_2 t^{\alpha} \left[\vec{a}\sin(\beta \ln(t)) + \vec{b}\cos(\beta \ln(t)) \right]$$

   - Single Eigenvalue - Single Eigenvector: Find the generalized eigenvector $\vec{w}$:

     $$\vec{x}_{gh} = C_1 \vec{v} t^{\lambda} + C_2 \left(\vec{v} t^{\lambda} \ln(t) + \vec{w} t^{\lambda} \right)$$

4. **Diagonalization** or **Jordan Eigenform** for nonhomogeneous $\vec{x}' = A\vec{x} + \vec{g}(t)$:

   - Diagonalizable: eigenvalues λ_1, λ_2 and eigenvectors $\vec{v}_1, \vec{v}_2$. Then $T = (\vec{v}_1, \vec{v}_2), D = diag(\lambda_1, \lambda_2)$: Assume $\vec{x} = T\vec{y}$. Then $\vec{y}' = D\vec{y} + T^{-1}\vec{g}$. Both equations will be linear on a single variable. Recall that $y' + p(t)y = g(t)$ solves as $y = e^{-\int p(t)dt} \int e^{\int p(t)dt} g(t) dt$.
   - Nondiagonalizable: eigenvalue λ, eigenvector $\vec{v}$, generalized eigenvector: $\vec{w}$. Then $T = (\vec{v}, \vec{w})$ and $J = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$: Assume $\vec{x} = T\vec{y}$. Then $\vec{y}' = J\vec{y} + T^{-1}\vec{g}$. Solve first the second equation which only involves y_2, and then plug it in the first equation. Both equations will be then of the form $y' + p(t)y = g(t)$. Solve it as in the previous case.

   At the end, we can assume all constants of integration zero and obtain the particular solution $\vec{y}_p$, and $\vec{x}_p = T\vec{y}_p$.

5. **Variations of Parameters** for the non-homogeneous system: Assume Ψ is a fundamental matrix, where each column is an element of a fundamental set of solutions for the homogeneous corresponding system.

$$\vec{x}_p = \Psi(t) \int \Psi^{-1}(t)\vec{g}(t)dt$$

Note: the indefinite integral happens on each entry of the product $\Psi^{-1}(t)\vec{g}(t)$.

6. **Exponential of a Matrix**: Given Matrix A, we want to find $\exp(At)$.

   - Step 1: Solve $\vec{x}' = A\vec{x}$. The general solution is of the form $\vec{x}_{gh}(t) = C_1\vec{x}_1(t) + C_2\vec{x}_2(t)$.
   - Step 2: Find constants α_1, α_2 (replacing the parameters C_1, C_2) so that $\vec{x}_{gh}(t=0) = \vec{e}_1 = (1,0)^T$. Then $\alpha_1\vec{x}_1(t) + \alpha_2\vec{x}_2(t)$ will be the first column of the solution.
   - Step 3: Find constants β_1, β_2 (replacing the parameters C_1, C_2) so that $\vec{x}_{gh}(t=0) = \vec{e}_2 = (0,1)^T$. Then $\beta_1\vec{x}_1(t) + \beta_2\vec{x}_2(t)$ will be the second column of the solution.

$$\exp(At) = \begin{pmatrix} \alpha_1\vec{x}_1(t) + \alpha_2\vec{x}_2(t) & \beta_1\vec{x}_1(t) + \beta_2\vec{x}_2(t) \end{pmatrix}$$

   - If A is diagonalizable, then $A = TDT^{-1} \Rightarrow \exp(At) = T\exp(Dt)T^{-1}$ where $\exp(diag(d_1, d_2)t) = diag(e^{d_1 t}, e^{d_2 t})$.

7. **Fourier Series Expansion**: The set $\{\sin(n\pi x/L), \cos(n\pi x/L) : n \in \mathbb{Z}^+\}$ is a basis for the set of periodic functions f of period $2L$ where f, f' are piecewise continuous (ignoring the points of discontinuity, where the series converges to the average of the left and right limit of the function).

$$f(x) = \frac{a_0}{2} + \sum_{n\geq 1} a_n \cos(n\pi x/L) + \sum_{n\geq 1} b_n \sin(n\pi x/L)$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos(n\pi x/L)dx$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin(n\pi x/L)dx$$

Moreover, if f is even, then $\forall n : b_n = 0$; and if f is odd, then $\forall n : a_n = 0$. Also Parseval's identity:

$$\frac{1}{L}\int_{-L}^{L} f(x)^2 dx = \frac{a_0^2}{2} + \sum_{n\geq 1}(a_n^2 + b_n^2)$$

8. **Eigen-pairs**: $y'' + \lambda y = 0$.

   - $y(0) = 0$ and $y(L) = 0$: $(\lambda_n, y_n)_{n>0} = (n^2\pi^2/L^2, \sin(n\pi x/L))$.
   - $y(0) = 0$ and $y'(L) = 0$: $(\lambda_n, y_n)_{n:odd>0} = (n^2\pi^2/4L^2, \sin(n\pi x/2L))$.
   - $y'(0) = 0$ and $y(L) = 0$: $(\lambda_n, y_n)_{n:odd>0} = (n^2\pi^2/4L^2, \cos(n\pi x/2L))$.
   - $y'(0) = 0$ and $y'(L) = 0$: $(\lambda_n, y_n)_{n\geq 0} = (n^2\pi^2/L^2, \cos(n\pi x/L))$.

9. **Extensions**:

   - f symmetric about $x = L$ on the interval $(0, 2L)$, i.e.: $f(2L-x) = f(x)$; odd, and of period $4L$ makes f have a Fourier expansion involving only $\sin(\pi nx/2L)$ where n is odd with coefficients $\frac{2}{L}\int_0^L f(x)\sin(\pi nx/2L)dx$
   - f anti-symmetric about $x = L$ on the interval $(0, 2L)$, i.e.: $f(2L - x) = -f(x)$; even, and of period $4L$ makes f have a Fourier expansion involving only $\cos(\pi nx/2L)$ where n is odd with coefficients $\frac{2}{L}\int_0^L f(x)\cos(\pi nx/2L)dx$

12.3.1 Applications of Separation of Variables

10. **Heat Equation**: $\alpha^2 u_{xx} = u_t$ where the initial temperature is known $u(x,0)$ for $x \in [0,L]$.

    (a) **Endpoints at constant temperatures** BC: $u(0,t) = T_1, u(L,t) = T_2$ (Non-homogeneous BC - Stabilizing assumption)

    $$u(x,t) = \frac{T_2 - T_1}{L}x + T_1 + \sum_{n\geq 1} c_n \exp\left(-\frac{\alpha^2 n^2 \pi^2}{L^2}t\right)\sin\left(\frac{n\pi x}{L}\right);$$

    Extend $u(x,0) - \left(\frac{T_2-T_1}{L}x + T_1\right)$ to be odd and of period $2L$.

    $$c_n = \frac{2}{L}\int_0^L \left[u(x,0) - \left(\frac{T_2 - T_1}{L}x + T_1\right)\right]\sin\left(\frac{n\pi x}{L}\right)dx$$

    (b) **Insulated at both Endpoints** BC: $u_x(0,t) = 0, u_x(L,t) = 0$ (Separation of Variables) Extend $u(x,0)$ to be even of period $2L$, then

    $$u(x,t) = \frac{a_0}{2} + \sum_{n\geq 1} a_n \exp\left(-\frac{\alpha^2 n^2 \pi^2}{L^2}t\right)\cos\left(\frac{n\pi x}{L}\right)$$

    $$a_n = \frac{2}{L}\int_0^L f(x)\cos(n\pi x/L)dx$$

    (c) **One endpoint fixed and the other is insulated**:

    i. BC: $u(0,t) = T$ and $u_x(L,t) = 0$:

    $$u(x,t) = T + \sum_{n:odd>0} c_n \sin\left(\frac{n\pi x}{2L}\right)e^{-\frac{n^2\pi^2\alpha^2}{4L^2}t}$$

    $$c_n = \frac{2}{L}\int_0^L (f(x) - T)\sin\left(\frac{n\pi x}{2L}\right)dx$$

    Extend $u(x,0) - T$ symmetric about $x = L$ on $(0,2L)$, odd, and of period $4L$.

    ii. BC: $u_x(0,t) = 0$ and $u_x(L,t) = T$:

    $$u(x,t) = T + \sum_{n:odd>0} c_n \cos\left(\frac{n\pi x}{2L}\right)e^{-\frac{n^2\pi^2\alpha^2}{4L^2}t}$$

    $$c_n = \frac{2}{L}\int_0^L (f(x) - T)\cos\left(\frac{n\pi x}{2L}\right)dx$$

    Extend $u(x,0) - T$ anti-symmetric about $x = L$ on $(0,2L)$, even, and of period $4L$.

11. **1-dimensional Wave Equation**: $a^2 u_{xx} = u_{tt}$ where a is the velocity of propagation of the wave. BC: $u(0,t) = u(L,t) = 0$; $u(x,0) = f(x)$; $u_t(x,0) = g(x)$. The general solution is of the form:

    $$u(x,t) = \sum_n \sin\left(\frac{n\pi x}{L}\right)\left[c_n \cos\left(\frac{n\pi a t}{L}\right) + d_n \sin\left(\frac{n\pi a t}{L}\right)\right]$$

    Extend both f and g to be odd and of period $2L$.

    $$c_n = b_n(f) = \frac{2}{L}\int_0^L f(x)\sin(n\pi x/L)dx; d_n = \frac{L}{n\pi a}b_n(g) = \frac{2}{n\pi a}\int_0^L g(x)\sin(n\pi x/L)dx$$

12. **Laplace's Equation**:

    (a) **Dirichlet Problem for a Rectangle**: $u_{xx} + u_{yy} = 0$ for $(x,y) \in (0,A) \times (0,B)$.

i. BC: $u(x,0)=0, u(x,B)=0, u(0,y)=f(y), u(A,y)=g(y)$ for $y\in[0,B]$.

$$u(x,y)=\sum_{n\geq 1}\left[c_n\sinh\left(\frac{n\pi x}{B}\right)+d_n\sinh\left(\frac{n\pi(x-A)}{B}\right)\right]\sin\left(\frac{n\pi y}{B}\right)$$

Extend both f and g to be odd and of period $2B$.

$$c_n\sinh\left(\frac{n\pi A}{B}\right)=\frac{2}{B}\int_0^B g(y)\sin\left(\frac{n\pi y}{B}\right)dy$$

$$-d_n\sinh\left(\frac{n\pi A}{B}\right)=\frac{2}{B}\int_0^B f(y)\sin\left(\frac{n\pi y}{B}\right)dy$$

ii. By symmetry of the expressions: if BC: $u(x,0)=i(x), u(x,B)=j(x), u(0,y)=0, u(A,y)=0$ for $x\in[0,A]$.

$$u(x,y)=\sum_{n\geq 1}\left[e_n\sinh\left(\frac{n\pi y}{A}\right)+f_n\sinh\left(\frac{n\pi(y-B)}{A}\right)\right]\sin\left(\frac{n\pi x}{A}\right)$$

Extend both f and g to be odd and of period $2A$.

$$e_n\sinh\left(\frac{n\pi B}{A}\right)=\frac{2}{A}\int_0^A j(x)\sin\left(\frac{n\pi x}{A}\right)dx$$

$$-f_n\sinh\left(\frac{n\pi B}{A}\right)=\frac{2}{A}\int_0^A i(x)\sin\left(\frac{n\pi x}{A}\right)dy$$

iii. Combining: if BC: $u(x,0)=i(x), u(x,B)=j(x), u(0,y)=f(y), u(A,y)=g(y)$, then just add the two previous solutions.

(b) **Dirichlet Problem for the Circle** centered at $(0,0)$ and of radius $r=a$: $u_{rr}+\frac{1}{r}u_r+\frac{1}{r^2}u_{\theta\theta}=0$ with IC $u(a,\theta)=f(\theta)$.

$$u(r,\theta)=\frac{c_0}{2}+\sum_{n\geq 1}r^n(c_n\cos(n\theta)+k_n\sin(n\theta))$$

where $a^nc_n=\frac{1}{\pi}\int_0^{2\pi}f(\theta)\cos(n\theta)d\theta$ and $a^nk_n=\frac{1}{\pi}\int_0^{2\pi}f(\theta)\sin(n\theta)d\theta$

13. Useful Integrals:

$$\begin{aligned}
\int x^2\cos(bx)\,dx &= \tfrac{1}{b}x^2\sin(bx)+\tfrac{2}{b^2}x\cos(bx)-\tfrac{2}{b^3}\sin(bx)\\
\int x\cos(bx)\,dx &= \tfrac{1}{b}x\sin(bx)+\tfrac{1}{b^2}\cos(bx)\\
\int x^2\, sin(bx)\,dx &= -\tfrac{1}{b}x^2\cos(bx)+\tfrac{2}{b^2}x\sin(bx)+\tfrac{2}{b^3}\cos(bx)\\
\int x\sin(bx)\,dx &= -\tfrac{1}{b}x\cos(bx)+\tfrac{1}{b^2}\sin(bx)
\end{aligned}$$