LINEAR ALGEBRA

ESSENCE OF THE SUBJECT.

BY

SHAURYA PRATAP SINGH

Department of Computer Science and Engineering
DIT University
Dehradun

Foreword

The following are my notes on Linear Algebra, as I have understood the subject so far. I am far from an expert on the subject and it is presumed that the reader has an elementary understanding of coordinate geometry, vectors, and matrix representation and multiplication. I shall try my very best to explain the concepts related to the subject, but please spare me the criticism in case of my failure to explain well. The notes are comprehensive but the order of introduction of topics isn't perfect. I have tried to incorporate some python code also, but, it is not very sophisticated since learning programming hasn't been my prime priority, but it does help understand the mathematical examples better. Above all these notes can be a good reference if not some good introductory material.

Contents

Chapter 1

Introduction

1.1 Prerequisites of Linear Algebra

In mathematics, algebra deals with **variables** that are unknown quantities on which we operate so as to find values, that satisfy certain equations. An **equation** can be defined as an equality statement containing one or more variables. Solving an equation or more appropriately a **system of equations** helps us determine the **solutions** of that system, or in other words, what values of the variables would make that equality true.

Algebra studies two main families of equations:

- **Polynomials:** Expressions that can be built from constants and variables by the means of addition, multiplication and exponentiation to a non-negative integer power.

  Polynomial equation: $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = 0$
  Summation notation: $\sum_{k=0}^{n} a_k x^k = 0$

  Each term is a product of a **coefficient** and a variable raised to some non-negative integer power known as the **degree** of the term.

- **Linear Equations:** Polynomials that have a degree of at-most 1.
  Example: $ax + by + cz = 0$
  A more intuitive way to think about linear equations is that the relationship among it's variables can be expressed in the form of a **line**.

The usefulness of linear equations lies in the fact that many non-linear systems in the real world can be studied by breaking them down into much simpler linear systems. (**linearization**)

1.1.1 Common forms of Two-Dimensional Linear Equations

Often called *equations on a straight line*, all of the equations can be geometrically interpreted as a line in the xy plane.

- **General Form (Standard Form):** $Ax + By = C$
  where $A, B \neq 0$
- **Slope-Intercept Form:** $y = mx + c$
  where m = slope, c = $y-$intercept.
- **Point-Slope Form:** $y - y_0 = m(x - x_0)$
  where m = slope, (x_0, y_0) = a point on the line.
- **Two-Point Form:** $y - y_0 = \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$
  (x_0, y_0) and (x_1, y_1) are two points on a line with $x_0 \neq x_1$
- **Intercept Form:** $\frac{x}{a} + \frac{y}{b} = 1$
  where a, b are x and y intercepts respectively and $a, b \neq 0$
- **Parametric Form:** $x = (p - h)t + h$ and $y = (q - k)t + k$
  Two simultaneous equations are expressed in terms of a variable parameter t.
- **Matrix Form:** For the equations $A_1x + B_1y = C_1$ and $A_2x + B_2y = C_2$. The matrix form is given by: $\begin{bmatrix} A_1 & B_1 \\ A_2 & B_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$

It is crucial so as to interpret what form of a linear equation we are provided and how can we change from one form to another as per our requirements. Linear Algebra will be mostly concerned with using the *Standard Form* and the *Matrix Form.*

1.2 Why Linear Algebra?

Linear Algebra is that branch of mathematics that deals with the study of linear equations and their representations through **matrices** and **vector spaces.**

As we will come to know, the ideas of linear algebra can be applied to almost all of mathematics. It is also heavily used in the domains of science and engineering due its ability to efficiently model and solve natural phenomena and provide approximations for non-linear phenomena.

Some domains that find application of Linear Algebra: Linear Programming, Quantum Mechanics, Machine Learning, Computer Graphics, Finite Element Analysis.

Chapter 2

Vectors

Before we proceed to learn about vectors, I would like to stress upon that necessity for understanding both the geometric interpretation and the numerical interpretation of the subject. Geometric interpretation is crucial because it helps us build an intuition so as to why a concept works in the first place by our ability to visualize it, and numeric interpretation is an absolute necessity after that so as to abstract the objects we operate on and focus only on our method.

Another important fact to elucidate is that Linear Algebra is a very practical subject and that makes it very crucial to learn the computational aspect of the subject in a programming language. Thus, we will also learn how to visualize and solve our problems in python.

2.1 Three Intuitions About Vectors

There is no simple answer to the question "What is a vector?" but here are some ideas so as to provide an intuitive idea behind what a vector is before we formalize the idea of vectors and vector spaces.

- **Physics:** Vectors in physics are quantities that have a certain direction and magnitude. They are depicted by **arrows** pointing in a space with a given length and direction. They are used to depict forces, fields etc.

- **Computer Science:** Ordered Lists of numbers used for representing data. Example: Real-Estate data (land area, cost, location, ...) The vectors have an n number of variables (It's components).

- **Mathematics:** A vector can be anything that obeys a set of axioms (which we shall describe later). Informally speaking, a vector could be anything having a sensible notion of adding another vector to it or scaling it by some number.

Some important points:

- This **abstract** definition of a vector is one of the central ideas of linear algebra.
- When we geometrically represent vectors, one difference from vectors in physics would be that all of our vectors would be tailed at origin.

2.2 Geometric Interpretation of Vectors

A vector in a two dimensional space has two **components** v_x, v_y that represent it's spacial arrangement. It is written as: $\vec{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}$

It can also be represented as $\vec{v} = (v_x, v_y)$.

The interesting thing to note here is that vector can be broken down as follows, $v = c\hat{v_x} + d\hat{v_y}$, where $\hat{v_x}, \hat{v_y}$ are the unit vectors or more generally speaking the **basis vectors** and c, d are **scalars.** We shall only deal with the special case of basis vectors i.e unit vectors for now, changing the basis vectors would essentially change our coordinate system.

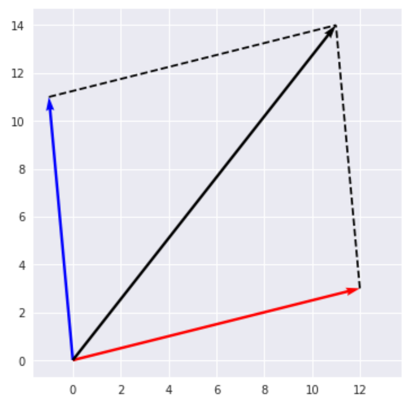

Figure 2.1: Vectors in a 2-D plane.

- **Unit Vectors:** These are vectors of unit magnitude in the direction of the coordinate axes in both 2-D and 3-D space and are represented by $\hat{i}, \hat{j}, \hat{k}$.
- **Scalars:** Scalars are just numbers that allow us to *stretch* or *squish* any vector, thus the name.

From this we can easily conclude:

- Any vector with no matter how many components is just a *sum of scaled unit vectors.*
- Adding two vectors means adding the *scaled components*.

These two key facts are central to linear algebra as they define what is known as taking a linear combination:

Linear Combination: $c\vec{u} + d\vec{v} = c \begin{bmatrix} u_x \\ u_y \end{bmatrix} + d \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} cu_x + dv_x \\ cu_y + dv_y \end{bmatrix}$

The visual image of taking linear combinations can be thought of as moving in the coordinate plane in the direction described by the vectors. Where ever we stop at the end, are the coordinates of the resultant vector.

Here is a small python program that can plot Fig 2.1

```
%matplotlib inline

# imports for pretty plots
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# default settings for seaborn
sns.set()

# u, v: 2D vectors
def vectorplot2D(u, v):
    plt.figure(figsize=(6,6))

    # calculate the resultant
    r = [u[0] + v [0], u[1] + v[1]]

    # quiver allows us to plot arrowheads
    plt.quiver([0, 0, 0],
               [0, 0, 0],
               [u[0], v[0], r[0]],
               [u[1], v[1], r[1]],
               angles='xy',
               scale_units='xy',
               scale=1,
               color=['r', 'b', 'k'])

    # generation of a range for dotted lines
    l1 = [np.linspace(v[0], r[0]),
          np.linspace(v[1], r[1])]

    l2 = [np.linspace(u[0], r[0]),
          np.linspace(u[1], r[1])]

    # plot dotted lines
    plt.plot(l1[0], l1[1], 'k--')
    plt.plot(l2[0], l2[1], 'k--')

    # maintain the ratio of visible area to the plot
    plt.axis('square')
    plt.show()

vectorplot2D([12, 3],[-1, 11])
```

Given the definition of linear combinations, I want you to pause and ponder. If we are allowed to freely select the scalars and vectors i.e not fix them to a particular value, what would the geometric image look like? (For vectors with 2 components)

It would be something in which we could obtain every possible 2-D vector or in other words, a plane that would contain each and every resultant vector. For the 2-D coordinate system, it would be the xy plane. In case of the 3-D system, the vectors are free to move in the three dimensional space.

Another important thing to know is that the above statement is not complete by itself, in case one of the two vectors is the scaled version of another i.e $\vec{v}, c\vec{v}$, their linear combinations would only result in another scaled version of vector $\vec{v}$ forming a line. Thus for two vectors to form a 2-D plane, its absolutely necessary for them to be **distinct.**

This situation can also arise in the case of 3-D coordinates. Given three vectors, if the third vector lies in the plane formed by the linear combinations of the first two, the three vectors will only **span** a plane and not the entire three dimensional space. For them to cover the entire 3-D space, it is necessary that the three vectors are **linearly independent.** (Explained later in Vector Spaces.)

2.3 Dot Products

Geometrically speaking, dot products can be defined as taking the projection of one vector on another and then scaling it by the magnitude of the second vector. When the direction of the two vectors is opposite, it yields a negative dot product and **zero** when they are perpendicular. The notion for dot product is as follows:

Cosine Formula: $\vec{u}.\vec{v} = ||\vec{u}||\ ||\vec{v}|| \cos\theta$

A noticeable fact here is that the order does not matter while taking the dot product and that has a very interesting geometric explanation.

Think about it this way, suppose we were to take dot products of two unit vectors $\vec{u}, \vec{v}$, the order would not matter because both of them have equal magnitude, thus projection would be same on one another. Now if they were not unit vectors, but had a scalar also, it really wouldn't matter where the scalar is because it would either scale the projection that is to be applied, or in the other case, it would scale the magnitude to which the projection is applied.

$\vec{u}.\vec{v} = K||\hat{u}||\ ||\hat{v}|| \cos\theta$; It doesn't matter where the K comes from.

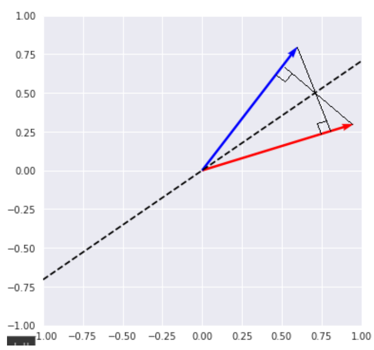

Figure 2.2: Symmetry in dot products.

Dot product in terms of individual vector components can be written as:

Dot Product: $\vec{u}.\vec{v} = \begin{bmatrix} u_x \\ u_y \end{bmatrix} \cdot \begin{bmatrix} v_x \\ v_y \end{bmatrix} = u_x v_x + u_y v_y$

Now it might look very odd and one might think, "How does that have something to do with projections?" but it can be explained easily by having an intuitive understanding of what is known as **linear transformations.**

2.3.1 An Intuitive Idea of Linear Transformations

A formal introduction to linear transformations will be covered in the later chapters. Informally, a linear transformation can be thought of as a function that takes a vector as an argument and spits out another changed vector. If we were to imagine a 2-D space with grid lines, any transformation to this space would not change their *straight* nature i.e. preserve the linearity of the space and the origin would still remain fixed.

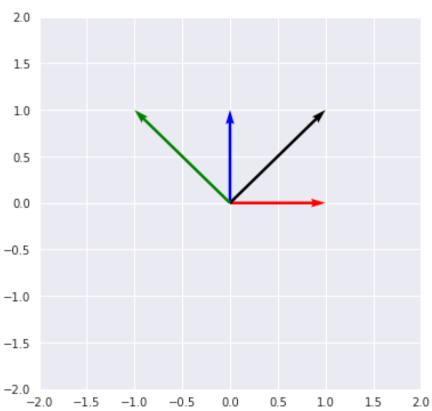

Figure 2.3: Green: Transformed Vector $\vec{w}$

For transformations to occur numerically, one needs to know where our basis $(\hat{i}, \hat{j})$ will land on on our grid once the transformation takes place. Let us take an example, given the vector:

$\vec{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, we wish to apply such a transform that it rotates anti-clockwise by 90°. We can clearly deduce from the figure that our transformation matrix has to be such that it changes our vector to $\vec{w} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Putting it in mathematical terms, $T \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, where T is called our **transformation matrix.**

In this case, $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, T = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$, the **rotation** operation.

> **Note:** If the rows or columns of the transformation matrix are **linearly dependent**, the 2-D space squishes into a single line.(reduction of dimension)

2.3.2 Establishing the connection between the geometric and numeric interpretations of dot products.

Think of a linear transformation in this manner:
n-dimension input $\rightarrow L(\vec{v}) \rightarrow$ 1-dimension output (number line)

Formally speaking, linear transformations must obey two properties:

- $L(\vec{v}+\vec{w}) = L(\vec{v}) + L(\vec{w})$
- $L(c\vec{v}) = cL(\vec{v})$

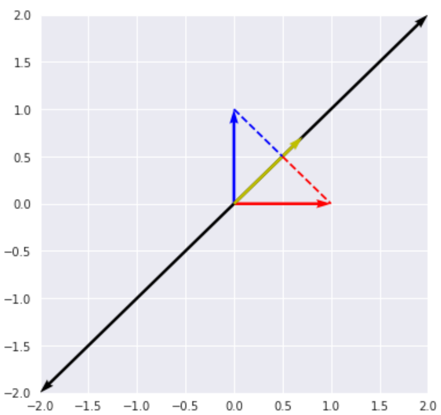

Figure 2.4: Red: $\hat{\imath}$ Blue: $\hat{\jmath}$ Yellow: Unit vector $\vec{u}$ Black: Number Line

A visual equivalent of these is taking evenly spaced points in our space and applying the transformation so as to represent them all on a number line. If the dots are still evenly spaced, then the transformation is linear. Both $\hat{\imath}, \hat{\jmath}$ can be reduced in a way that they are contained on a number line after the linear transformation is applied. This is shown in Fig. 3.4. Also, no matter how we assume our number line's orientation, *due to symmetry in projections of unit vectors* (explained earlier), whatever component $\hat{u}$ projects onto either of the unit vectors, they also project the same magnitude of component onto $\hat{u}$ respectively. This helps us find where $\hat{\imath}, \hat{\jmath}$ *land after our transformation is applied.*

Thus, we know that our transformation matrix has to be, $T = \begin{bmatrix} u_x & u_y \end{bmatrix}$ **i.e. the changed basis after transformation.** Assuming a vector $\begin{bmatrix} x \\ y \end{bmatrix}$, undergoes transformation, we can say that $\begin{bmatrix} u_x & u_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = u_x x + u_y y$, which is computationally equivalent to $\begin{bmatrix} u_x \\ u_y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}$. This is an example of what is known as **duality.**

Another important thing to note is the association between 1x2 matrices and 2-D vectors. $\begin{bmatrix} x & y \end{bmatrix} \underset{\longleftrightarrow}{flip} \begin{bmatrix} x \\ y \end{bmatrix}$

2.3.3 Use of Dot Products

- To compare two vectors, to what extent are the two alike.
- By definition, it gives a single number that indicates the component of a vector in the direction of another. This property is especially useful while changing basis. Let us have $\vec{v}$ to be changed to $\vec{w}$ using a set of orthogonal basis given $\{u_1, u_2, \cdots, u_n\}$, we can transform $\vec{v}$ component by component:

$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} u_1.\vec{v} \\ u_2.\vec{v} \\ \vdots \\ u_n.\vec{v} \end{bmatrix} = \begin{bmatrix} u_1^T\vec{v} \\ u_2^T\vec{v} \\ \vdots \\ u_n^T\vec{v} \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} \vec{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}^T \vec{v}$$

2.3.4 Properties of Dot Products

Given vectors $\vec{a}, \vec{b}, \vec{c}$ and a scalar r:

- **Commutative:** $\vec{a}.\vec{b} = \vec{b}.\vec{a}$
- **Distributive:** $\vec{a}.(\vec{b} + \vec{c}) = \vec{a}.\vec{b} + \vec{a}.\vec{c}$
- **Bi-linear:** $\vec{a}.(r\vec{b} + \vec{c} = r(\vec{a}.\vec{b}) + \vec{a}.\vec{c})$
- **Scalar multiplication:** $(r_1\vec{a}).(r_2\vec{b}) = r_1 r_2(\vec{a}\vec{b})$
- **Not Associative:** Dot product of three is not defined.
- **Orthogonal:** Two no-zero vectors $\vec{a}, \vec{b}$ are orthogonal if and only if $\vec{a}.\vec{b} = 0$.
- **Not Cancellation:** If $\vec{a}.\vec{b} = \vec{a}.\vec{c}$, $\vec{b} \neq \vec{c}$ is also possible.
- **Product rule:** Derivative of $\vec{a}.\vec{b}$ is $\vec{a}'.\vec{b} + \vec{a}\vec{b}'$

2.3.5 Important Results

- **Schwarz inequality:** $||\vec{v}.\vec{w}|| \leq ||\vec{v}||||\vec{w}||$
- **Triangle inequality:** $||\vec{v} + \vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||$
- $c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_n\vec{v_n} = 0$, implies **dependence.**

Chapter 3

Matrices

3.1 Definition

A matrix is a rectangular array of numbers or other mathematical objects, arranged in rows and columns, and on which operations such as *addition* and *multiplication* are defined. Matrices are usually defined over a field $\mathbb{F}$ of scalars like real numbers($\mathbb{R}$) or complex numbers($\mathbb{C}$).

3.2 Notation

The individual terms of the matrix are denoted by a_{ij} and are called the **elements** of the matrix. Here, i represents the row position and j represents the column position. Sometimes the entries of the matrix are also depicted by a formula such as $a_{ij} = f(i,j)$, example: $a_{ij} = i - j$.

Given below is the representation of a matrix M_{mxn}.

$$\mathbf{M} = \begin{array}{cc} & \begin{array}{cccc} n & cols & i & \rightarrow \end{array} \\ \begin{array}{c} m \\ rows \\ j \\ \downarrow \end{array} & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \end{array} = (a_{ij}) \in \mathbb{R}^{mxn}$$

3.3 Size

The **order** of a matrix is given by its m rows and n columns and is written as mxn and m,n are individually referred to as **dimensions**. Matrices that have a single row are called **row vectors**, and those that have a single column are known as **column vectors.** A matrix with same number or rows and columns is known as a **square matrix.**

3.4 Basic Operations

3.4.1 Matrix Addition: Entry wise Sum

The operation of adding two matrices by adding corresponding entries together. Two matrices must have the **same dimensions** to be added.

$(\mathbf{A}+\mathbf{B})_{ij} = \mathbf{A}_{ij} + \mathbf{B}_{ij}$, where $1 \le i \le m$ and $1 \le j \le n$.

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix} =$$

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

3.4.2 Matrix Addition: Direct Sum

Another operation, which is used less often, is the direct sum (denoted by $\oplus$). The direct sum of any pair of matrices A of size $m \times n$ and B of size $p \times q$ is a matrix of size $(m+p) \times (n+q)$ defined as

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & b_{11} & \cdots & b_{1q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & b_{p1} & \cdots & b_{pq} \end{bmatrix}$$

The direct sum of matrices is a special type of **block matrix**, , in particular the direct sum of square matrices is a block diagonal matrix.

In general, direct sum of n matrices is:

$$\bigoplus_{i=1}^{n} \mathbf{A}_i = diag(\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_n) = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_n \end{bmatrix}$$

3.4.3 Scalar Multiplication

The **left scalar multiplication** of a matrix $\mathbf{A}$ with a scalar λ gives another matrix $\lambda\mathbf{A}$ of the same size as $\mathbf{A}$. The entries of $\lambda\mathbf{A}$ are defined by $(\lambda\mathbf{A})_{ij} = \lambda(\mathbf{A}_{ij})$.

$$\lambda\mathbf{A} = \lambda \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{pmatrix} = \begin{pmatrix} \lambda A_{11} & \lambda A_{12} & \cdots & \lambda A_{1m} \\ \lambda A_{21} & \lambda A_{22} & \cdots & \lambda A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda A_{n1} & \lambda A_{n2} & \cdots & \lambda A_{nm} \end{pmatrix}$$

Similarly, **right scalar multiplication** of a matrix $\mathbf{A}$ with a scalar λ is defined to be $(\mathbf{A}\lambda)_{ij} = (\mathbf{A}_{ij})\lambda$.

$$\mathbf{A}\lambda = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{pmatrix} \lambda = \begin{pmatrix} A_{11}\lambda & A_{12}\lambda & \cdots & A_{1m}\lambda \\ A_{21}\lambda & A_{22}\lambda & \cdots & A_{2m}\lambda \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1}\lambda & A_{n2}\lambda & \cdots & A_{nm}\lambda \end{pmatrix}$$

When the underlying ring is commutative, for example, the real or complex number field, these two multiplications are the same, and are simply called **scalar multiplication**. However, for matrices over a more general ring that are not commutative, such as the *quaternions*, they may not be equal.

3.4.4 Transposition

The transpose of a matrix is an operator which flips a matrix over its diagonal, that is it switches the row and column indices of the matrix by producing another matrix denoted as $\mathbf{A}^T$.

If $\mathbf{A}$ is $m \times n$, $\mathbf{A}^T$ will be $n \times m$

$$\left[\mathbf{A}^T\right]_{ij} = [\mathbf{A}]_{ji} \quad ex: \begin{bmatrix} a_0 & a_1 \\ a_2 & a_3 \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} a_0 & a_2 \\ a_1 & a_3 \end{bmatrix}$$

Properties of Transpose

For matrices $\mathbf{A}, \mathbf{B}$ and scalar c we have the following properties of transpose

- $(\mathbf{A}^T)^T = \mathbf{A}$, transpose is it's own inverse.
- $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$, distributive over addition.
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$, the order of factors reverses.
- $(c\mathbf{A})^T = c\mathbf{A}^T$, transpose of a scalar is a scalar.
- $\det(\mathbf{A}^T) = \det(\mathbf{A})$, the determinant of a square matrix is the same as its transpose.
- $[\mathbf{a} \cdot \mathbf{b}] = \mathbf{a}^T\mathbf{b}$
- If $\mathbf{A}$ has only real entries, then $\mathbf{A}^T\mathbf{A}$ is a **positive-semidefinite matrix.**
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$, The transpose of an invertible matrix is also invertible, and its inverse is the transpose of the inverse of the original matrix.
- If $\mathbf{A}$ is a square matrix, then its eigenvalues are equal to the eigenvalues of its transpose since they share the same *characteristic polynomial*.

Points 2 and 4 together tell us that a transpose is a *linear map* from the space of $m \times n$ matrices to the space of all $n \times m$ matrices.

Special Transpose Matrices

- *Symmetric Matrix:* $\mathbf{A}^T = \mathbf{A}$
- *skew-Symmetric Matrix:* $\mathbf{A}^T = -\mathbf{A}$
- *Orthogonal Matrix:* $\mathbf{A}^T = \mathbf{A}^{-1}$
- *Hermartian Matrix:* $\mathbf{A}^T = \overline{\mathbf{A}}$
- *skew-Hermitian Matrix:* $\mathbf{A}^T = -\overline{\mathbf{A}}$
- *Unitary Matrix:* $\mathbf{A}^T = \overline{\mathbf{A}^{-1}}$

3.4.5 Matrix Multiplication

Matrix multiplication or **matrix product** is a binary operation that produces a matrix from two matrices with entries in a field, or, more generally, in a ring. The matrix product is designed for representing the **composition of linear maps** that are represented by matrices.

If $\mathbf{A}$ is an $n \times m$ matrix and $\mathbf{B}$ is an $m \times p$ matrix, $\mathbf{AB} = \mathbf{C}$, where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mp} \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{np} \end{pmatrix}$$

Such that $c_{ij} = a_{i1}b_{1j} + \cdots + a_{im}b_{mj} = \sum_{k=1}^{m} a_{ik}b_{kj}$.
for $i = 1, \cdots, n$ and $j = 1, \cdots, m$.

The product is defined if and only if columns in $\mathbf{A}$ = rows in $\mathbf{B}$.

Properties of Matrix Multiplication

- **Not Commutative: $\mathbf{AB} \neq \mathbf{BA}$**
- **Distributive: $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$** (left distributivity),
  $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{BA} + \mathbf{CA}$ (right distributivity)
- **Associative: $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$**
  $\prod_{i=1}^{n} A_i = A_1 A_2 \cdots A_n$, where multiplication order does not matter.
- **Transpose:** $(\mathbf{AB})^T = \mathbf{B}^\mathbf{T}\mathbf{A}^\mathbf{T}$, when scalars are commutative.
- **Complex Conjugate:** $(\mathbf{AB})^* = \mathbf{A}^*\mathbf{B}^*$
- **Application to Similarity:** $\mathbf{S_P}(\mathbf{AB}) = \mathbf{S_P}(\mathbf{A})\mathbf{S_P}(\mathbf{B})$
  $\mathbf{P}^{-1}(\mathbf{AB})\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}(\mathbf{PP}^{-1})\mathbf{BP} = (\mathbf{P}^{-1}\mathbf{AP})(\mathbf{P}^{-1}\mathbf{BP})$

Powers of a Matrix

As for any ring, one may raise a square matrix to any nonnegative integer power multiplying it by itself repeatedly in the same way as for ordinary numbers. $\mathbf{A}^k = \mathbf{AA}\cdots\mathbf{A}$, k times.

Let $\mathbf{D}$ be a diagonal matrix with elements a_{ii} on the diagonal, then, $\mathbf{D}^k$ has ${a_{ii}}^k$ as its diagonal elements.

Views on Matrix Muliplication

- **rows times columns**
  Where $\mathbf{A}$ is $m \times p$ and $\mathbf{B}$ is $p \times n$ and,
  $\mathbf{AB}_{ij} = (i^{th} row\ of\ \mathbf{A}) \cdot (j^{th} column\ of\ \mathbf{B})$

$$\mathbf{AB} = \begin{bmatrix} \\ \cdots & \cdots & \cdots \\ \\ \end{bmatrix} \begin{bmatrix} & \vdots & \\ & \vdots & \\ & \vdots & \end{bmatrix}$$

- **matrix times a column**

$$\mathbf{AB} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{bmatrix} \vdots & & \\ \vdots & & \\ \vdots & & \end{bmatrix}; \quad \mathbf{AB} = \begin{bmatrix} \mathbf{A}b_1 & \cdots & \mathbf{A}b_n \end{bmatrix}$$

- **rows times a matrix**

$$\begin{bmatrix} row_i & of & \mathbf{A} \end{bmatrix} \begin{bmatrix} & & \\ & \mathbf{B} & \\ & & \end{bmatrix} = \begin{bmatrix} row_i & of & \mathbf{AB} \end{bmatrix}$$

- **column times rows**

$$\mathbf{AB} = \sum_{i=1}^{n} (col_i)(row_i)$$

Note on Complexity

The conventional algorithm for matrix multiplication results in a worst case complexity of $\Theta(n^3)$, since it requires, n^3 multiplications and $n(n^2 - 1)$ additions.

3.4.6 Row Operations

There are three types of row operations that are given as multiplication with an **elementary matrix.**

- *Row Addition:* Adding one row to another
- *Row Multiplication:* Multiplying all row entries by a non-zero constant factor.
- *Row Switching:* Interchanging rows with the help of a **permutation matrix.**

3.4.7 Subsection

A submatrix of a matrix is obtained by deleting any collection of rows and/or columns. For example, from the following 3×4 matrix, we can construct a 2×3 submatrix by removing row 3 and column 2:

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & \color{red}{2} & 3 & 4 \\ 5 & \color{red}{6} & 7 & 8 \\ \color{red}{9} & \color{red}{10} & \color{red}{11} & \color{red}{12} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 4 \\ 5 & 7 & 8 \end{bmatrix}$$

The minors and cofactors of a matrix are found by computing the determinant of certain submatrices.

3.5 Types of Matrices

3.5.1 Terms Associated With Matrices

Main Diagonal

The **main diagonal** of a matrix $\mathbf{A}$ is the collection of entries $\mathbf{A}_{ij}$ where $i = j$. It is also known as principal diagonal, primary diagonal, leading diagonal, or major diagonal.

Anti-Diagonal

The **antidiagonal** of a square matrix $\mathbf{B}_{n\times n}$ is the collection of entries $\mathbf{B}_{ij}$ such that $i + j = n + 1$, if $i, j = (1..N)$. That is, it runs from the top right corner to the bottom left corner. It is also known as counter diagonal, secondary diagonal, trailing diagonal or minor diagonal.

3.5.2 Square Matrix

A **square matrix** is a matrix with the same number of rows and columns. An $n \times n$ matrix is known as a square matrix of order n. Any two square matrices of the same order can be added and multiplied. Square matrices are often used to represent simple linear transformations, such as shearing or rotation.

Example:

$$\mathbf{A}_{3\times 3} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

3.5.3 Zero Matrix

The set of $m \times n$ matrices with entries in a ring $\mathbf{K}$, forms a ring $\mathbf{K}_{m,n}$. The zero matrix $\mathbf{0}_{K_{m,n}}$ in $\mathbf{K}_{m,n}$, is the matrix with all entries equal to $\mathbf{0}_K$, where $\mathbf{0}_K$, is the additive identity in $\mathbf{K}$.

Notation:

$$\mathbf{0}_{K_{m,n}} = \begin{bmatrix} 0_K & 0_K & \cdots & 0_K \\ 0_K & 0_K & \cdots & 0_K \\ \vdots & \vdots & \ddots & \vdots \\ 0_K & 0_K & \cdots & 0_K \end{bmatrix}_{m\times n} \qquad Ex: \mathbf{0}_{2,2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, where\ \mathbf{K} = \mathbb{R}$$

$$\forall \mathbf{A} \in \mathbf{K}_{m\times n}, \quad \mathbf{0}_{K_{m,n}} + \mathbf{A} = \mathbf{A} + \mathbf{0}_{K_{m,n}} = \mathbf{A}$$

3.5.4 Diagonal Matrix

The matrices whose entries other than the main diagonal are zero. Square diagonal matrices are also **symmetric**. The term can also sometimes refer to **rectangular diagonal matrices**.

Example:

$$D_{square} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad D_{rectangle} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}$$

Matrix Operations

The operations of matrix addition and matrix multiplication are especially simple for symmetric diagonal matrices.

- *Addition:* $diag(a_1, \cdots, a_n) + diag(b_1, \cdots, b_n) = diag(a_1 + b_1, \cdots, a_n + b_n)$
- *Multiplication:* $diag(a_1, \cdots, a_n) \cdot diag(b_1, \cdots, b_n) = diag(a_1 b_1, \cdots, a_n b_n)$
- *Inverse:* $diag(a_1, \cdots, a_n)^{-1} = diag(a_1^{-1}, \cdots, a_n^{-1})$.
  A diagonal matrix is invertible if and only if all its diagonal entries are non-zero.

We can thus say that diagonal matrices form their own subring of the ring of all $n \times n$ matrices.

Properties

- The determinant of $diag(a_1 \cdots a_n)$ is $\prod_{i=1}^{n} a_i$
- The **adjugate** of a diagonal matrix is again diagonal.
- A square matrix is diagonal if and only if it is triangular and normal.
- Any square diagonal matrix is also a symmetric matrix.

3.5.5 Identity Matrix

The **identity matrix**, or sometimes ambiguously called a **unit matrix**, of size n is the $n \times n$ square matrix with ones on the main diagonal and zeros elsewhere. It is denoted by $\mathbf{I}_n$, or simply by $\mathbf{I}$.

Example:

$$I_1 = \begin{bmatrix}1\end{bmatrix},\ I_2 = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},\ I_3 = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix},\ \cdots,\ I_n = \begin{bmatrix}1 & 0 & 0 & \cdots & 0\\ 0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\end{bmatrix}$$

A property of identity matrices is that for $\mathbf{A}_{m\times n}$, $\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$.

In particular, the identity matrix serves as the unit of the ring of all $n \times n$ matrices, and as the identity element of the general linear group $GL(n)$ consisting of all invertible n×n matrices (The identity matrix is its own inverse). The identity matrix of a given size is the only idempotent matrix of that size having full rank.

3.5.6 Scalar Matrix

A square diagonal matrix with all its main diagonal entries equal is a **scalar matrix**, that is, a scalar multiple $\lambda\mathbf{I}$ of the identity matrix $\mathbf{I}$.

Example: For a scalar field $\mathbb{R}$

$$\begin{bmatrix}\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda\end{bmatrix} \equiv \lambda\mathbf{I}_3$$

3.5.7 Antidiagonal Matrix

An anti-diagonal matrix is a matrix where all the entries are zero except those on the diagonal going from the lower left corner to the upper right corner (↗), known as the anti-diagonal.

Example:

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

Properties

- All antidiagonal matrices are **persymmetric.**
- The product of two antidiagonal matrices is a diagonal matrix.

- An antidiagonal matrix is invertible if and only if all its diagonal entries are non zero.
- $|\mathbf{D}| = \prod_{i=1}^{n} \mathbf{A}_{ii}$, where $\mathbf{A}$ is an $n \times n$ antidiagonal matrix. The sign of this determinant is dependent on the triangular number corresponding to n.

3.5.8 Elementary Matrix

An e**elementary matrix** is a matrix which differs from the identity matrix by one single elementary row operation. The elementary matrices generate the **general linear group** of **invertible matrices**. Left multiplication (pre-multiplication) by an elementary matrix represents elementary **row operations**, while right multiplication (post-multiplication) represents elementary **column operations**.

Elementary row operations are used in Gaussian elimination to reduce a matrix to row echelon form. They are also used in Gauss-Jordan elimination to further reduce the matrix to reduced row echelon form.

Row Operations

There are three types of elementary matrices, which correspond to three types of row operations (respectively, column operations):

- **Row switching:** A row within the matrix can be switched by another row. $\mathbf{R}_i \leftrightarrow \mathbf{R}_j$
  This kind of an elementary matrix is obtained by swapping row i and row j of the identity matrix. This will intern swap the same rows in matrix A when left multiplied.

  Notation:

  $$\mathbf{T}_{i,j} = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

  Properties:

  - $T_{ij} = T_{ij}^{-1}$ It is the inverse of itself.
  - $det(T_{ij}) = -1$, making $det(T_{ij}A) = det(-A)$

- **Row multiplication:** Each element in a row can be multiplied by a non-zero constant. $k\mathbf{R}_i \rightarrow \mathbf{R}_i$, where $k \neq 0$.

The corresponding elementary matrix is a diagonal matrix, with all entries 1 except the ith position, where it is m.

Notation:

$$\mathbf{D}_i(m) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & m & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

Properties:

- The inverse of this matrix is $D_i(m)^{-1} = D_i(1/m)$
- The matrix and it's inverses are diagonal matrices.
- $det(D_i(m)) = m$, giving us $det(D_i(m)\ A) = m\ det(A)$

- **Row addition:** A row can be replaced by the sum of that row and a multiplie of another row. $\mathbf{R}_i + k\mathbf{R}_j \to \mathbf{R}_i$, where $i \neq j$.

Notation:

$$\mathbf{L}_{ij}(m) = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & m & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$$

Properties:

- These transformations are a kind of **shear mapping**, also known as a *transvections*.
- $L_{ij}(m)^{-}1 = L_{ij}(-m)$ (inverse matrix).
- Row-addition transforms satisfy the **Steinberg relations**.

3.5.9 Triangular Matrix

A **triangular matrix** is a special kind of square matrix. A square matrix is called **lower triangular** if all the entries above the main diagonal are zero. Similarly, a square matrix is called **upper triangular** if all the entries below the main diagonal are zero. A triangular matrix is one that is either lower triangular or upper triangular. A matrix that is both upper and lower triangular is called a diagonal matrix.

Because matrix equations with triangular matrices are easier to solve, they are very important in numerical analysis. By the **LU decomposition** algorithm, an invertible matrix may be written as the product of a lower triangular matrix L and an upper triangular matrix U if and only if all its **leading principal minors** are non-zero.

Lower Triangular or Left Triangular

$$\begin{bmatrix} \ell_{1,1} & & & & 0 \\ \ell_{2,1} & \ell_{2,2} & & & \\ \ell_{3,1} & \ell_{3,2} & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ \ell_{n,1} & \ell_{n,2} & \ldots & \ell_{n,n-1} & \ell_{n,n} \end{bmatrix}$$

Upper Triangular or Right Triangular

$$\begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \ldots & u_{1,n} \\ & u_{2,2} & u_{2,3} & \ldots & u_{2,n} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & u_{n-1,n} \\ 0 & & & & u_{n,n} \end{bmatrix}$$

Matrices that are similar to triangular matrices are called **triangularisable**. Many operations on upper triangular matrices preserve the shape (same can be said for lower triangular matrices):

- The sum of two upper triangular matrices is upper triangular.
- The product of two upper triangular matrices is upper triangular.
- The inverse of an invertible upper triangular matrix is upper triangular.
- The product of an upper triangular matrix by a constant is an upper triangular matrix.

3.5.10 Invertible Matrix

An n-by-n square matrix **A** is called **invertible** (also **non-singular** or **non-degenerate**) if there exists an $n \times n$ square matrix **B** such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I_n}$. In case of rectangular matrices, they have separate left and right inverses. For an $m \times n$ matrix **A**, if matrix **B** is a **left inverse**, $\mathbf{BA} = \mathbf{I_n}$ and if matrix **B** is a **right inverse**, $\mathbf{AB} = \mathbf{I_m}$.

A square matrix that is **not** invertible is called **singular** or **degenerate**. A square matrix is singular if and only if its determinant is **0**. Singular matrices are rare in the sense that a square matrix randomly selected from a continuous uniform distribution on its entries will almost never be singular.

Properties

- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(k\mathbf{A})^{-1} = k^{-1}A^{-1}$ for nonzero scalar k
- $(\mathbf{A}x)^{+} = x^{+}\mathbf{A}^{-1}$ where + denotes the **Moore-Penrose pseudo inverse** and x is a vector
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
- $det(\mathbf{A}^{-1}) = (det(\mathbf{A}))^{-1}$
- If $\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_n$ are invertible matrices, then $(\mathbf{A}_1\mathbf{A}_2\cdots\mathbf{A}_n)^{-1} = \mathbf{A}_n^{-1}\cdots\mathbf{A}_2^{-1}\mathbf{A}_1^{-1}$

In relation to Adjugate

The adjugate of a matrix **A** can be used to find the inverse of **A** as follows:

$\mathbf{A}^{-1} = \frac{1}{det(\mathbf{A})}\,\mathrm{adj}(\mathbf{A})$

Invertible matrix theorem

Let **A** be a square $n \times n$ matrix over a field $\mathbb{K}$ (for example the field $\mathbb{R}$ of real numbers). The following statements are equivalent, i.e., for any given matrix they are either all true or all false:

- **A** is invertible, i.e. **A** has an inverse, is non-singular, or is non-degenerate.
- **A** is row-equivalent to the $n \times n$ identity matrix $\mathbf{I_n}$.
- **A** is column-equivalent to the $n \times n$ identity matrix $\mathbf{I_n}$.
- $det(\mathbf{A}) \neq 0$. In general, a square matrix over a commutative ring is invertible if and only if its determinant is a **unit** in that ring.
- **A** has full rank; that is, $rank(\mathbf{A}) = n$.
- The equation $\mathbf{A}x = 0$ has only the trivial solution $x = 0$.
- $ker(\mathbf{A}) = 0$.
- The equation $\mathbf{A}x = b$ has exactly one solution for each $b \in \mathbb{K}_n$.
- The columns of **A** are linearly independent.
- The columns of **A** span $\mathbb{K}_n$.
- $col(\mathbf{A}) = \mathbb{K}_n$.
- The columns of **A** form a basis of $\mathbb{K}_n$.
- The linear transformation mapping x to $\mathbf{A}x$ is a bijection: $\mathbb{K}_n \to \mathbb{K}_n$.
- There is an $n \times n$ matrix **B** such that $\mathbf{AB} = \mathbf{I_n} = \mathbf{BA}$.
- The transpose $\mathbf{A}^T$ is an invertible matrix (hence rows of **A** are linearly independent, span $\mathbb{K}_n$, and form a basis of $\mathbb{K}_n$).
- The number 0 is not an eigenvalue of **A**.
- The matrix **A** can be expressed as a finite product of elementary matrices.
- The matrix **A** has a left inverse (i.e. there exists a **B** such that $\mathbf{BA} = \mathbf{I}$) or a right inverse (i.e there exists a **C** such that $\mathbf{AC} = \mathbf{I}$), in which case both left and right inverses exist and $\mathbf{B} = \mathbf{C} = \mathbf{A}^{-1}$.

Chapter 4

System of Linear Equations

A s**system of linear equations** (or linear system) is a collection of two or more linear equations involving the same set of variables.

4.1 Representation of Linear Systems

General Form

A general system of m linear equations with n unknowns can be written as

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

Vector Equation Form

This view helps us show unknowns as weights of different vectors. It helps bring to life ideas such as *span* of a linear combination.

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

Matrix Form

The vector equation for is equivalent to a matrix form of $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{x}$ is a *column vector* with n entries, and $\mathbf{b}$ is also a column vector with m entries.

$$\mathbf{Ax} = \mathbf{b} \quad \Rightarrow \quad \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

The solution to $\mathbf{Ax} = \mathbf{b}$ can be written as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, provided A^{-1} exists.

4.2 Solution Set

A linear system may behave in one of three possible ways:

- The system has *infinitely many* solutions.
- the system has a *single unique* solution.
- the system has *no* solution.

4.2.1 Geometric Interpretation

For a system involving two variables (x and y), each linear equation determines a line on the xy−plane. Because a solution to a linear system must satisfy all of the equations, the solution set is the intersection of these lines, and is hence either a line, a single point, or the empty set.

For three variables, each linear equation determines a plane in three-dimensional space, and the solution set is the intersection of these planes. Thus the solution set may be a plane, a line, a single point, or the empty set. For example, as three parallel planes do not have a common point, the solution set of their equations is empty; the solution set of the equations of three planes intersecting at a point is single point; if three planes pass through two points, their equations have at least two common solutions; in fact the solution set is infinite and consists in all the line passing through these points.

For n variables, each linear equation determines a hyper plane in n-dimensional space. The solution set is the intersection of these hyper planes, and is a flat, which may have any dimension lower than n.

4.2.2 General Behavior

In general, the behavior of a linear system is determined by the relationship between the number of equations and the number of unknowns.

- In general, a system with fewer equations than unknowns has infinitely many solutions, but it may have no solution. Such a system is known as an **under determined system.**
- In general, a system with the same number of equations and unknowns has a single unique solution.
- In general, a system with more equations than unknowns has no solution. Such a system is also known as an **over determined system.**

4.3 Properties

4.3.1 Independence

The equations of a linear system are independent if none of the equations can be derived algebraically from the others. When the equations are independent, each equation contains new information about the variables, and removing any of the equations increases the size of the solution set. For linear equations, logical independence is the same as **linear independence.** (Described later)

Example: $x+2y=3$ and $2x+4y=6$ are not independent, since, it's the same equation scaled by a factor of two.

4.3.2 Consistency

A linear or nonlinear system of equations is **consistent** if there is at least one set of values for the unknowns that satisfies every equation in the system — that is, that when substituted into each of the equations makes each equation hold true as an identity. In contrast, an equation system is **inconsistent** if there is no set of values for the unknowns that satisfies all of the equations.

Cases for Linear Equations

- **Under determined and Consistent:** Infinite solutions.
- **Under determined and Inconsistent:** No solution.
- **Exactly determined and Consistent:** One solution.
- **Exactly determined and Inconsistent:** No solution.
- **Over determined and Consistent:** At least one solution.
- **Over determined and Inconsistent:** No solution.

Rouché-Capelli Theorem

A system of linear equations with n variables has a solution if and only if the rank of its *coefficient matrix* **A** is equal to the rank of its *augmented matrix* $mathbf[A|b]$. If there are solutions, they form an **affine subspace** of $\mathbb{R}^n$ of dimension $n - rank(\mathbf{A})$. In particular:

- if $n = rank(\mathbf{A})$, the solution is unique,
- otherwise there is an infinite number of solutions.

Consider the system of equations

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

The coefficient matrix becomes

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

and the augmentation matrix is

$$[A|b] = \left[\begin{array}{ccc|c} 1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \end{array}\right]$$

Since both of these have the same rank, namely 2, there exists at least one solution; and since their rank is less than the number of unknowns, the latter being 3, there are an infinite number of solutions.

4.3.3 Equivalence

Two linear systems using the same set of variables are equivalent if each of the equations in the second system can be derived algebraically from the equations in the first system, and vice versa. Two systems are equivalent if either both are inconsistent or each equation of each of them is a linear combination of the equations of the other one. It follows that two linear systems are equivalent if and only if they have the same solution set.

4.4 Solving System of Linear Equations

Gaussian elimination (also known as **row reduction**) is an algorithm for solving systems of linear equations. It is usually understood as a sequence of operations performed on the corresponding matrix of coefficients. This method can also be used to find the rank of a matrix, to calculate the determinant of a matrix, and to calculate the inverse of an invertible square matrix.

To perform row reduction on a matrix, one uses a sequence of **elementary row operations** to modify the matrix until the lower left-hand corner of the matrix is filled with zeros, as much as possible. Using these operations, a matrix can always be transformed into an upper triangular matrix, and in fact one that is in row echelon form. We can then solve it using back substitution. Elimination can only take place, provided the **pivot elements** (the elements that remain on the diagonal) are not zero.

Steps for Gauss Elimination

Step I: Write the given the equations:

$$\begin{gathered} x+3y+z=9 \\ x+y-z=1 \\ 3x+11y+6z=36 \end{gathered}$$

Step II: Write the **augmented matrix**

$$\left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 6 & 36 \end{array}\right]$$

Step III: Eliminate variables using **elementary operations**

$E_{13}E_{12}A \to A'$ (Eliminating elements under *first pivot*).

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 1 & 1 & -1 & 1 \\ 3 & 11 & 6 & 36 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 3 & 9 \end{array}\right]$$

$E_{23}A' \to U$ (Eliminating elements under *second pivot*).

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 2 & 3 & 9 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 3 & 1 & 9 \\ 0 & -2 & -2 & -8 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

Thus, we find U which is an upper triangular matrix that can be solved $E_{23}E_{13}E_{12}A \to U$, which can be written as $\mathbf{EA} = \mathbf{U}$

Note: refer to Elementary Matrices to know more about elementary operations.

Gauss-Jordan Form

Step I: Augment an identity matrix to our matrix

$$[\mathbf{A} \;\; \mathbf{I}] = \left[\begin{array}{ccc|ccc} 2 & -1 & 0 & 1 & 0 & 0 \\ -1 & 2 & -1 & 0 & 1 & 0 \\ 0 & -1 & 2 & 0 & 0 & 1 \end{array}\right]$$

Step II: Reduce to upper triangular form **U**

$$r_2 \to r_2 + 1/2 r_1$$
$$r_3 \to r_3 + 2/3 r_2$$

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

Step III: Carry out upward elimination to remove the elements above the pivots in a similar manner.

$$r_2 \to r_2 + 3/4 r_3$$
$$r_1 \to r_1 + 2/3 r_2$$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 3/2 & 1 & 1/2 \\ 0 & 3/2 & 0 & 3/4 & 3/2 & 3/4 \\ 0 & 0 & 4/3 & 1/3 & 2/3 & 1 \end{array}\right]$$

Step IV: Multiplying each row by the scalar multiplicative inverse of it's respective pivot.

$$\left[\mathbf{I} \quad \mathbf{A}^{-1}\right] = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3/4 & 1/2 & 1/4 \\ 0 & 1 & 0 & 1/2 & 1 & 1/2 \\ 0 & 0 & 1 & 1/4 & 1/2 & 3/4 \end{array}\right]$$

Some Important Notes

- It is worthwhile to fine $\mathbf{A}^{-1}$ for small matrices, not for large.
- If **A** is symmetric across it's diagonal, so is $\mathbf{A}^{-1}$.
- In our example we see that although, **A** has three diagonals (1 primary and 2 offset) but $\mathbf{A}^{-1}$ is a **dense matrix** with no zeros.
- Diagonally dominating matrices are usually invertible.

4.4.1 Applications of Gauss-Jordan Elimination

Computing Determinants

To explain how Gaussian elimination allows the computation of the determinant of a square matrix, we have to recall how the elementary row operations change the determinant:

- Swapping two rows multiplies the determinant by -1
- Multiplying a row by a nonzero scalar multiplies the determinant by the same scalar
- Adding to one row a scalar multiple of another does not change the determinant.

If d is the product of all the scalars by which the determinant has been multiplied, using the above rules. Then the determinant of $\mathbf{A}$ is the quotient by d on the product of its diagonal elements.

Determinant: $det(\mathbf{A}) = \frac{\prod diag(\mathbf{A})}{d}$

Finding the Inverse

We can use the Gauss-Jordan method to find the inverse of a given matrix as shown by the example above.

Finding the Rank of a matrix

As we will see in the later chapters, conversion to **row reduced echelon form** (*rref*) helps us determine the rank.

4.4.2 Computational Efficiency

To solve a system with n equations and n unknowns by performing row operations, till it reaches echelon form and then solving it in reverse order requires:

- $\frac{n(n+1)}{2}$ divisions
- $\frac{2n^3+3n^2-5n}{6}$ multipications
- $\frac{2n^3+3n^2-5n}{6}$ subtractions
- for a total of approximately $\frac{2n^3}{3}$ operations.

Thus giving it a *worst case complexity* of $\Theta(n^3)$. One problem Gaussian elimination can suffer from is numerical instability, caused by the possibility of dividing by very small numbers. If, for example, the leading coefficient of one of the rows is very close to zero, then to row reduce the matrix one would need to divide by that number so the leading coefficient is 1. This means any error that existed for the number which was close to zero would be amplified. Gaussian elimination is numerically stable for diagonally dominant or positive-definite matrices.

4.5 LU Decomposition

Chapter 5

Vector Spaces

5.1 Spaces

A **space** is a set (often of points of the space) with some added structure defined by it's operations and it's properties. Examples: Linear Space, Metric Space, Topological Space, etc.

Mathematical spaces form a hierarchy in such a way that one space can inherit all the characteristics of a parent space. Inner product induces a norm, norm induces a metric and the metric induces a topology.

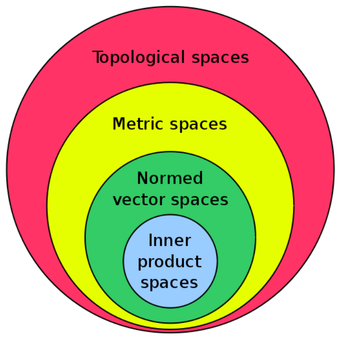

Figure 5.1: A hierarchy of mathematical spaces.

5.2 Vector Space

A vector space, also called a **linear space** is a collection of abstract mathematical objects that are called vectors, which may be added together or multiplied ("scaled") by numbers called scalars, often defined by a field (like $\mathbb{R}$ of real numbers). The operations for vector addition and scalar multiplication must satisfy certain axioms. In linear algebra, vector spaces are characterized by their **dimension**, usually the number of independent vectors in a space.

Formal definition: A vector space over a field F is a set V together with two operations that satisfy the eight axioms listed below.

(Closure properties)
Vector Addition: $V * V \to V$, $\vec{v} + \vec{w} \to \vec{x}$; $v, w, x \in V$
Scalar Multiplication: $F * V \to V$, $a.\vec{v} \to a\vec{v}$, $a\vec{v} \in V$, $a \in F$

Axioms for operations on a vector space.

- **Associativity of addition:** $\vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$
- **Commutativity of addition:** $\vec{u} + \vec{v} = \vec{v} + \vec{u}$
- **Identity element of addition:** $\vec{v} + \vec{0} = \vec{v}$, $\forall \vec{v} \in V$, $\vec{0} \in V$
- **Inverse element of addition:** $\forall \vec{v} \in V \ \exists \ \vec{-v} \in V \ni \vec{v} + (\vec{-v}) = \vec{0}$
- **Compatibility of scalar and field multiplication:** $a(b\vec{v}) = (ab)\vec{v}$
- **Identity element of scalar multiplication:** $1\vec{v} = \vec{v}$, $1 \in F$, $\forall \vec{v} \in V$
- **Distributivity of scalars w.r.t vector addition:** $a(\vec{u}+\vec{v}) = a\vec{u}+a\vec{b}$
- **Distributivity of scalars w.r.t field addition:** $(a + b)\vec{v} = a\vec{v} + b\vec{v}$

An example of a vector space is the Euclidean Space $\mathbb{R}^n$ where vectors are $n-$tuples of real numbers; like a vector in $\mathbb{R}^2$ would be $(1,2)$. To verify the validity of this, we can check the above axioms hold true for both vector addition and scalar multiplication.

Vector spaces often arise as solution sets to various problems involving linearity, such as the set of solutions to homogeneous system of linear equations and the set of solutions of a homogeneous linear differential equation.

Let A be an mxn matrix, then the columns of Ax or AB are **linear combinations of $n-$vectors.** The notation $\mathbb{R}^n$ depicts all column vectors in an $n-$dimensional space that have n components. $\mathbb{R}$ tells us that the components of $\vec{v}$ are real numbers. A vector whose n components are complex lie in the space $\mathbb{C}^n$. The word **space** here asks us to think about all possible vectors.

Examples: $\begin{bmatrix} 4 \\ \pi \end{bmatrix}$ in $\mathbb{R}^2$; $(1,1,0,1,1)$ in $\mathbb{R}^5$; $\begin{bmatrix} 1+i \\ 1-i \end{bmatrix}$ in $\mathbb{C}^2$

- Linear combinations of vectors within a given space will have their result in that space.
- Some more examples of vector spaces:
  M : All $2x2$ matrices. (Satisfy all eight axioms)
  F : All real functions. $f(x)$ (infinite dimensional)
  Z : Zero vector space.
- No space can be without the zero vector.

5.3 Subspace

A subspace can be thought of as a vector space within a vector space.

Let $W \neq \emptyset$ be a subset of vector space V over the field F with the vector operations from V restricted to W. Then W is a vector sub-space if and only if the following properties hold:

- $\vec{u} + \vec{v} \in W; \forall \vec{u}, \vec{v} \in W$
- $c\vec{u} \in W; \forall \vec{u} \in W, c \in F$

A subspace follows all the rules of the host space. Examples:

- $p(x)$: A subspace of $f(x)$ containing all the polynomials of degree n.
- Inside a vector space $\mathbb{M}$ of all $2x2$ matrices some sub-spaces are:
  $\mathbb{U}$: Upper Triangular
  $\mathbb{L}$: Lower Triangular
  $\mathbb{D}$: Diagonal Matrices

5.4 The Four Fundamental Subspaces

When studying a particular matrix, we are often interested in determining vector spaces associated with the matrix so as to better understand how the corresponding linear transformation operates The fundamental subspaces are vector spaces defined by an mxn matrix A (and it's transpose):

- **Column Space**
- **Null Space or Kernel**
- **Row Space**
- **Left Null Space**

The fundamental subspaces are useful for a number of linear algebra applications, including analyzing the rank of a matrix. The subspaces are also closely related by the fundamental theorem of linear algebra.

5.5 Column Space

The column space of a matrix A is the vector space formed by the linear combinations of columns of A. Equivalently, column space consists of all matrices Ax for some vector x. For this reason, the column space is also known as the **image** of A (denoted $im(A)$), as it is the result when A is viewed as a linear transformation of the vector space $\mathbb{R}^m$ (where m is the number of rows of A). In particular, the image of A is necessarily the subspace of $\mathbb{R}^m$, hence the term "fundamental subspace."

For example, consider the matrix

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 0 & 3 & 3 \end{bmatrix}, \text{ here the column vectors are } \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

But the basis is given by $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 3 \end{bmatrix}$

Since there are two vectors in the basis of the column space, the column space is given by $\mathbb{R}^2$ which is a subspace of $\mathbb{R}^3$.

> **Column Space:** Suppose A is an mxn matrix, columns $c_1, \cdots, c_n \in \mathbb{R}^m$, the column space $C(A)$ is the subspace of $\mathbb{R}^m$ spanned by vectors $\{c_i\}_{1 \le i \le n}$.

To solve $Ax = b$ is to express b as a linear combination of columns of A.The right hand side has to lie in the column space of A, or $Ax = b$ has no solution. In the case A is not invertible, then the system has solutions for some b and no solutions for others.

Let A be an mxn matrix:

- Its columns have m components and belong to $\mathbb{R}^m$.
- The column space of A is a subspace of $\mathbb{R}^m$.

Note:

- If $\mathbb{R}^n$ has two subspaces P and L then, $P \cap L$ is also a subspace but $P \cup L$ is not necessarily.
- The dimension of the column space is given by the number of independent column vectors.

5.6 Null Space of A: Solving $Ax = 0$ and $Rx = 0$

The null space or the **kernel** of a matrix A (denoted by $ker(A)$) is the set of all vectors x for which $Ax = 0$. Since it is a vector space, any linear combination of vectors x, y in the null space will also be in the null space as
$A(c_1x + c_2y) = c_1(Ax) + c_2(Ay) = 0 + 0 = 0$
If A is viewed as a linear transformation, the nullspace is the subspace of $\mathbb{R}^n$ that is sent to $\vec{0}$ under the map A, hence the name "fundamental subspace."

> **Nullspace:** The kernel (or nullspace) of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the set $\ker(T)$ of vectors $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{0}$. It is a subspace of $\mathbb{R}^n$ whose dimension is called the nullity. The rank-nullity theorem relates this dimension to the rank of T.

In $Ax = 0$, $x = 0$ will always be a solution (the **trivial** solution). It will be the **only solution** for invertible matrices. For other non-invertible matrices, there will be non-zero solutions. **Elimination helps us identify this space.**

Steps to solving $Ax = 0$ (Homogeneous Form)

- Reducing A to its **row reduced echelon form** R.
- Finding special solutions to $Ax = 0$

5.7 Pivot Columns and Free Columns

Free columns correspond to the columns with no pivots. Example:

$$A = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix}$$

We can deduce from the matrix that some columns are linear combinations of others. Applying elimination, we get

$$U = \begin{bmatrix} \boxed{1} & 2 & 2 & 4 \\ 0 & \boxed{2} & 0 & 4 \end{bmatrix}$$

The **rank** of the matrix here is the number of pivots. In this case $rank(A) = 2$

$rank(A) = n(A_{pivots}) = n(A_{colums}) - n(A_{freevariables})$

Let the above equations have four variables x_1, x_2, x_3, x_4. The elimination procedure tells us that we are free to pick two of them (x_3, x_4). We take them to be $(1,0)$ and $(0,1)$ to find our *special solutions.* The null space then becomes linear combination of these solutions. For the matrix given above

$$S_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}, S_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \text{ special solutions found using back substitution.}$$

Thus, the null space is given by $cS_1 + dS_2$.

5.8 Reduced Row Echelon form

Definition: The reduced row echelon form of a matrix A (denoted as $rref(A)$) is a matrix of equal dimensions that satisfies:

- The left most non-zero element is a 1. This element is known as a **pivot.**
- Any column can have at most 1 pivot. If a column has a pivot, then the rest of the elements of the column are 0.
- For any two columns C_1 and C_2 that have pivots in rows R_1 and R_2 respectively, if pivot in C_1 is to the left of pivot in C_2, then R_1 is above R_2.
- Rows that consist of only zeros are at the bottom of the matrix.

Steps to reach $rref$:

- Carry out *Gaussian Elimination* to reach the upper triangular matrix U.
- Produce 0s above pivots, using pivot rows to eliminate upwards.
- Produce 1s in pivots, by dividing the whole row by it's pivot to reach R.

For systems having the solution only as $x = 0$, we learn an important fact, i.e., $N(A) = Z$, thus we know that all the columns of a matrix are independent if it's null space is Z.

Important: Null space stays the same ($N(A) = N(U) = N(R)$). The reason for this being that since, $Rx = 0$ and $R = EA$, where E is *non-singular square matrix* that is the product of all elementary operations to reach R, we can say that $EAx = 0$. Since E is *invertible*, we can say, $E^{-1}EAx = E^{-1}0$, thus giving $Ax = 0$. Which then has the same null space as $Rx = 0$.

With $n_{rows} > m_{cols}$ there is atleast 1 free variable, thus the system $Ax = 0$ will have atleast 1 special solution, which is not $x = 0$ (A short wide matrix always has non-zero vectors in its null space).

5.9 Rank-Nullity Theorem

The rank nullity theorem states that the rank and the nullity sum to give the number of columns in a matrix.

$rank(M) + nullity(M) = n_{cols}(M)$

5.10 The Rank of a Matrix

The **rank** of a matrix is the dimension of it's row space or column space. This corresponds to the maximum number of linearly independent rows or columns.

Theorem: Let A be a matrix. If $R(A)$ denotes its row space and C(A) denotes its column space, $dim(R(A)) = dim(C(A))$. This quantity is called the rank of A, and is denoted by $rk(A)$.

Proposition: For a given matrix A, $rk(A) = dim(C(A)) = dim(R(A))$

Let A be an mxn matrix where $a_{ij} \in \mathbb{R}$
Let $row_{rank}(A) = r$, therefore, $dim(R(A)) = r$
Let $x_1, x_2, \cdots, x_r$ be the basis vectors for $R(A)$

Claim: $Ax_1, Ax_2, \cdots, Ax_r$ are linearly independent.

Proof:
Assuming a *Homogeneous Relation* with scalar coefficients $c_1, c_2, \cdots, c_r \in \mathbb{R}$
$c_1Ax_1 + c_2Ax_2 + \cdots + c_rAx_r = 0$
$A(c_1x_1 + c_2x_2 + \cdots + c_rx_r) = 0$
$Av = 0$, where $v = c_1x_1 + c_2x_2 + \cdots + c_rx_r$

v is a linear combination of vectors of $R(A)$ which implies $v \in R(A)$.

$Av = 0$ suggests that vector v is orthogonal to every row vector of A.

This holds true only when $v \in R(A)$ and $v \in N(A)$(null space of A) i.e when, $v = 0$

Thus, $c_1x_1 + c_2x_2 + \cdots + c_rx_r = 0$

Since we assumed $x_1, x_2, \cdots, x_r$ to be the basis vectors of $R(A)$, they are linearly independent.

We can thus conclude that $c_1 = c_2 = \cdots = c_r = 0$.
Which makes $Ax_1, Ax_2, \cdots, Ax_r$ a set of r linearly independent vectors that belong to the column space of A.

From this we can say that
$col_{rank}(A) \geq r$, where r is $row_{rank}(A)$

Applying the same procedure to A^T will show us
$row_{rank}(A) \geq r$, where r is $row_{rank}(A^T)$ or the $col_{rank}(A)$,
Since $C(A^T) = R(A)$

Hence we conclude that, $row_{rank}(A) = col_{rank}(A) = rk(A)$

The true size of a system $Ax = 0$ is given by $rk(A) = r_{pivots}$. To find the rank of a matrix, we reduce it to the $rref$.

5.10.1 Rank One matrices

Matrices of rank one only have 1 pivot. Example:

$$A = \begin{bmatrix} 1 & 3 & 10 \\ 2 & 6 & 20 \\ 3 & 9 & 30 \end{bmatrix}, R = \begin{bmatrix} 1 & 3 & 10 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{A can be written as } A = uv^T$$

$$A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 3 & 10 \end{bmatrix}$$

An important consequence: Since $Ax = 0$, we can say that $uv^T x = 0$ or, $u(v^T)x = 0$. This clearly shows us that $v^T x = 0$ or $v.x = 0$, meaning, all vectors x in the null space will be orthogonal to the vector v, in the row space. If $r = 1$ the row space is a *line* and the null space is a *perpendicular hyper plane*.

5.11 The Bigger Picture of Elimination

Elimination answers two major questions:

- Is a column a combination of previous columns?
- Is a row a combination of previous rows?

One pass through the matrix to reach the upper triangular matrix (echelon form) $\mathbf{A} \to \mathbf{U}$ tells us which columns are the combinations. The next pass from $\mathbf{U} \to \mathbf{R}$ (rref form) goes bottom up, telling us what those combinations are.

R reveals basis for the three fundamental subspaces:

- Column space of **A**: Choose the pivot columns of **A** (basis vectors).
- Row space of **A**: Choose the non-zero rows of **R** as it's basis.
- Null Space of **A**: Choose the solutions to $\mathbf{Rx} = \mathbf{0}$.

Note: Reducing $[\mathbf{A}\ \ \mathbf{I}]$ to $[\mathbf{R}\ \ \mathbf{E}]$ tells virtually everything about **A**. The matrix **E** keeps a record otherwise lost. When **A** is square and invertible, $[\mathbf{A}\ \ \mathbf{I}]$ becomes $[\mathbf{I}\ \ \mathbf{A}^{-1}]$.

5.12 The Complete Solution to $Ax = b$

Steps to finding the complete solution:

- **Step I:** Write down the system of linear equations as an *augmented matrix* $[\mathbf{A}\ \ \mathbf{b}]$. Example:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix} \Longrightarrow \left[\begin{array}{cccc|c} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 1 & 3 & 1 & 6 & 7 \end{array}\right]$$

- **Step II:** Apply *elimination* to reach $[\mathbf{R}\ \ \mathbf{d}]$. For a solution to exist, the system of linear equations must be **consistent.** The computational check for this is that when a row becomes a *zero row*, its *augmented part* must also be a zero. Example:

$$[\,\mathbf{R}\ \ \mathbf{d}\,] = \left[\begin{array}{cccc|c} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & \boxed{0} & \boxed{0} \end{array}\right]$$

- **Step III:** *One Particular Solution* $\mathbf{Ax_p} = \mathbf{b}$. For a particular solution, set all *free variables* to zero and then solve $\mathbf{Ax_p} = \mathbf{b}$ for pivot variables, using simple back-substitution. (pivot variables come from $\mathbf{d}$)

- **Step IV:** Complete solution also includes the many solutions from the null space (which we know how to solve). Thus we have, $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_n$, finally giving us:

$$\boxed{\mathbf{A}(\mathbf{x}_p + \mathbf{x}_n) = \mathbf{b}}$$

- For the above example, the complete solution thus becomes:

$$\mathbf{x} \quad = \quad \mathbf{x}_p + \mathbf{x}_n \quad = \quad \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} \quad + \quad \mathbf{x}_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad + \quad \mathbf{x}_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

The idea behind distinguishing particular (also known as *in-homogeneous*) and homogeneous solution is that, the homogenous solutions $\mathbf{Ax} = \mathbf{0}$ constitute a whole *linear subspace*, also known as the *kernel*.

Particular solutions on the other hand, cannot form such a subspace; consider the solutions x_1, x_2 to $\mathbf{Ax} = \mathbf{b}$. We see then that $A(\alpha x_1 + \beta x_2) = \alpha A x_1 + \beta A x_2 = (\alpha + \beta) b \neq b$ in general.

In fact, the set of solutions to $\mathbf{Ax} = \mathbf{b}$ is not a linear subspace but an **affine** one – here, a particular solution x_p to $\mathbf{Ax} = \mathbf{b}$ (our 'origin' for our affine space) gives rise to **displacement vectors** $x - x_p$ which then comprise a linear subspace of solutions to the homogeneous problem $\mathbf{Ax} = \mathbf{0}$.

Four Possible linear equations depending on rank r

- $r = m = n$
  This kind of a matrix is **square** and **invertible**, with $rref(\mathbf{A}) = \mathbf{I}$. $\mathbf{A}x = b$ has only one **unique** solution.

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} \end{bmatrix}$$

- $r = m < n$
  This is the case of **full row rank**, where number of pivots = number of non-zero rows. This is the case where our system has ∞ solutions for every right hand side b. The column space is whole $\mathbb{R}^m$ and there are $n - r$ **special solutions** of $\mathbf{N}(\mathbf{A})$.

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix}$$

  Where **F** is the **special solution matrix** or the **free matrix.** A part of **F** can mix into **I**.

- $r = n < m$
  This is the case for **full column rank**, where number of pivots = number of columns. It has **no free variables** and therefore **no special solutions**. $\mathbf{A}x = b$ has 0 or 1 solutions.

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$$

- $r < n$ and $r < m$
  The case of **not full rank.** $\mathbf{A}x = b$ has 0 or ∞ solutions.

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$