1. Reviews and Miscellanea

This chapter is aimed towards helping students revise the basic concepts in undergraduate linear algebra that are relevant in the context of data science. We start with the presumption that the students already have the basic knowledge of vectors and matrices, and of various operations such as addition, scalar multiplication, transpose etc. We will revisit matrix multiplication with the intention of understanding the various algorithmic approaches for efficiently finding the product and for having different interpretation of the product in different contexts.

1.1 Fundamental Operations on Vectors and Matrices

Please note that by scalars we only mean real numbers. We will not be considering complex numbers for operations such as scalar multiplication of a vector. Similarly, we will mostly focus on vectors, matrices and tensors made of real numbers as most of the data-science applications involve non-complex data.

Please note that by a vector we will always mean a column vector irrespective of how we write it. This is mainly due to the convenience in writing that we will also be using notations such as $\mathbf{v} = \langle 1, 2, 3, 4 \rangle = (1, 2, 3, 4)^T$, rather than $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$.

The transpose of a matrix $A \in \mathbb{R}^{m \times n}$ is a matrix $A^T \in \mathbb{R}^{n \times m}$ such that the columns of A becomes the rows of A^T and the rows of A becomes the columns of A^T.

1.1.1 Dot Product of Vectors

We will define the general inner product of two vectors later. Here is the most commonly used inner product of two vectors that we need for various tasks.

Definition 1.1.1 — Dot Product. The dot product (or scalar product, or inner product) of two vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is a scalar given by

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \mathbf{v}^T \mathbf{u} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

■ **Example 1.1** Find the dot product of $\mathbf{u} = (1,2,3,4,5)^T$ and $\mathbf{v} = (3,-2,-1,0,1)^T$.

$$\mathbf{u}^T \mathbf{v} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \\ -1 \\ 0 \\ 1 \end{pmatrix} = (1)(3) + (2)(-2) + (3)(-1) + (4)(0) + (5)(1) = 2.$$

■

R The length or magnitude of a vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is a non-negative number, also called the Euclidean norm, which is given by

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

Clearly, $\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = \mathbf{v}^T \mathbf{v}$.

R The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ (that have the same tail) could be defined by the following relation

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$$

■ **Example 1.2** Find the angle between the vectors $\langle 2, 2, 1 \rangle$ and $\langle 1, 1, 2 \rangle$.

From above, we have

$$\cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{2+2+2}{3\sqrt{6}} = \frac{2}{\sqrt{6}}.$$

This implies that $\theta = \arccos \dfrac{2}{\sqrt{6}}$. ■

■ **Example 1.3** Pick up the correct option for the angle between the vectors $\langle 1, 3, -2 \rangle$ and $\langle 2, 2, 1 \rangle$:
(A) 0,
(B) $\pi/2$,
(C) acute angle,
(D) obtuse angle. ■

1.1.2 Outer Product of Vectors

Definition 1.1.2 — Outer Product. Given two vectors $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$, their outer product $\mathbf{u} \otimes \mathbf{v}$ is a rank-1 matrix of order $m \times n$ given by

$$\mathbf{u} \otimes \mathbf{v} := uv^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} (v_1, v_2, \cdots, v_n) = \begin{pmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & u_2v_2 & \cdots & u_2v_n \\ \vdots & \vdots & \cdots & \vdots \\ u_mv_1 & u_mv_2 & \cdots & u_mv_n \end{pmatrix}$$

■ **Example 1.4** Find the outer product of the vectors $\langle 2, 2, 1 \rangle$ and $\langle 3, -1, 2, 1 \rangle$

$$\mathbf{u} \otimes \mathbf{v} = \begin{pmatrix} 6 & -2 & 4 & 2 \\ 6 & -2 & 4 & 2 \\ 3 & -1 & 2 & 1 \end{pmatrix}$$

■

1.1.3 Matrix Multiplication

Definition 1.1.3 — Inner Product form for Matrix Multiplication. If $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{p \times n}$, then the product $C = AB \in \mathbb{R}^{m \times n}$ where the elements of C could be given as the inner product of rows of A with columns of B as

$$C_{ij} = A_{i:}B_{:j} = A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{ip}B_{pj} = \sum_{k=1}^{p} A_{ik}B_{kj}$$

where $A_{i:}$ gives the i-th row of A and $B_{:j}$ denotes the j-th column of B.

R The number of long floating point operations in the matrix product could be calculated to be $\mathscr{O}(mpn)$ which is $\mathscr{O}(n^3)$ for square matrices of order $n \times n$.

R The j-th column of the product AB is the product of A with the j-th column of B.

$$AB = A[B_{:1} \;\; B_{:2} \;\; \cdots \;\; B_{:n}] = [AB_{:1} \;\; AB_{:2} \;\; \cdots \;\; AB_{:n}]$$

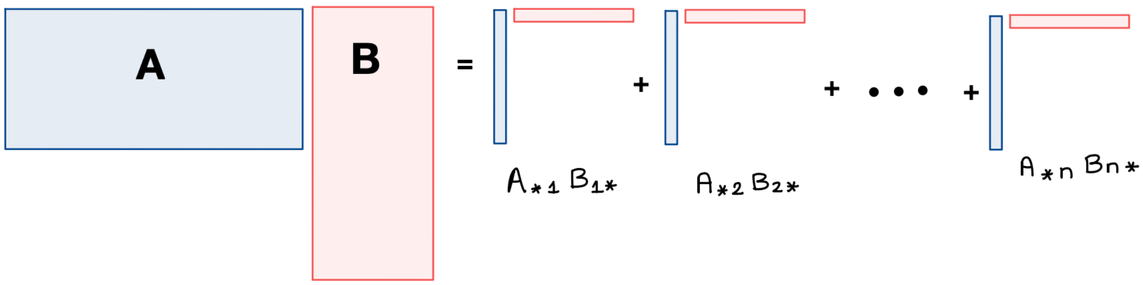
R The i-th row of the product AB is the product of the i-th row of A with B.

$$AB = \begin{bmatrix} A_{1:}B \\ A_{2:}B \\ \vdots \\ A_{m:}B \end{bmatrix}$$

Definition 1.1.4 — Outer Product Form of Matrix Multiplication. If $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{p \times n}$, then the product $C = AB \in \mathbb{R}^{m \times n}$ where C could be given as the sum of rank-1 matrices in $\mathbb{R}^{m \times n}$

found by taking the outer product of columns of A with rows of B

$$C = A_{:1}B_{1:} + A_{:2}B_{2:} + \cdots + A_{:p}B_{p:} = \sum_{k=1}^{p} A_{:k}B_{k:}$$



R The outer product form will be quite useful later especially in the context of dimension reduction and randomized matrix operations.

■ **Example 1.5** Write the following product as a sum of two rank-1 matrices.

$$\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

■

1.2 Motivating Examples

1.2.1 Manipulating Images

Please check iPython Notebook 1.

1.2.2 Working with Text Data

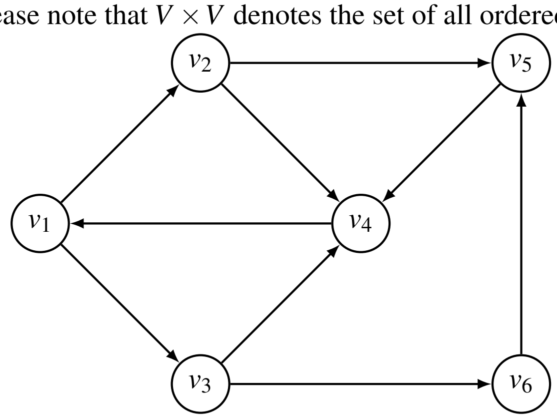
Please check iPython Notebook 2.

1.2.3 Matrix Representation of Graphs

A graph G is a collection of nodes (or vertices), V, and edges, E, between the nodes.

$$G = (V,\ E), \text{ where } V = \{v_1, v_2, \cdots, v_n\}, \text{ and } E \subset V \times V.$$

Please note that $V \times V$ denotes the set of all ordered pairs of nodes.



In the above graph, the nodes are

$$G = \{v_1, v_2, v_3, v_4, v_5, v_6\}$$

and the edges are

$$E = \{(v_1, v_2), (v_1, v_3), (v_2, v_4), (v_2, v_5), (v_3, v_4), (v_3, v_6), (v_4, v_1), (v_5, v_4), (v_6, v_5)\}$$

Definition 1.2.1 — Adjacency Matrix. The adjacency matrix of a **directed graph** with n number of nodes is a square matrix $A \in \mathbb{R}^{ntimesn}$,

$$A = [a_{ij}]_{n\times n} \text{ where } a_{ij} = \begin{cases} 1, & (i, j) \in E \\ 0, & \text{otherwise} \end{cases}$$

where (i, j) denotes an edge from node i to node j.

R The adjacency matrix of an **undirected graph** could be formed by treating every edge as bidirectional. So, for node $(v_i, v_j) \in E$, we also have $(v_j, v_i) \in E$. Therefore $a_{ij} = a_{ji} = 1$ and the adjacency matrix is symmetric.

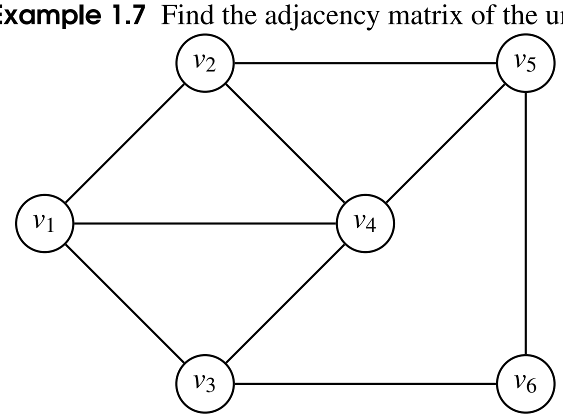
R If the graph has edges with weights, then we can create a **weighted adjacency matrix** for the graph where a_{ij} is the weight for node (v_i, v_j).

■ **Example 1.6** Find the adjacency matrix of the graph given above.

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

■

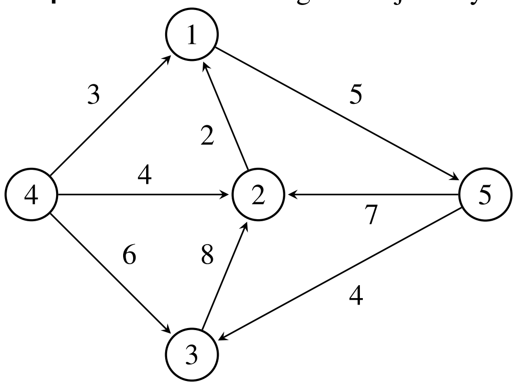
■ **Example 1.7** Find the adjacency matrix of the undirected graph given below



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

■

■ **Example 1.8** Find the weighted adjacency matrix of the following graphs



$$A = \begin{bmatrix} 0 & 0 & 0 & 5 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 8 & 0 & 0 & 0 \\ 3 & 4 & 6 & 0 & 0 \\ 0 & 7 & 4 & 0 & 0 \end{bmatrix}$$

■

Fact: For a simple graph, the ij-th element of the power of the adjacency matrix, $[A^m]_{ij}$, is equal to the number of different paths of length m between vertices i and j.

R The matrix defined by e^A, where A is the adjacency matrix of a simple graph could be computed by using spectral decomposition of $A = QDQ^T$ as

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k = \sum_{k=0}^{\infty} \frac{1}{k!} QD^kQ^T = Q\left(\sum_{k=0}^{\infty} \frac{1}{k!} D^k\right) Q^T = Qe^DQ^T.$$

where $e^D = \text{diag}(e^{\lambda_1}, e^{\lambda_2}, \cdots, e^{\lambda_n})$.

R The diagonal entries in e^A provide a measure of well-connectedness or *centrality* of different nodes, where as the off-diagonal entries provide a measure of *communicability* between every pair of nodes.

1.2.4 Structured Data as Matrices

Please check the iPython Notebooks for now.

1.3 Review of Linear Systems

Consider the following system of m linear equations in n variables.

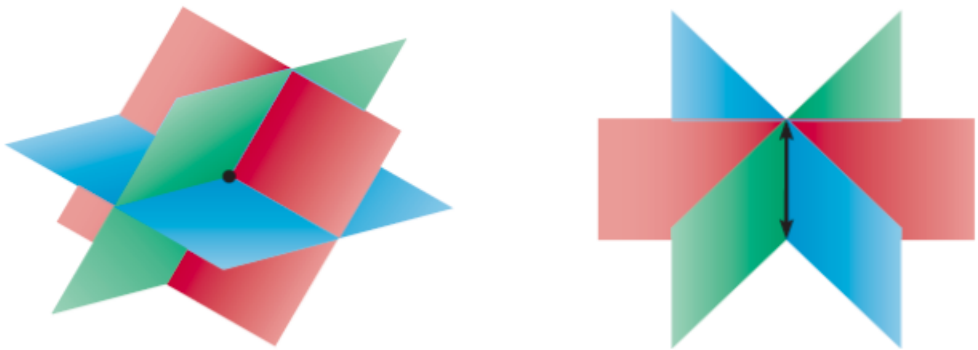
$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

The system could be represented in a compact form as $A\mathbf{x} = \mathbf{b}$, where

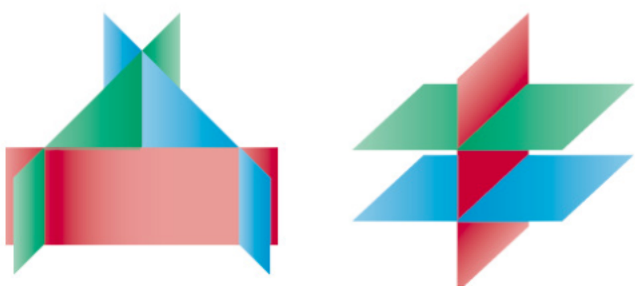
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\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

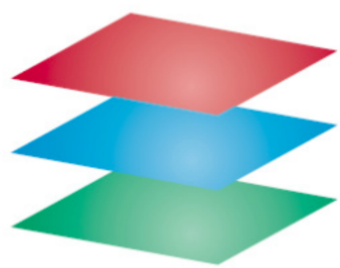
R The solution, if any, of a linear system represents the **points of intersection of hyper-planes** given by the linear equations.[1]

R A system of equations is called **consistent** if there exists at least one solution.



R A system of equations is called **inconsistent** if there is no solution that is satisfied by all the equations.





R A linear system is called *over-determined* if there are more equations than unknown $m > n$. It is called *under-determined* if there are more variables than equations $m < n$. We have a square system with equal number of equations and variables when $m = n$.

Column-view of a linear system
The solutions of a linear system also represent the **linear combination of the columns** of a matrix A to obtain a vector $\mathbf{b}$.

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{1.1}$$

[1]The images are from some source on web

or simply as

$$x_1A_1 + x_2A_2 + \cdots + x_nA_n = \mathbf{b}$$

■ **Example 1.9** A special solution of the system

$$\begin{pmatrix} 2 & 0 & 5 \\ 0 & 1 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 5 \end{pmatrix}$$

which has the column-view

$$x_1 \begin{pmatrix} 2 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} 5 \\ 9 \end{pmatrix} = \begin{pmatrix} 8 \\ 5 \end{pmatrix}$$

could be found just by a visual inspection of the system as $x_1 = 4$, and $x_2 = 5$, and $x_3 = 0$. ■

1.3.1 Gaussian Elimination

You are advised to review the Gaussian elimination method for solving a linear system. You will find this link of a Georgia tech online textbook useful.

Another resource is Ken Kuttler

The central idea is to create an augmented matrix from the system and convert it into an upper triangular format by using Gaussian elimination. Subsequently, one can use back-substitution to find all the unknowns.

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right]$$

Applications of Gaussian Elimination

- Solving linear systems.
- Finding rank of a rectangular matrix.
- Finding basis vectors for a set of given vectors.
- Finding the determinant of a matrix.
- Finding LU decomposition of a matrix.
- Determining if a symmetric matrix is positive definite.

■ **Example 1.10** Find the solution of the following linear system by Gaussian elimination.

$$\begin{array}{rrrr} 3x & -6y & +9z = & 0 \\ 4x & -6y & +8z = & -4 \\ -2x & -y & +z = & 7 \end{array}$$

■

■ **Example 1.11** Find a basis representation of the following subspace V of $\mathbb{R}^4$.

$$V = \text{span}\left\{ \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \\ 4 \end{pmatrix}, \begin{pmatrix} 0 \\ 7 \\ 5 \\ 15 \end{pmatrix}, \begin{pmatrix} -4 \\ 1 \\ -5 \\ 1 \end{pmatrix} \right\}$$

■

1.3.2 Gauss-Jordan Elimination

You are advised to review the Gauss-Jordan elimination method for solving a linear system. You will find this resource by Ken Kuttler useful.

The basic idea in Gauss-Jordan elimination is to bring the augmented matrix from the system to reduced-row-echelon-form (RREF) which has a special staircase patterns as shown below.

$$B = \begin{bmatrix} 1 & 0 & * & 0 & * & * \\ 0 & 1 & * & 0 & * & * \\ 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

■ **Example 1.12** Find the solution of the following linear system by Gauss-Jordan elimination.

$$\begin{aligned} 3x \quad -6y \quad +9z &= \quad 0 \\ 4x \quad -6y \quad +8z &= -4 \\ -2x \quad -y \quad +z &= \quad 7 \end{aligned}$$

■

1.3.3 Under-determined Systems and General Solutions

When the number of unknown variables are more than the number of equation in a linear system, it is called under-determined system. If such a system is consistent, there are infinitely many solutions. However, one can find a general family of solutions by (A) finding a special or particular solution of the system $A\mathbf{x} = \mathbf{b}$ and (B) by finding a general solution of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$, and finally taking their combination.

■ **Example 1.13** Find the general solution of the following *homogeneous system* $A\mathbf{x} = \mathbf{0}$ where

$$A = \begin{pmatrix} 1 & -2 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 1 & 5 \end{pmatrix}$$

Please note that the first, third, and forth columns above are pivot columns. This implies that x_1, x_2 and x_4 are pivot (basic) variables, whereas x_2 and x_5 are two free variables. This implies that there will be exactly two independent solutions of the homogeneous system.

Let's rewrite the above equation in the column view form of a linear system as in Equation (1.1).

$$x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} + x_5 \begin{pmatrix} 4 \\ -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

One can see that the second column is just a multiple of the first (pivot) column. By taking a contribution of zero from other columns one can form a solution $(-2,-1,0,0,0)^T$. Similarly, the fifth column could be given as a linear combination of all the pivot columns that leads to an independent solution $(4,0,-3,5,-1)^T$.

Hence the general solution of the system $A\mathbf{x} = \mathbf{0}$ is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = k_1 \begin{pmatrix} -2 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + k_2 \begin{pmatrix} 4 \\ 0 \\ -3 \\ 5 \\ -1 \end{pmatrix}$$

where k_1 and k_2 are two arbitrary constants. ■

■ **Example 1.14** Find the general solution of the following *non-homogeneous system* $A\mathbf{x} = \mathbf{b}$ given below

$$\begin{bmatrix} 1 & 0 & -3 & 4 \\ 0 & 1 & 3 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}.$$

Once again, by writing the equation in column view form as in Equation (1.1),

$$x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + x_3 \begin{pmatrix} -3 \\ 3 \end{pmatrix} + x_4 \begin{pmatrix} 4 \\ -7 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

we can find a special (or particular) solution right away $(3, 5, 0, 0)^T$. Furthermore, two independent solutions of the corresponding homogeneous system $A\mathbf{x} = \mathbf{0}$ could be found as in the previous example by a visual inspection. They are $(-3,3,-1,0)^T$ and $(4,-7,0,-1)^T$. Hence the general solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 0 \\ 0 \end{pmatrix} + k_1 \begin{pmatrix} -3 \\ 3 \\ -1 \\ 0 \end{pmatrix} + k_2 \begin{pmatrix} 4 \\ -7 \\ 0 \\ -1 \end{pmatrix}$$

where k_1 and k_2 are to arbitrary constants. ■

R The examples above easy to solve as they are already in the reduced row echelon form (RREF). If a given system is not in RREF, then we should first use Gauss-Jordan elimination to bring the rectangular system in the RREF form.

■ **Example 1.15** Find the general solution of the following *non-homogeneous system* $A\mathbf{x} = \mathbf{b}$ given below

$$\begin{bmatrix} 3 & 0 & -9 & 12 \\ 2 & 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 9 \\ 11 \end{bmatrix}.$$

The augmented matrix for the above system changes as follows by elementary row operations

$$\left(\begin{array}{cccc:c} 3 & 0 & -9 & 12 & 9 \\ 2 & 1 & -3 & 1 & 11 \end{array}\right) \xrightarrow{\frac{1}{3}R_1} \left(\begin{array}{cccc:c} 1 & 0 & -3 & 4 & 3 \\ 2 & 1 & -3 & 1 & 11 \end{array}\right) \xrightarrow{R_2+(-2)R_1} \left(\begin{array}{cccc:c} 1 & 0 & -3 & 4 & 3 \\ 0 & 1 & 3 & -7 & 5 \end{array}\right)$$

which is one of the solved examples given above. ∎

1.4 Vector Spaces and Subspaces

As a first introduction to vectors, you may have worked with vectors in two or three dimensions in earlier college mathematics or physics courses. They represent physical quantities such as force, momentum, velocity etc. These vectors have both magnitudes and directions. One can find the sum and difference of these vectors as a third vector by using triangle law or parallelogram law. One can also find their scalar product by multiplying the magnitude by the given scalar. These vectors satisfy certain properties that are more general and we can talk about vectors in higher dimension, and in a more abstract sense.

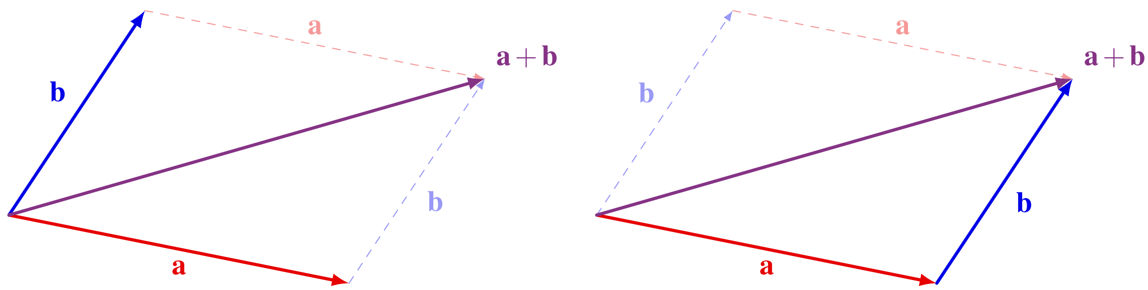


Figure 1.1: Parallelogram law and triangle law for vector addition.

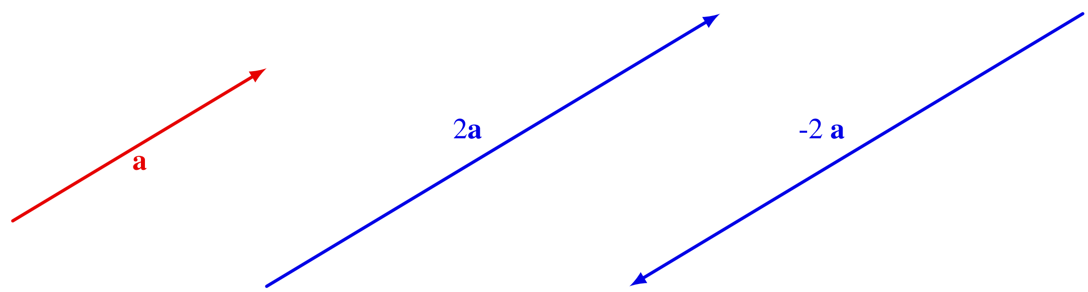


Figure 1.2: Scalar product of a vector with a positive real number and a negative real number.

A real vector space (a vector space defined over the set of real numbers $\mathbb{R}$) is a nonempty set V whose members are certain mathematical constructs called vectors. This set is equipped with two operations: Addition of members, '+', and multiplication '$\cdot$' of members by scalars from $\mathbb{R}$. The set V is closed under these two operations, meaning that the result of these operations are also members of V.

In additions, the following properties are also satisfied. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in V, and α and β be two scalars.

1. Associativity: $\mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}$.

2. Commutativity: $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$.

3. Additive identity: There exists a vector $\mathbf{0} \in V$ such that $\mathbf{u}+\mathbf{0} = \mathbf{0}+\mathbf{u} = \mathbf{u}$.

4. Additive inverse: For every $\mathbf{u} \in V$ there exists a $-\mathbf{u} \in V$ such that $\mathbf{u}+(-\mathbf{u}) = -\mathbf{u}+\mathbf{u} = \mathbf{0}$.

5. $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$.

6. Multiplicative identity: There exist a scalar 1 such that $1\mathbf{u} = \mathbf{u}$.

7. Distributive property of scalars over vector addition: $\alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u}+\alpha\mathbf{v}$.

8. Distributive property of vectors on scalar addition: $(\alpha+\beta)\mathbf{u} = \alpha\mathbf{u}+\beta\mathbf{u}$.

The set, V, of these vectors combined with the operations '+', and , '$\cdot$', is called a vector space over $\mathbb{R}$ and is denoted by $(V, +, \cdot)$.

■ **Example 1.16** One can verify that the set of n-tuples with element-wise addition and scalar multiplication operations define a vectors space over the set of real numbers? This is usually denoted by $\mathbb{R}^n$

$$V = \{\mathbf{v} : \mathbf{v} = (v_1, v_2, \cdots, v_n),\ \text{where } v_i \in \mathbb{R}\}$$

■

■ **Example 1.17** Set of continuous functions over interval $[0, 1]$ together with the usual addition/subtraction of functions and multiplication by real numbers form a vector space. ■

Subspace

Definition 1.4.1 — Subspace. A nonenpty subset U of a vector space V is called a subspace if for every $\mathbf{u}, \mathbf{v} \in U$ and scalars α and β, the vector $\alpha\mathbf{u}+\beta\mathbf{v} \in U$. A subspace is a vector space in its own right over the same underlying field.

For more detailed explanation of subspaces please check this Georgia Tech text.

R Determination of whether a given subset U of a vector space V is a subspace or not could be made by checking the following in a sequence

Step1: Is $\mathbf{0}$ a member of the subset?

Step 2: Is $\alpha\mathbf{u} \in U$ for all $\mathbf{u} \in U$?

Step 3: Is $\mathbf{u}+\mathbf{v} \in U$ for all $\mathbf{u}, \mathbf{v} \in U$?

If any the three steps does not hold, the given subset is not a subspace.

■ **Example 1.18** Which of the following sets are vector subspaces of $\mathbb{R}^3$?

A. $S = \{(\lambda, \lambda+\mu^3, \lambda-\mu^3) \in \mathbb{R}^3 \mid \lambda, \mu \in \mathbb{R}\}$

B. $T = \{(\lambda^2, -\lambda^2, 0) \in \mathbb{R}^3 \mid \lambda \in \mathbb{R}\}$

C. $U = \{(x, y, z) \in \mathbb{R}^3 \mid x-2y+3z = 0\}$

D. $V = \{(x, y, z) \in \mathbb{R}^3 \mid y \in \mathbb{Z}\}$

■

■ **Example 1.19** Let A be an $m \times n$ matrix. Then we may think of A as a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ (why?). Let $V \subseteq \mathbb{R}^m$ be the range, or image, of this transformation. That is,

$$V = \{\mathbf{v} \in \mathbb{R}^m \mid A\mathbf{u} = \mathbf{v} \text{ for some } \mathbf{u} \in \mathbb{R}^n\}.$$

Show that V is a vector space. ■

Inner Product

In the previous sections we saw the definition of dot product of two vectors. This was one special kind of inner product. Here is the general definition of the inner product of vector over real or complex fields of scalars.

Definition 1.4.2 — Inner Product. The inner product of two vectors in a vector space V over the field F of real or complex numbers is a binary operation $\langle \cdot, \cdot \rangle : V \times V \to F$ satisfying the following properties

1. (Conjugate symmetry) $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
2. (Positive definiteness) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$, and $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$.
3. $\langle \mathbf{u} + \mathbf{w}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle$.
4. $\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$.

A vector space that has an inner product defined in the above way is called an *inner product space*. inner product grants a vector space the concepts of length, angle and orthogonality.

■ **Example 1.20** Prove the following

- $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.
- $\langle \mathbf{u}, \alpha \mathbf{v} \rangle = \overline{\alpha} \langle \mathbf{u}, \mathbf{v} \rangle$.

■

■ **Example 1.21** The Euclidean space $\mathbb{R}^n$ is an inner product space where the inner product is given by the dot product (or scalar product)

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T \mathbf{u} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

■

R We can define an angle between two nonzero vectors in an inner product space whose cosine is given by

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}}.$$

R The notion of inner product could be used to define the concept of lengths of vectors called norms. For example in Euclidean space $\mathbb{R}^n$, a special norm called Euclidean norm or l_2-norm could be defined as

$$\|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

which implies that

$$\|\mathbf{v}\|_2^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \mathbf{v}^T \mathbf{v}.$$

We will visit the idea of norms in more detail in later sections.

Fact 1.4.1 — Cauchy-Schwarz Inequality. Let $\mathbf{u}$, and $\mathbf{v}$ be vectors in a given vector space with certain inner product and associated norm defined appropriately. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|.$$

where equality holds in the case of parallel vectors.

Linear Independence, Span and Basis

Given a set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_r}\}$, their linear combination is a vector $\mathbf{v} = \alpha_1 \mathbf{v_1} + \alpha_2 \mathbf{v_2} + \cdots + \alpha_r \mathbf{v_r}$ for the scalars α_i $(i = 1,\ 2,\ \cdots,\ r)$.

Definition 1.4.3 — Linear Independence. A set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_r}\}$ are said to be linearly independent if

$$\alpha_1 \mathbf{v_1} + \alpha_2 \mathbf{v_2} + \cdots + \alpha_r \mathbf{v_r} = 0 \Longrightarrow \alpha_i = 0 \ \text{ for } \ i = 1, \cdots, r.$$

In other words, linearly independent vectors can not be combined to form a zero vector, or, one of these vectors can not be given as a linear combination of the remaining.

■ **Example 1.22** Determine whether the following set of vectors are linearly independent or not.

$$T = \{\langle 1,2,1,0,0\rangle, \langle 1,1,0,1,1\rangle, \langle 1,0,0,1,1\rangle\}$$

Let us check if the following systems has any non-trivial solutions or not

$$\alpha_1 \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

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

As all columns are pivot columns, they can not be combined to produce a zero vector in any way other than $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$. So the given set of vectors are linearly independent. ■

Definition 1.4.4 — Span. Given a set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_m}\}$, the span of these vectors is the set of all vectors that could be written as their linear combination, i.e.,

$$\text{span}\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_m}\} = \{\mathbf{v} : \mathbf{v} = \alpha_1 \mathbf{v_1} + \alpha_2 \mathbf{v_2} + \cdots + \alpha_m \mathbf{v_m}\} \quad \text{ for } \ \alpha_i \in \mathbb{R}.$$

Definition 1.4.5 — Basis. A set of linearly independent vectors $\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_m}\}$ is said to form a basis of a vector space V if every element in V could be written as a linear combination of vectors from this given set, in other words

$$V = \text{span}\{\mathbf{v_1}, \mathbf{v_2}, \cdots, \mathbf{v_m}\}.$$

■ **Example 1.23** Find a basis for the following subspaces of $\mathbb{R}^4$.

A. All vectors whose components are equal.

B. All vectors whose components add to zero.

C. The column space of I_4.

■

Definition 1.4.6 — Orthogonality. Two vectors **u** and **v** are mutually orthogonal or perpendicular to each other if their dot product is zero.

$$\mathbf{u} \perp \mathbf{v} \iff \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T \mathbf{u} = 0.$$

A vector $\mathbf{u} \in \mathbb{R}^n$ is orthogonal to a subspace V of $\mathbb{R}^n$ written as $\mathbf{u} \perp V$ if for every $\mathbf{v} \in V$, we have $\mathbf{u} \perp \mathbf{v}$.

Two subspaces U and V of a vector space are mutually orthogonal if every member of one subspace is orthogonal to every member of the other, or

$$U \perp V \iff \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{v}^T \mathbf{u} = 0, \quad \forall (\mathbf{u} \in U,\ \mathbf{v} \in V).$$

Definition 1.4.7 — Orthogonal Complement. If U is a subspace of $\mathbb{R}^n$ then

$$U^\perp := \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \perp U\}$$

is called the *orthogonal complement* of U.

■ **Example 1.24** Prove that $U^\perp$ defined above is a subspace of $\mathbb{R}^n$. ■

R It could be verified easily that

$$(U^\perp)^\perp = U$$

and that $\mathbb{R}^n$ is a **direct sum of the subspaces** U and $U^\perp$. This means that any vector $\mathbf{x} \in \mathbb{R}^n$ could be written uniquely as $\mathbf{x} = \mathbf{y} + \mathbf{z}$ where $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$.

$$U \oplus U^\perp = \mathbb{R}^n.$$

■ **Example 1.25** Verify that the following subset of $\mathbb{R}^3$ is a subspace. Find its orthogonal complement.

$$V = \{(x, y, z) \in \mathbb{R}^3 : 2x - y + 3z = 0\}$$

■

■ **Example 1.26** Let $\mathbf{v} = \langle a, b, c \rangle \in \mathbb{R}^3$. How can you describe $(\text{span}\{\mathbf{v}\})^\perp$. ■

■ **Example 1.27** Find a basis representation of U and $U^{\perp}$, where

$$U = \left\{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1 + 2x_2 + 3x_3 + 4x_4 = 0\right\}$$

■

1.5 Vectors, Matrices and Tensors

1.5.1 Rank

Definition 1.5.1 — Rank of a matrix. Suppose matrix $A \in \mathbb{R}^{m \times n}$. Its rank is the number of linearly independent rows or columns. Clearly $\text{rank}(A) \leq \min\{m, n\}$.

One can find the rank by using Gaussian elimination on the matrix for its row echelon form which also informs about the independent columns of the matrix.

■ **Example 1.28** Find the rank of the matrix by using Gaussian elimination

$$A = \begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & -1 & -4 & 8 \\ -1 & 1 & 3 & -5 \\ -1 & 2 & 5 & -6 \\ -1 & -2 & -3 & 1 \end{bmatrix} \xrightarrow{REF} \begin{bmatrix} 1 & 2 & 3 & -1 \\ 0 & -5 & -10 & 10 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Three pivot columns implies that the rank in this example is 3.

■

1.5.2 Some Special Matrices

Study about the following types of matrices.

- Permutation Matrices.
- Upper and Lower Triangular Matrices
- Diagonal Matrices
- Symmetric Matrices
- Orthogonal Matrices
- Normal Matrices

1.5.3 The Four Fundamental Subspace

A matrix $A \in \mathbb{R}^{m\times n}$ represents a linear transformation from a space $\mathbb{R}^n$ to anther space $\mathbb{R}^m$. We can define the following four subspaces associated with matrix A.

Definition 1.5.2 — Column Space. The column space (or image) col(A) of A is a subspace of $\mathbb{R}^m$ spanned by all the columns of the matrix:

$$\mathscr{C}(A) = \{\mathbf{y} \in \mathbb{R}^m : \mathbf{y} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}.$$

Definition 1.5.3 — Null Space. The null space (also called kernel) of A is a subspace of $\mathbb{R}^n$:

$$\mathscr{N}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}.$$

Definition 1.5.4 — Row Space. The row space row(A) of A is a subspace of $\mathbb{R}^n$ which is spanned by the row-vectors of A:

$$\mathscr{R}(A) = \mathscr{C}(A^T) = \left\{\mathbf{z} \in \mathbb{R}^n : \mathbf{z} = A^T\mathbf{y} \text{ for some } \mathbf{y} \in \mathbb{R}^m\right\}.$$

Definition 1.5.5 — Left Null Space. Left null space of A, or null space of A^T, is a subspace of $\mathbb{R}^m$:

$$\mathscr{N}(A^T) = \left\{\mathbf{y} \in \mathbb{R}^m : A^T\mathbf{y} = \mathbf{0}\right\}.$$

Theorem 1.5.1 — Fundamental Theorem of Linear Algebra. The above four subspaces satisfy the following properties:

- $\mathscr{R}(A) \perp \mathscr{N}(A)$, or $\mathscr{C}(A^T) \perp \mathscr{N}(A)$.
- $\mathscr{C}(A) \perp \mathscr{N}(A^T)$.
- $\mathscr{R}(A) \oplus \mathscr{N}(A) = \mathbb{R}^n$.
- $\mathscr{C}(A) \oplus \mathscr{N}(A^T) = \mathbb{R}^m$.
- If rank(A) = r, then
  - $\dim \mathscr{R}(A) = \dim \mathscr{C}(A) = r$,
  - $\dim \mathscr{N}(A) = n - r$,
  - and $\dim \mathscr{N}(A^T) = m - r$.

Here $\perp$ stands for orthogonal (perpendicular), and $\oplus$ stands for direct sum of subspaces. For

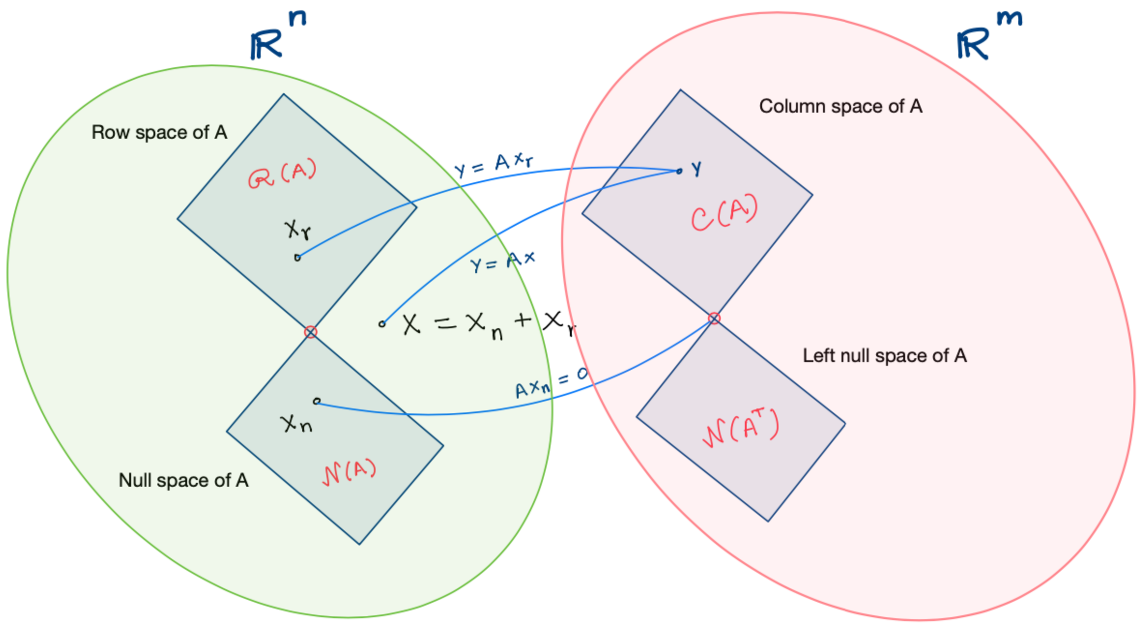


Figure 1.3: Relationship between the four fundamental subspaces. Notice that the row-space and null space are orthogonal complements. Similarly column-space and left null space are orthogonal complements.

example, $\mathscr{R}(A) \oplus \mathscr{N}(A) = \mathbb{R}^n$ means that every element $x \in \mathbb{R}^n$ could be written as $x = x_r + x_n$, where $x_r \in \mathscr{R}(A)$ and $x_n \in \mathscr{N}(A)$.

■ **Example 1.29** Prove the following relation between the two subspaces of $\mathbb{R}^n$:

$$\mathscr{R}(A)^{\perp} = \mathscr{N}(A)$$

■

■ **Example 1.30** Prove the following relation between the two subspaces of $\mathbb{R}^m$:

$$\mathrm{col}(A)^{\perp} = \mathscr{N}(A^T)$$

■

■ **Example 1.31** Find the null space of the matrix A as a span of some basis vectors

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

■

■ **Example 1.32** Consider the following matrix. What is its rank? $A = \begin{pmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

Find the dimensions of the four subspaces associated with this matrix. ■

■ **Example 1.33** Find the basis representation of the four fundamental subspaces of the following matrix. The echelon form of the matrix is also provided

$$A = \begin{bmatrix} 1 & -2 & 1 & -5 \\ 2 & 1 & 7 & 5 \\ 1 & -1 & 2 & -2 \end{bmatrix} \rightarrow RREF \rightarrow \begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

■

1.6 Determinants

The determinant of a square matrix A, denoted by $|A|$ or by $\det(A)$ is a special number that could be thought of as the higher dimensional equivalence of the volume of a parallelepiped in hyperspace ($\mathbb{R}^n$) with a given orientation (represented by the sign). For $2x2$ and $3x3$ matrices, we can find this number easily from the formulas

$$det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

and

$$det \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}.$$

The determinant for higher order matrices could be defined in a similar manner recursively given by the following.

Definition 1.6.1 — Laplace's Formula for Determinant. Let $A \in \mathbb{R}^{n\times n}$

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij} \quad \text{for a fixed row } i,$$

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij} \quad \text{for a fixed column } j.$$

where minor M_{ij} is the determinant of the submatrix of A having all rows and columns but the i-th row and j-th column.

R Using Laplace expansion for finding the determinant of a square matrix is not an efficient way as one can see that the recursive definition has a complexity $\mathcal{O}(n!)$ in long operations. A practical way to calculate determinant is use to Gaussian elimination. In this process the product of the pivots provide the determinant for even number of row swaps and negative of the determinant for odd number of row swaps.

Properties of Determinants Let $A \in \mathbb{R}^{n\times n}$

- $\det(I_n) = 1$, where I_n is the $n \times n$ identity matrix.
- $\det\left(A^\top\right) = \det(A)$, where $A^\top$ denotes the transpose of A.
- When a row or column is scaled by a number k, the determinant is multiplied by a the factor k.
- The determinant changes to its negative for every row or column swap.
- The determinant remains unchanged by the elementary row operation of adding a multiple of a row to another row, i.e. $R_j \Leftarrow R_j + kR_i$ where $i \neq j$.
- For square matrices A and B of equal size,
$$\det(AB) = \det(A)\det(B).$$
- $\det\left(A^{-1}\right) = \dfrac{1}{\det(A)} = [\det(A)]^{-1}$.
- $\det(cA) = c^n \det(A)$, for an $n \times n$ matrix A.
- Determinant of a matrix is the product of all its eigenvalues.

1.6.1 Trace

Another important function of a square vector is the trace function that is defined below.

Definition 1.6.2 — Trace. Trace of a square matrix is the sum of its diagonal elements

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$$

Properties of Trace: Let $A \in \mathbb{R}^{n \times n}$

- $\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A})$,
- $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$,
- $\mathrm{tr}(k\mathbf{A}) = k\,\mathrm{tr}(\mathbf{A})$,
- $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$,
- For any nonsingular matrix $\mathbf{S}$: $\mathrm{tr}(\mathbf{SAS}^{-1}) = \mathrm{tr}(\mathbf{A})$ (i.e., invariance under change of base)
- Trace is the sum of the eigenvalues of a matrix.
- $\mathrm{tr}(\mathbf{AB}) \neq \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B})$

R Consider the rectangular matrices $A, B \in \mathbb{R}^{m \times n}$, then

$$\mathrm{tr}\left(\mathbf{A}^\top \mathbf{B}\right) = \mathrm{tr}\left(\mathbf{A}\mathbf{B}^\top\right) = \mathrm{tr}\left(\mathbf{B}^\top \mathbf{A}\right) = \mathrm{tr}\left(\mathbf{B}\mathbf{A}^\top\right) = \sum_{i,j} A_{ij} B_{ij} = \sum_{i,j} (A \circ B)_{ij}.$$

where '$\circ$' stands for element-wise product of two matrices of same dimension

■ **Example 1.34** Verify the above remark for the following matrices

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} -2 & 1 & 1 \\ 1 & 3 & 2 \end{pmatrix}.$$

■

1.7 Inverse and Nonsingularity

Definition 1.7.1 A square matrix is said to have an inverse, or is non-singular (invertible) if all its rows and all its columns are linearly independent. The inverse of a square matrix A is denoted by A^{-1}. It has the following property:

$$A^{-1}A = AA^{-1} = I$$

■ **Example 1.35** Are the matrices $A = \begin{pmatrix} 1 & 2 & 1 \\ 4 & 4 & 5 \\ 6 & 7 & 7 \end{pmatrix}$ and $B = \begin{pmatrix} -7 & -7 & 6 \\ 2 & 1 & -1 \\ 4 & 5 & -4 \end{pmatrix}$ inverses of one another? Explain.

The product of the above two matrices is indeed an identity matrix. So, they are inverses of each other. ■

R Verify that the following inverse formula for 2×2 matrices hold

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

R In general one can find the inverse by

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A)$$

where

$$\operatorname{adj}(\mathbf{A}) = \mathbf{C}^{\mathsf{T}} = \left((-1)^{i+j}\mathbf{M}_{ji}\right)_{1 \le i,j \le n}.$$

where M_{ji} is the determinant of the submatrix of A without the j-th row and i-th column.
In practice this method is not efficient. So, we use Gauss-Jordan elimination to find the inverse of a matrix. See an example later.

■ **Example 1.36** Use Gaussian elimination (also known as Gauss-Jordan elimination in this context) to find the inverse of the matrix.

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

The basic idea behind using Gauss-Jordan elimination for finding the inverse of a matrix A is to solve the matrix equation $AX = I$. At the end of the process the right half of the augmented matrix gives the inverse.

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 2 & 1 & 3 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & -3 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

Hence

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

■

Properties of matrix inverses
Let $A, B \in \mathbb{R}^{n\times n}$ and $k \neq 0$ be scalar. Then

- $$(A^{-1})^{-1} = A$$
- $$(kA)^{-1} = k^{-1}A^{-1}$$
- $$(A^\top)^{-1} = (A^{-1})^\top$$
- $$(AB)^{-1} = B^{-1}A^{-1}.$$
- $$\det\left(A^{-1}\right) = (\det(A))^{-1} = \frac{1}{\det(A))}.$$

■ **Example 1.37** Prove the following for the inverse of the given block-matrix. Suppose that both A and D square matrices and are invertible. Then

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}\right)^{-1} & \mathbf{0} \\ \mathbf{0} & \left(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}\right)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{C}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}.$$

■

Conditions for invertibility If $A \in \mathbb{R}^{n\times n}$ is a square matrix A, then the following conditions are equivalent.

- A is an invertible matrix.
- $\operatorname{rank}(A) = n$.
- All the rows of A are linearly independent.
- All the columns of A are linearly independent.
- The homogeneous system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
- The system $A\mathbf{x} = \mathbf{b}$ has a unique solution.
- Determinant of A is non-zero, $\det(A) \neq 0$.
- All the eigenvalues of A are non-zeros.

1.8 Eigenvalues and Eigenvectors

Definition 1.8.1 — Eigenvalues and Eigenvectors. An eigenvector of a square matrix $A \in \mathbb{R}^{n\times n}$ is a special non-zero vector $\mathbf{v} \in$
$mathbbC^n$ such that

$$A\mathbf{v} = \lambda\mathbf{v},$$

where λ is called the associated eigenvalue. We will refer to $(\lambda, \mathbf{v})$ as an eigen-pair of A.

R In the process of finding the eigen-pairs of A, one needs to solve a system of equations

$$A\mathbf{v} = \lambda\mathbf{v} \Longrightarrow (A - \lambda I)\mathbf{v} = \mathbf{0}$$

This implies that $A - \lambda I$ is a singular (non-invertible) matrix. So, $\det(A - \lambda I) = 0$.

The left side of this equation is a polynomial in λ that we call the **characteristic polynomial** of the square matrix A, denoted by $p_A(x)$.

One can find all the eigenvalues by solving the **characteristic equation** given by $p_A(\lambda) = 0$.

Subsequently, one can solve the system $(A - \lambda I)\mathbf{v} = \mathbf{0}$ to find the associated eigenvectors for every eigenvalue, one after another.

■ **Example 1.38** Prove that the diagonal elements of a triangular matrix are its eigenvalues. ■

■ **Example 1.39** Find the eigenvalues and eigenvectors of both A and A^2, where $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. ■

Theorem 1.8.1 — Cayley-Hamilton Theorem. Every square matrix satisfies its characteristic polynomial. Let $A \in \mathbb{R}^{n\times n}$

$$p(\lambda) = det(A - \lambda I) = 0 \Rightarrow p(A) = 0 \in \mathbb{R}^{n\times n}$$

■ **Example 1.40** Express the inverse of A in terms of powers of A if the characteristic polynomial of A is $p_A(x) = x^3 - 4x + 7$.

Please note that zero is not a solution of the characteristic polynomial. So, the matrix is A invertible. From the Cayley-Hamilton Theorem $A^3 - 4A + 7I = 0$. This implies that $I = \frac{1}{7}(4A - A^3)$. On multiplying both side by A^{-1} we get

$$A^{-1} = \frac{4}{7}(A^{-1}A - A^{-1}A^3) = \frac{4}{7}I - \frac{1}{7}A^2.$$

■

■ **Example 1.41** Prove that the eigenvalues of a complex Hermitian matrix ($A^H = \bar{A}^T = A$) and therefore of a symmetric matrix are real.

As $A\mathbf{v} = \lambda\mathbf{v}$. Taking Hermitian transpose on both sides gives $\mathbf{v}^H A^H = \mathbf{v}^H A = \bar{\lambda}\mathbf{v}^H$. Now,

$$A\mathbf{v} = \lambda\mathbf{v} \implies \mathbf{v}^H A\mathbf{v} = \lambda\mathbf{v}^H\mathbf{v} \implies \bar{\lambda}\mathbf{v}^H\mathbf{v} = \lambda\mathbf{v}^H\mathbf{v} \implies (\bar{\lambda} - \lambda)\mathbf{v}^H\mathbf{v} = 0 \implies \bar{\lambda} - \lambda = 0.$$

The last equation is implied from the fact that $\mathbf{v}^H\mathbf{v} \neq 0$ in the definition of eigenvectors.

■

■ **Example 1.42** Find the eigenvalues and eigenvectors of the matrix

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

■

1.8.1 Diagonalization and Eigen-decomposition

Definition 1.8.2 — Similar Matrices. Two square matrices of the same order $A, B \in \mathbb{R}^{n\times n}$ are called similar if there exists a non-singular matrix $P \in R^{n\times n}$ such that $A = PBP^{-1}$.

■ **Example 1.43** Prove that two similar matrices have the same characteristic polynomials.

Let $A = PBP^{-1}$. Then

$$\begin{aligned}
\det[A - \lambda I] &= \det\left[PBP^{-1} - \lambda I\right] \\
&= \det\left[P(B - \lambda I)P^{-1}\right] \\
&= \det(P)\det(B - \lambda I)\det\left(P^{-1}\right) \\
&= \det(P)\det(B - \lambda I)\frac{1}{\det(P)} \\
&= \det(B - \lambda I)
\end{aligned}$$

■

Properties of similarity of two matrices Let $A = PBP^{-1}$. Then

- Two similar matrices have the same characteristic equations

$$\det(A - \lambda I) = \det(B - \lambda I) = 0.$$

- Same characteristic equations imply that similar matrices have the same set of eigenvalues and their multiplicities.
- This implies that similar matrices have the same trace and determinant.
- If the eigen-pairs of A are $(\lambda_i, \mathbf{v}_i)$, then the eigen-pairs of B are $(\lambda_i, P^{-1}\mathbf{v}_i)$.
- Ranks of two similar matrices are the same.

Once we have all these nice properties of tow similar matrices, it is easy to see that it would be great to have a given matrix similar to a diagonal matrix.

Definition 1.8.3 — Diagonalization. A matrix $A \in C^{m\times n}$ is called diagonalizable if it is similar to a diagonal matrix $D \in C^{m\times n}$. This means that there exists a non-singular matrix P such that $A = PDP^{-1}$, where D is a diagonal matrix.

R Finding the above diagonalization of the matrix is also called **eigen-decomposition** for the simple reason that one finds all the eigen-pairs of the matrix in the process of such a factorization.

As the diagonal entries of a diagonal matrix D happen to be its eigenvalues, we have

$$
\begin{aligned}
A = P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} P^{-1} \\
\Longleftrightarrow AP = P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\
\Longleftrightarrow A[\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots\ \mathbf{v}_n] = [\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots\ \mathbf{v}_n] \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \\
\Longleftrightarrow A[\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots\ \mathbf{v}_n] = [\lambda_1\mathbf{v}_1,\ \lambda_2\mathbf{v}_2,\ \cdots\ \lambda_n\mathbf{v}_n] \\
A\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad \text{for } i = 1, 2, \cdots, n.
\end{aligned}
$$

The above implies that the columns of P are formed by the eigenvectors of A and the diagonal entries of D are the associated eigenvalues.

R A matrix of order $n \times n$ is diagonalizable if and only if there are n linearly independent eigenvectors which is certainly the case when all the eigenvalues are distinct. However, even if some eigenvalues are repeated, there could be n linearly independent eigenvectors and the matrix could be diagonalizable.

■ **Example 1.44** Which of the following matrices is are diagonalizable?

$$
\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.
$$

■

R If a square matrix A has the eigen-decomposition $A = PDP^{-1}$, then its powers, polynomials, power series and some other functions of A could be calculated easily by using the fact that

$$
A^k = PD^kP^{-1}, \quad \text{where } D^k = \text{diag}(\lambda_1^k,\ \lambda_2^k,\ \cdots,\ \lambda_n^k).
$$

■ **Example 1.45** Find A^9 when $A = \begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix}$

The eigenvalues of this matrix could be calculated to be 2 and 3.

When $\lambda = 3$: $(A - 3I)\mathbf{v} = \mathbf{0} \implies \begin{pmatrix} -4 & 2 \\ -6 & 3 \end{pmatrix}\mathbf{v} = \mathbf{0}$, giving $\mathbf{v} = (1, 2)^T$.

When $\lambda = 2$: $(A - 2I)\mathbf{v} = \mathbf{0} \implies \begin{pmatrix} -3 & 2 \\ -6 & 4 \end{pmatrix}\mathbf{v} = \mathbf{0}$, giving $\mathbf{v} = (2, 3)^T$.

Hence the eigen-decomposition could be given by

$$A = \begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$$

where

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

By using the above diagonalization, we have

$$A^9 = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 3^9 & 0 \\ 0 & 2^9 \end{pmatrix} \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$$

■

■ **Example 1.46** Find 2^A when $A = \begin{pmatrix} -1 & 2 \\ -6 & 6 \end{pmatrix}$. Please note [2]

Continuing from the last example, we have

$$2^A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 2^3 & 0 \\ 0 & 2^2 \end{pmatrix} \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} -8 & 8 \\ -24 & 20 \end{pmatrix}$$

■

R Let $f(x)$ be a polynomial. If $(\lambda, \mathbf{v})$ is an eigenpair of A, then $(f(\lambda), \mathbf{v})$ is an eigenpair of $f(A)$.

1.8.2 Symmetric Matrices

A symmetric matrix is a square matrix that equals its transpose $A^T = A$. A skew-symmetric matrix is such that $A^T = -A$. Any square matrix could be written as the sum of a symmetric and a skew-symmetric matrix as

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T).$$

where the first term on the right is symmetric and the second one is skew-symmetric.

Theorem 1.8.2 — Spectral Decomposition. Every real symmetric matrix $A \in \mathbb{R}^{n\times n}$ is orthogonally diagonalizable. This means that there exist an orthonormal set of eigenvectors (real) of A that forms the columns of an orthogonal matrix Q such that

$$A = QDQ^T,$$

where the diagonal matrix D has real diagonal entries for the eigenvalues.

R **Spectral Decomposition and Projections** When A is a real and symmetric matrix and the eigen-decomposition is written in outer product form, we have

$$A = QDQ^T = \sum_{k=1}^{n} \lambda_k Q_{:k} Q_{:k}^T$$

where $P_k = Q_{:k}Q_{:k}^T$ is a rank-one projection matrix that orthogonally projects any vector $v \in \mathbb{R}^n$ onto the k-th eigenspace E_{λ_k} of A.

[2] I found this problem on a YouTube video.

Definition 1.8.4 — Eigen-space. The eigenspace of a square matrix $A \in \mathbb{C}^{n \times n}$ associated with an eigenvalue λ is given by

$$E_\lambda = \{\mathbf{v} \in \mathbb{C}^n : A\mathbf{v} = \lambda \mathbf{v}\}$$

R The dimension of E_λ is called the geometric multiplicity of the eigenvalue λ. A necessary and sufficient condition for diagonalizability is that the dimensions of the eigen-spaces are the same as the (algebraic) multiplicity of the associated eigenvalues

R From the spectral decomposition, we can conclude that every symmetric matrix is diagonalizable.

■ **Example 1.47** Are the following matrices diagonalizable? If so, find a diagonalization PDP^{-1}. If not, explain why a diagonalization does not exist.

(A)

$$\begin{pmatrix} 0 & 1 \\ -8 & 4 \end{pmatrix}$$

(B)

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

■

1.9 Vector and Matrix Norms

Norms provide the meaning of length and size to vectors and matrices. There are various types of norms that all should agree to some common notions.

Definition 1.9.1 — Vector Norm. Associated with a given vectors space V, a norm is a function that assigns a real number to its vectors, $\|\cdot\| : V \to \mathbb{R}$, and that satisfies the following properties

1. Nonnegativity: $\|\mathbf{v}\| \geq \mathbf{0}$.
2. Definiteness: $\|\mathbf{v}\| = \mathbf{0} \iff \mathbf{v} = \mathbf{0}$.
3. Homegeneity: For any real number α, $\|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\|$
4. Triangle inequality: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$

■ **Example 1.48** Let $\mathbf{v} = (v_1, v_2, \cdots, v_n)$. Prove that the following defines a norm

$$\|\mathbf{v}\| = |v_1| + |v_2| + \cdots + |v_n|$$

■

Another important idea is to have a sense of distance between pairs of vectors in a vectors space which is conveyed by a metric.

Definition 1.9.2 — Metric. A metric on a vector space associates a notion of distance to every pair of its elements. It is defined as $d : V \times V \to \mathbb{R}$ and satisfies the following properties for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$

1. Identity: $d(\mathbf{u}, \mathbf{v}) = 0 \Longleftrightarrow \mathbf{u} = \mathbf{v}$
2. Symmetry: $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$
3. Triangle inequality: $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

R One can verify that $d(\mathbf{u}, \mathbf{v}) \geq 0$.

R In the case of a normed vector space we can define a metric as

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{v} - \mathbf{u}\|.$$

We can use different norms, to get different metrics.

Examples of some commonly used norms on the space of n-dimensional vectors.

l_1 norm:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| = |x_1| + |x_2| + \cdots + |x_n|$$

l_2 norm:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \left(x_1^2 + x_2^2 + \cdots + x_n^2\right)^{1/2}$$

l_p norm:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} = \sqrt[p]{(|x_1|^p + |x_2|^p + \cdots + |x_n|^p)}$$

l_∞ norm:

$$\|x\|_\infty = \max_{i=1,\cdots,n} |x_i|$$

■ **Example 1.49** Find the l_1, l_2 and l_∞ norms of the vector

$$\mathbf{v} = (2, 3, -2, 5, -3, 0, -4, 2, -5, 2)^T$$

$$\|\mathbf{v}\|_1 = |2| + |3| + |-2| + |5| + |-3| + |0| + |-4| + |2| + |-5| + |2| = 28$$

$$\|\mathbf{v}\|_2 = \sqrt{(2)^2 + (3)^2 + (-2)^2 + (5)^2 + (-3)^2 + (0)^2 + (-4)^2 + (2)^2 + (-5)^2 + (2)^2} = \sqrt{100} = 10$$

$$\|\mathbf{v}\|_\infty = \max\{|2|, |3|, |-2|, |5|, |-3|, |0|, |-4|, |2|, |-5|, |2|\} = 5$$

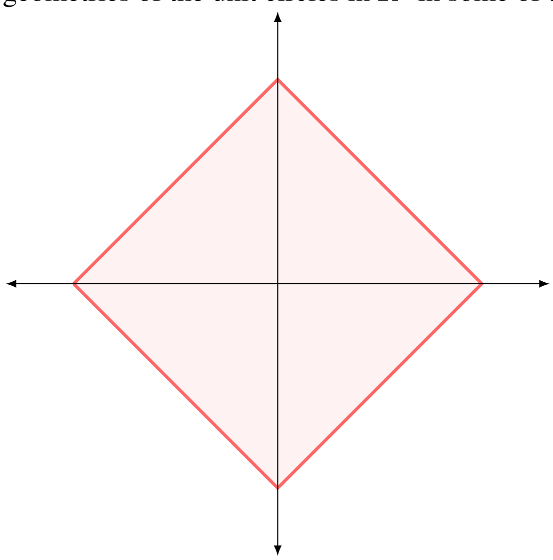
∎

∎ **Example 1.50** Determine if the expression defines a norm on $\mathbb{R}^n$:

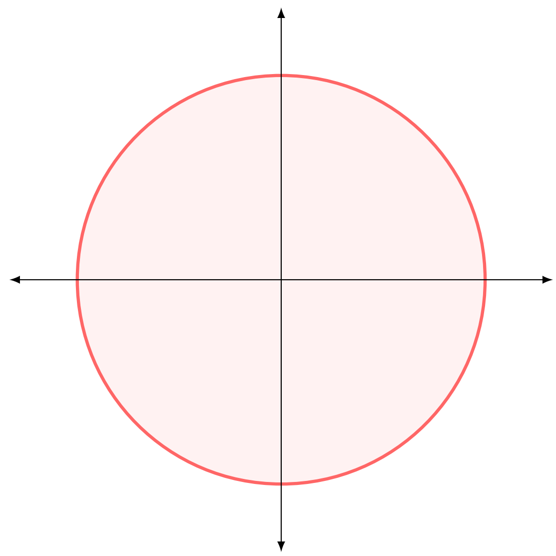
$$f(\mathbf{x}) = \sum_{i=1}^{n} |x_i|^3.$$

No. It violate the homogeneity property. ∎

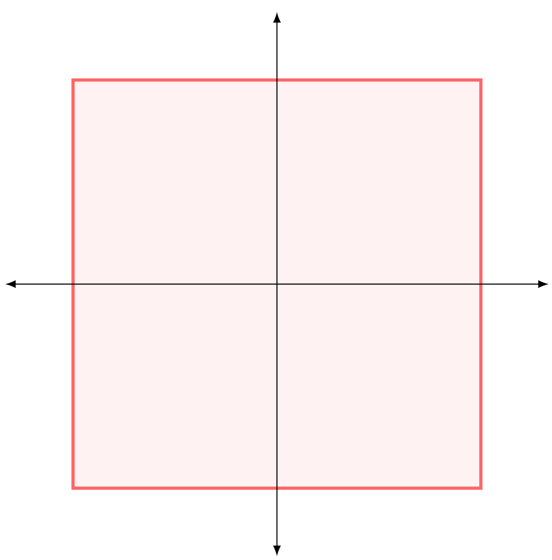
A Unit Circles is the set of all vectors that satisfy $\|\mathbf{v}\|_p \leq 1$ in some given norm. Following are the geometries of the unit circles in $\mathbb{R}^2$ in some of the common norms as discussed above.



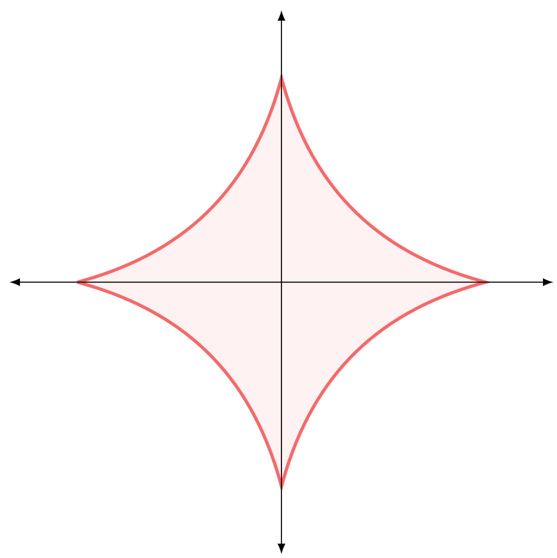
$\|\mathbf{v}\|_1 \leq 1$



$\|\mathbf{v}\|_2 \leq 1$



$\|\mathbf{v}\|_\infty \leq 1$



$\|\mathbf{v}\|_{1/2} \leq 1$

Some commonly used distances (metrics)
Euclidean Distance

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{v} - \mathbf{u}\|_2 = \sqrt{(v_1 - u_1)^2 + \cdots + (v_n - u_n)^2}.$$

Taxicab or Manhattan Distance

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{v} - \mathbf{u}\|_1 = |v_1 - u_1| + \cdots + |v_n - u_n|.$$

Chebyshev Distance

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{v} - \mathbf{u}\|_\infty = \max\{|v_1 - u_1|, \cdots, |v_n - u_n|\}.$$

1.9.1 Matrix Norms

We can treat a matrix as a vector and extend the definition of vector norms to matrices. For example the Euclidean norm could be extended to matrices to get what is called the **Frobenius norm** of a matrix defined as below.

Definition 1.9.3 — Frobenius Norms.

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j}^2 \right)^{1/2}.$$

■ **Example 1.51** Verify that $\|A\|_F = \sqrt{\operatorname{tr}(A^T A)}$.

$$\operatorname{tr}(A^T A) = \sum_{k=1}^{n} [A^T A]_{kk} = \sum_{k=1}^{n} [A^T]_{k:} A_{:k} = \sum_{k=1}^{n} \|A_{:k}\|_2^2 = \sum_{k=1}^{n} \sum_{i=1}^{n} A_{ik}^2$$

which is exactly the square of the Frobenius norms.

■

Definition 1.9.4 — Induced or Subordinate Matrix Norms. A matrix norm induced by a vector norm is defined as follows

$$\|A\| = \sup\{\|A\mathbf{x}\| : \mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| = 1\}$$

R An important consequence is that the subordinate norms also must satisfy the following properties.

- From definition of subordinate norms

$$\|A\mathbf{x}\| \le \|A\| \|\mathbf{x}\|$$

- Submultiplicative:

$$\|AB\| \le \|A\| \|B\|$$

- Subordinate matrix norms of identity matrix is always 1:

$$\|I\| = 1$$

R The subordinate matrix norm induced by vector norm $\|\cdot\|_\infty$, called **max-abs-row-sum** norm, is given by

$$\|A\|_\infty = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$$

R The subordinate matrix norm induced by vector norm $\|\cdot\|_1$, called **max-abs-column-sum** norm, is given by

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$$

R The subordinate matrix norm induced by vector norm $\|\cdot\|_2$, called **spectral norm**, is given by

$$\|A\|_2 = \sqrt{\rho(A^T A)}$$

where $\rho(B)$ denotes the spectral radius which is the magnitude of the largest eigenvalue of B.

■ **Example 1.52** Prove that the Frobenius norm is not an induced norm?
As $\|I_n\|_F = \sqrt{n} \neq 1$, the Frobenius norm is not an induced norm. ■

■ **Example 1.53** Consider the following matrix to find the norms below

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

- $\|A\|_\infty = \max\{|-3|+|-8|+|2|+|4|,\ |7|+|-4|+|-6|+|1|,\ |1|+|-7|+|3|+|9|,\ |-5|+|4|+|-3|+|7|\} = \max\{17,18,20,21\} = 21$

- $\|A\|_1 = \max\{|-3|+|7|+|1|+|-5|,\ |-8|+|-4|+|-7|+|4|,\ |2|+|-6|+|3|+|-3|,\ |4|+|1|+|9|+|7|\} = \max\{16,23,14,21\} = 23$

- $\|A\|_F = \sqrt{9+64+4+16+49+16+36+1+1+49+9+81+25+16+9+49} = \sqrt{434}$

■

R A vector norm associates a sense of size and distance to vectors in a vector space. There could be several different norms providing different interpretation of this length or distance. In a finite dimensional space all these norms are equivalent in some sense.

■ **Example 1.54** Find the Frobenius norm and spectral norms of the following matrix

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

■

■ **Example 1.55** Find the Euclidean and Manhattan distance($|d_1 - d_2|_1$) between the following two documents that share a common corpus of terms.

$$d_1 = (2,0,1,0,0,0,3,5,1,0,4)^T, \quad d_2 = (1,0,0,2,0,0,2,0,1,0,2)^T.$$

■

■ **Example 1.56** Prove that Euclidean norms are invariant under multiplication by orthogonal matrices. ■

■ **Example 1.57** Prove that Frobenius norms are invariant under multiplication by orthogonal matrices. ■

1.10 Orthogonal Matrices

Orthogonality is one of the most important ideas in linear algebra. In this section we will define orthogonal matrices and study their properties.

Definition 1.10.1 — Orthogonal Matrices. A square matrix, $Q \in \mathbb{R}^{n \times n}$, is called orthogonal if it satisfies the following

$$Q^T Q = I_n, \quad \text{and} \quad QQ^T = I_n$$

where I_n is the identity matrix of order n-by-n.

R If Q is orthogonal matrix, then $Q^{-1} = Q^T$

R Multiplication by an orthogonal matrix preserves length (l_2 norm) as

$$\|Q\mathbf{x}\|_2^2 = (Q\mathbf{x})^T Q\mathbf{x} = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T \mathbf{x} = \|\mathbf{x}\|_2^2.$$

R The dot product between two vectors is invariant under multiplication by an orthogonal matrix. This also preserves the angle between two vectors.

$$(Q\mathbf{u})^T Q\mathbf{v} = \mathbf{u}^T Q^T Q\mathbf{v} = \mathbf{u}^T \mathbf{v}$$

The cosine of the angle between two vectors $Q\mathbf{u}$ and $Q\mathbf{v}$ is

$$\cos\theta = \frac{(Q\mathbf{u})^T Q\mathbf{v}}{\|Q\mathbf{u}\|\|Q\mathbf{v}\|} = \frac{\mathbf{u}^T Q^T Q\mathbf{v}}{\|Q\mathbf{u}\|\|Q\mathbf{v}\|} = \frac{\mathbf{u}^T \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

the same as between $\mathbf{u}$ and $\mathbf{v}$.

R As a linear transformation, an orthogonal matrix either rotates or reflects a vector, or does a combination of both.

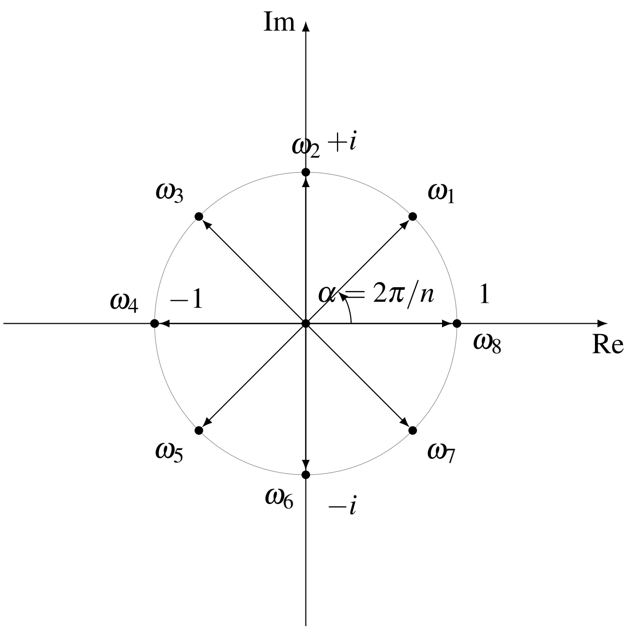
R Determinant of an orthogonal matrix is either +1 or -1 as

$$\det(Q^T Q) = \det(Q^T)\det(Q) = \det(Q)\det(Q) = 1 \Longrightarrow \det(Q) = \pm 1.$$

R Eigenvalues of an orthogonal matrix are of unit magnitude as

$$Q\mathbf{v} = \lambda\mathbf{v} \Longrightarrow \|Q\mathbf{v}\| = \|\lambda\mathbf{v}\| \Longrightarrow \|\mathbf{v}\| = |\lambda|\|\mathbf{v}\| \Longrightarrow |\lambda| = 1$$

These eigenvalues are the n-th complex roots of unity.



1.10.1 Orthogonal Projection

Definition 1.10.2 — Projection Matrix or Projector. A square matrix $P \in \mathbb{R}^{n\times n}$ serves as a projection matrix that projects elements in $\mathbb{R}^n$ to a subspace V of $\mathbb{R}^n$ (the image of P) if $P^2 = P$. In addition, if P is symmetric, $P^T = P$, then the projection is also orthogonal that means $x - Px \perp Px$ for all $x \in \mathbb{R}^n$.

■ **Example 1.58** Find a 3×3 matrix that...

(A) Projects vectors in $\mathbb{R}^3$ onto the z-axis.

$$P = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(B) Projects vectors in $\mathbb{R}^3$ onto the xy-plane.

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

■

■ **Example 1.59** Prove that the matrix $P = QQ^T$, where Q is a rectangular matrix made of orthonormal columns, is an orthogonal projection matrix. What is the image of P?

Check that

$$P^2 = QQ^T QQ^T = QQ^T = P.$$

Note that $Q^T Q = I$ but $QQ^T \neq I$. In addition,

$$P^T = (QQ^T)^T = QQ^T = P$$

.

So, we have an orthogonal projection matrix. As

$$P\mathbf{x} = QQ^T\mathbf{x} \in \mathrm{col}(Q),$$

the image of this projection is the column space of Q. ■

■ **Example 1.60** Find the projection p of the vector $\mathbf{u} = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$ onto the line passing through $\mathbf{v} = \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$. Then verify that the vector $\mathbf{e} = \mathbf{u} - \mathbf{p}$ is orthogonal to $\mathbf{v}$. ■

1.11 Positive Definite Matrices

Definition 1.11.1 — Positive Definite Matrices. A real symmetric matrix $A \in \mathbb{R}^{n\times n}$ is called positive definite if for every $\mathbf{x} \neq \mathbf{0}$ (of compatible dimension) we have $\mathbf{x}^T A\mathbf{x} > 0$.

■ **Example 1.61** ANSWER: Prove that the matrix $A = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$ is positive definite.

Let $\mathbf{x} = (x_1, x_2)^T$, then

$$\mathbf{x}^T A\mathbf{x} = (3x_1, 2x_2)(x_1, x_2)^T = 3x_1^2 + 2x_2^2 > 0 \quad \text{for } x_1, x_2 \neq 0.$$

■

(R) Above example could be generalized to show that every diagonal matrix with positive entries is symmetric positive definite.

(R) The term **positive semi-definite** is used for symmetric matrices satisfying $\mathbf{x}^T A\mathbf{x} \geq 0$ for all non-zero $\mathbf{x} \neq$.

(R) The term **negative semi-definite** and **negative definite** are used for real symmetric matrices satisfying $\mathbf{x}^T A\mathbf{x} \leq 0$, and $\mathbf{x}^T A\mathbf{x} < 0$ respectively.

(R) A symmetric matrix is called **indefinite** if it is neither positive semi-definite nor negative semi-definite.

■ **Example 1.62** Prove that the diagonal elements of every real symmetric positive definite matrix is positive and

$$a_{ij} < \sqrt{a_{ii}a_{jj}}, \quad \text{for } i \neq j.$$

■

■ **Example 1.63** The covariance and correlation matrices of a dataset are positive semidefinite. What conditions should be satisfied for it to be positive definite?

The sample covariance matrix of a data matrix X where columns are the features is given by

$$S = \frac{1}{m-1}(X - \bar{X})^T(X - \bar{X}).$$

This implies that

$$\mathbf{y}^T S\mathbf{y} = \frac{1}{m-1}\|(X - \bar{X})\mathbf{y}\|_2^2 \geq 0.$$

Hence, S is positive semidefinite. For positive definite-ness we must have the columns of $(X - \bar{X})$ linearly independent. ■

Theorem 1.11.1 — Diagonally dominant and Positive Definite. Every symmetric diagonally dominant (strictly) matrix with positive diagonal elements is positive definite.

■ **Example 1.64** The real symmetric matrix :

$$M = \begin{bmatrix} 5 & -2 & 0 \\ -2 & 5 & -2 \\ 0 & -2 & 5 \end{bmatrix}$$

is positive definite since for any non-zero column vector $\mathbf{x} = (a, b, c)^T$, we have:

$$\begin{aligned} \mathbf{x}^{\mathsf{T}} M \mathbf{x} = &= \begin{bmatrix} (5a-2b) & (-2a+5b-2c) & (-2b+5c) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\ &= (5a-2b)a + (-2a+5b-2c)b + (-2b+5c)c \\ &= 5a^2 - 2ba - 2ab + 5b^2 - 2cb - 2bc + 5c^2 \\ &= 5a^2 - 4ab + 5b^2 - 4bc + 5c^2 \\ &= 3a^2 + 2a^2 - 4ab + 2b^2 + b^2 + 2b^2 - 4bc + 2c^2 + 3c^2 \\ &= 3a^2 + 2(a-b)^2 + 2b^2 + 2(b-c)^2 + 3c^2 \end{aligned}$$

This result is a sum of squares, and therefore non-negative; and is zero only if $a = b = c = 0$, that is, when $\mathbf{x}$ is the zero vector. ■

R Above example serves to show that it will be considerably difficult to prove that a matrix is positive definite by following the definition as proving such inequalities could be challenging. There are other more practical methods to prove positive definiteness.

Theorem 1.11.2 — Test for Positive Definite Matrices. A real symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if and only if each of its leading principal submatrices have positive determinant, i.e.,

$$\det(A[1:k, 1:k]) > 0, \quad \text{for } k = 1, 2, \cdots, n.$$

■ **Example 1.65** Determine if the matrix $A = \begin{pmatrix} 3 & -2 & 0 \\ -2 & 4 & -3 \\ 0 & -3 & 5 \end{pmatrix}$ is positive definite.

Find the determinants of the three leading principal submatrices:

$$(3)\,, \quad \begin{pmatrix} 3 & -2 \\ -2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 3 & -2 & 0 \\ -2 & 4 & -3 \\ 0 & -3 & 5 \end{pmatrix}.$$

This matrix is real symmetric and the requisite determinants (also called leading principal minors) are 3, 8, and $3(20-9) - (-2)(-10-0) = 13$; all positive. Therefore the matrix is symmetric positive definite. ■

■ **Example 1.66** For what values of α is the following matrix a symmetric positive definite one?

$$M = \begin{bmatrix} 3 & -2 & \alpha \\ -2 & 3 & -2 \\ \alpha & -2 & 4 \end{bmatrix}$$

■

Fact — When A is symmetric positive definite, Gaussian elimination without row-interchanges could be performed for solving $Ax = b$ where all pivot elements are positive.

Theorem 1.11.3 — Cholesky Factorization. Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. There exist a unique real lower triangular matrix L and unique unit lower triangular matrix $\tilde{L}$ such that

$$A = LL^T = \tilde{L}D\tilde{L}^T$$

where D is a diagonal matrix such that $L = \tilde{L}D^{1/2}$.

Cholesky–Banachiewicz algorithm for Cholesky Decomposition

This algorithm finds the factor matrix L row-by-row starting from the top-left entry.

$$L_{i,i} = \sqrt{A_{i,i} - \sum_{j=1}^{i-1} L_{i,j}^2},$$

$$L_{i,j} = \frac{1}{L_{j,j}}\left(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k}\right) \quad \text{for } i > j.$$

```
#Cholesky-Banachiewicz algorithm for Cholesky Decomposition
from math import sqrt
import numpy as np

def cholesky(A):
    m,n = A.shape
    if m!=n:
        raise ValueError("Please provide a square matrix")

    # Start L with zero entries
    L = np.zeros((n,n),dtype=float)

    for i in range(n):
        for j in range(i+1):
            ele_sum = np.sum(L[i,0:j] * L[j,0:j])
            if i == j:
                L[i,j] = sqrt(A[i,i] - ele_sum)
            else:
                L[i,j] =  (A[i,j] - ele_sum)/L[j,j]
    return L
```

```
# This example is from WikiPedia
A = np.array([[4, 12, -16],[12, 37, -43],[-16, -43, 98]], dtype=float)
L = cholesky(A)

print ("A:", A)
print ("L:", L)
print("Check for zero norm of A-LL^T:",np.linalg.norm(A-L@L.T))

OUTPUT

A: [[  4.  12. -16.]
 [ 12.  37. -43.]
 [-16. -43.  98.]]
array([[  4.,  12., -16.],
       [ 12.,  37., -43.],
       [-16., -43.,  98.]])
L: [[ 2.  0.  0.]
 [ 6.  1.  0.]
 [-8.  5.  3.]]
Check for zero norm of A-LL^T: 0.0
```

Applications of Cholesky Factorization

Cholesky decomposition could be performed in half the operation counts compared to LU decomposition. A linear system $A\mathbf{x} = \mathbf{b}$ could then be solved by first solving $Lz = b$ by forward substitution and then solving $L^T x = z$ by backward substitution.

■ **Example 1.67** Find the Cholesky factorization of the following $\begin{pmatrix} 4 & 12 & 16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{pmatrix}$ ■

Theorem 1.11.4 — Positive Definiteness and Eigenvalue Signs. All the eigenvalues of a real symmetric positive definite matrix are positive. Conversely, if all the eigenvalues of a real symmetric matrix are positive, then it is a positive definite matrix.

■ **Example 1.68** Prove that the inverse of a symmetric positive definite matrix A is also symmetric positive definite.

Symmetry: $(A^{-1})^T = (A^T)^{-1} = A^{-1}$. The eigenvalues of A^{-1} are reciprocal of A and thus positive. Hence A^{-1} is symmetric positive definite too. ■

R **Determination of positive definiteness**: It is better to use Cholesky factorization than calculating the eigenvalues and checking their signs for determining whether a given real matrix is symmetric positive definite or not. This is so considering the number of operations and the stability of the process.

R A real symmetric matrix A is positive semidefinite if and only if there exists a real matrix M such that $A = M^T M$. Invertibility of M is required for positive definiteness.

R A real symmetric matrix A is positive semidefinite if and only if there exists a positive semidefinite matrix B such that $A = BB$. This unique matrix B is called the **square root** of A.

Problems

1. Find the Cholesky decomposition (in the form $A = LDL^T$) of the following matrix

$$A = \begin{pmatrix} 0.5 & 2 & -1 & 1 \\ 2 & 9 & -1 & 6 \\ -1 & -1 & 14 & 13 \\ 1 & 6 & 13 & 35 \end{pmatrix}$$

2. Prove that the inverse of a symmetric positive definite matrix is positive definite.
3. Under what condition could the matrix $A^T A$ be positive definite?
4. Determine whether the following matrix is positive definite.
5. Prove the Theorem 1.11.4.

1.12 Finite Precision Mathematics

How computers store and process numbers is of critical importance while writing numerical algorithms. Every computer has only finite number of binary places assigned to specific numbers depending on what kind of data type is used. For example, integers are stored in **int** and **unsigned int**; decimal numbers are stored in float32 and float64 formats that take 32 bits and 64 bits respectively. Here we will focus on float64 as the most common data type for numerical computations.

1.12.1 Computer Representation of Numbers

Please check the interactive media and the videos for 64-bit representation of decimals.

1.12.2 Floating Point Operations

Please check iPython Notebook 2.