Applied Linear Algebra in Data Analysis: Course Notes

Sivakumar Balasubramanian
CMC Vellore

Update on June 18, 2024

Contents

CONTENTS

List of Figures

Part I

Linear Algebra

Chapter 1

Vectors

1.1 n-Vectors

A collection of an ordered list of n numbers is called an n-vector. We will use bold lower case alphabets to represent such vectors, and we will represent these as a column of numbers, which is referred to as a *column vector*. We will look at *row vectors* at a later stage. Consider the following example:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

The elements of the n-vector $x_1, x_2, \ldots, x_n$ are called the *components* of the vector $\mathbf{x}$; x_i is the i^{th} component of the vector $\mathbf{x}$. If these components are all real numbers, the set of all such n-vectors is the set $\mathbb{R}^n$.

Where do we come across such n-vectors? In many places, such as in physics, engineering, economics, medicine, etc. Any application where we deal with multiple pieces of information that can be represented as a list of numbers can be represented as an n-vector. When we deal with systems with multiple inputs, multiple output, or multiple states, we can represent these as n-vectors. We talk about the state of a system in a later chapter.

1.2 Visualizing n-vectors

The n-vectors can be visualized as points in n-dimensional space. For example, A 1-vector or just single real number or a *scalar* can be thought of as a point on the real line. The 1-vector $x = 2.45$ is shown in Figure 1.1 is the red point. But we will find it useful to visualize a 1-vector as an arrow starting at the origin and ending at the point on the real line. The arrow is shown in blue in Figure 1.1.

The elements of $\mathbb{R}^2$ are points on the plane, and we can visualize them as points in the plane. The 2-vectors $\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ are shown in Figure 1.2a. A similar visualization is shown for $\mathbb{R}^3$

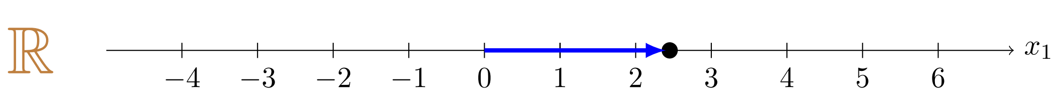
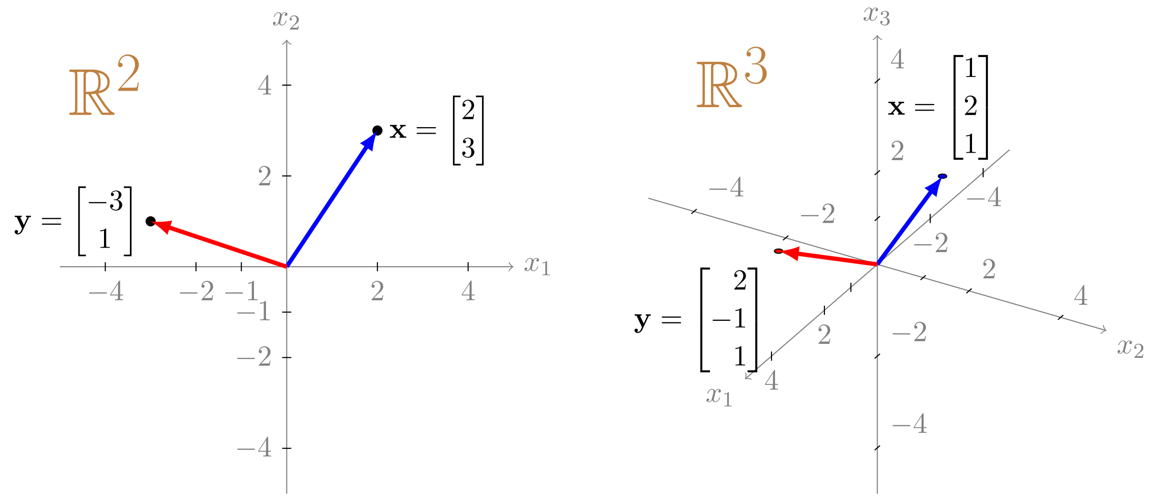


Figure 1.1: The real line $\mathbb{R}$ contains the 1-vectors.



(a) The $\mathbb{R}^2$ plane contains the 2-vectors.

(b) $\mathbb{R}^3$ contains the 3-vectors.

Figure 1.2: The $\mathbb{R}^2$ and $\mathbb{R}^3$ sets.

(Figure 1.2b), and for $\mathbb{R}^4$ and beyond you simply pretend that you can visualize things in your head like your instructor does.

1.3 Some Commonly Used n-vectors

We will now define a some commonly used n-vectors that we will use in the course.

- **Zero vector:** The n-vector whose components are all zeros is called the *zero vector*. $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

- **One vector:** The n-vector whose components are all ones is called the *one vector*. $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

- **Unit vectors:** The n-vectors whose components are all zeros except for one component which is 1. These are called the *standard basis vectors* and are denoted by $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. The n-vector $\mathbf{e}_i$ has all components as zeros except for the i^{th} component which is 1. For example, the unit vectors in $\mathbb{R}^2$ are:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

1.4 Operations on n-vectors

There are many operations we can perform on n-vectors, but we will only focus on two operations for this:

- **Scalar multiplication:** Given a scalar $c \in \mathbb{R}$ and an n-vector $\mathbf{x}$. The scalar multiplication operation produces another n-vector $c\mathbf{x}$ whose components are $cx_1, cx_2, \ldots, cx_n$.

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \longrightarrow 2\mathbf{x} = \begin{bmatrix} 2\,(1) \\ 2\,(2) \end{bmatrix} = \begin{bmatrix} 2 \\ 8.2 \end{bmatrix}$$

The geometric interpretation scalar multiplication is shown in Figure 1.3. Scalar multiplication stretches or shrinks the vector without rotating it. When the scalar is positive the direction of the scaled vector is the same as the original vector, and when the scalar is negative the direction is opposite. When the scalar is zero, the scaled vector is the zero vector $\mathbf{0}$.

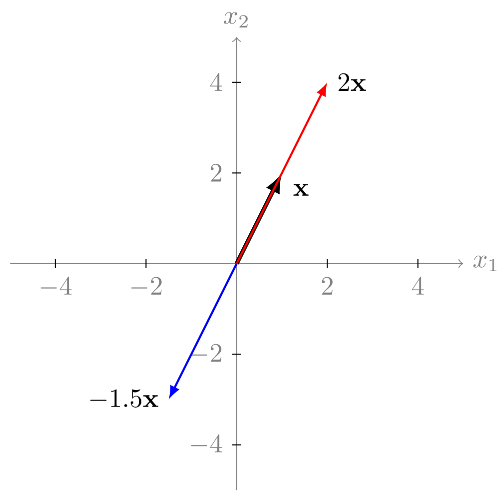


Figure 1.3: Scalar multiplication of a vector.

- **Vector Addition:** Given two n-vectors $\mathbf{x}$ and $\mathbf{y}$, the vector addition operation, represented by $\mathbf{x}+\mathbf{y}$, producs another n-vector whose components are $x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n$.

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \longrightarrow \mathbf{x}+\mathbf{y} = \begin{bmatrix} 1+2 \\ 3+1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

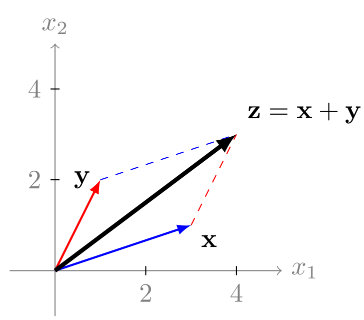


Figure 1.4: Vector addition.

The geometric interpretation the vector addition operation is shown in Figure 1.4. Geometrically, the vector addition operation follows the parallelogram law of addition, where the resulting vector $\mathbf{x}+\mathbf{y}$ is a diagonal of the parallelogram formed by the two vectors $\mathbf{x}$ and $\mathbf{y}$. Another way to think about this, is that you first move along $\mathbf{x}$ to its end point, and starting from there then move along $\mathbf{y}$ to its end point or vice versa.

You can add more than two vectors to obtain a new vector, like below:

$$\mathbf{w} = \mathbf{x} + \mathbf{y} + \mathbf{z}$$

Geometrically, we can first apply the parallelogram law to $\mathbf{x}$ and $\mathbf{y}$, and then apply the parallelogram law to $\mathbf{x} + \mathbf{y}$ and $\mathbf{z}$ to get $\mathbf{w}$.

1.5 Vector spaces

Vector spaces are *sets* with some special properites. More specifically, a vector space is a set V of elements called *vectors* that are closed under two operations called *addition* and *scalar multiplication*. This simply means that if you perfom these operations using elements from the set V, the result is also an element of the set V. A vector space must satisfy the following properties:

- **Closure under addition:** For any two vectors $\mathbf{x}, \mathbf{y} \in V$, the sum $\mathbf{x} + \mathbf{y} \in V$.
- **Closure under scalar multiplication:** For any scalar $c \in \mathbb{R}$ and any vector $\mathbf{x} \in V$, the product $c\mathbf{x} \in V$.
- **Additive identity:** There exists a vector $\mathbf{0} \in V$ such that for any vector $\mathbf{x} \in V$, $\mathbf{x} + \mathbf{0} = \mathbf{x}$.
- **Additive inverse:** For any vector $\mathbf{x} \in V$, there exists a vector $-\mathbf{x} \in V$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
- **Commutativity of addition:** For any two vectors $\mathbf{x}, \mathbf{y} \in V$, $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
- **Associativity of addition:** For any three vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$, $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
- **Distributive property:** For any scalar $c \in \mathbb{R}$ and any two vectors $\mathbf{x}, \mathbf{y} \in V$, $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$.
- **Distributive property:** For any two scalars $c, d \in \mathbb{R}$ and any vector $\mathbf{x} \in V$, $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$.
- **Associativity of scalar multiplication:** For any two scalars $c, d \in \mathbb{R}$ and any vector $\mathbf{x} \in V$, $(cd)\mathbf{x} = c(d\mathbf{x})$.
- **Multiplicative identity:** For any vector $\mathbf{x} \in V$, $1\mathbf{x} = \mathbf{x}$.

These properties are satisfied by the set $\mathbb{R}^n$ of n-vectors, and hence $\mathbb{R}^n$ is a vector space. Geometrically, the concept of a vector space corresponds to flat spaces that contain the origin. This will become more clear when we talk about subspaces. Notice that definition of the vector space given above does not make any specific mention of n-vectors. The definition is general and can be applied to any set of elements that satisfy the properties listed above. The following are some examples of vector spaces with the addition and scalar multiplication operations defined on them.

Example 1.1. Set of $m \times n$ matrices. The set M of all $m \times n$ matrices of real numbers is a vector space.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix}, \; a_{ij} \in \mathbb{R}$$

We define scalar multiplication and addition of matrices as follows:

$$c\mathbf{A} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{11} & ca_{12} & \cdots & ca_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{11} & ca_{12} & \cdots & ca_{1n} \end{bmatrix}, \ c \in \mathbb{R}$$

$$\mathbf{A}+\mathbf{B} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \end{bmatrix}, \ \mathbf{A},\mathbf{B} \in M$$

Since each element of $c\mathbf{A}$ and $\mathbf{A}+\mathbf{B}$ is a real number, M is a vector space.

Example 1.2. Set of polynomials of order $\leq n$. Now we look at strange example of a vector space. The set P_n of all polynomials of degree at most n with real coefficients, defined over an interval $[a,b]$.

$$p\left(x\right) = \sum_{k=0}^{n-1} a_k x^k, \ x \in [a,b], \ a_k \in \mathbb{R}$$

The set P_n contains all polynomials of the form shown above. We define scalar multiplication and addition of polynomials as follows:

$$cp\left(x\right) = c\sum_{k=0}^{n-1} a_k x^k = \sum_{k=0}^{n-1} ca_k x^k, \ p\left(x\right) \in P$$

$$p\left(x\right)+q\left(x\right) = \sum_{k=0}^{n-1} a_k x^k + \sum_{k=0}^{n-1} b_k x^k = \sum_{k=0}^{n-1} \left(a_k+b_k\right) x^k, \ p\left(x\right), q\left(x\right) \in P_n$$

The set P_n is a vector space because the sum and product of any two polynomials from P_n is also a polynomial of degree at most n with real coefficients.

Example 1.3. Set of continuous functions. The set $C\left[0,1\right]$ of all continuous functions $f\left(x\right)$ over the time interval $x \in [0,1]$ is a vector space. We define scalar multiplication and addition of functions as follows:

$$cf\left(x\right) = cf\left(x\right), \ f\left(x\right) \in C\left(0,1\right)$$

$$f\left(x\right)+g\left(x\right) = f\left(x\right)+g\left(x\right), \ f\left(x\right), g\left(x\right) \in C\left(0,1\right)$$

The set $C\left(0,1\right)$ is a vector space because the sum and product of any two continuous functions from $C\left(0,1\right)$ is also a continuous function.

1.6 Subspaces – “Little” Vector Spaces

These are little subspaces in the sense that they are subsets of a larger vector space that are themselves vector spaces. More formally, a subspace U of a vector space V is a subset of V that is itself a vector space. The subspace U of the vector space V must satisfy the following properties:

- **Closure under addition:** For any two vectors $\mathbf{x},\mathbf{y} \in U$, the sum $\mathbf{x}+\mathbf{y} \in U$.

- **Closure under scalar multiplication:** For any scalar $c \in \mathbb{R}$ and any vector $\mathbf{x} \in U$, the product $c\mathbf{x} \in U$.

One immediate consequence of the above definition is that hte zero element of the vector space V must be present in every subspace of V. If the zero element is not in a subset, then it cannot be a subspace. Geometrically subspace are flat structures (or surfaces or manifolds) in $\mathbb{R}^n$ (or the parent vector space) that contain the origin, and extends infinitely. Let's look at some examples of subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$, which are easier to visualize.

Example 1.4. A straight line through the origin. We know that $\mathbb{R}^2$ is a vector space. Now consider the set of all points in $\mathbb{R}^2$ that lie on a straight line passing through the origin, defined as follows:

$$S = \left\{ \mathbf{x} \,:\, x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2,\, x_1 = m \cdot x_2,\, m \in \mathbb{R} \right\}$$

How do we verfiy this is a subspace of $\mathbb{R}^2$? The definition above shows that any x in S comes from $\mathbb{R}^2$, which means its a subset of $\mathbb{R}^2$. Figure 1.5a shows the set S for $m = 2$. How do we verify if S is a subspace of $\mathbb{R}^2$? We need to now verify that S satisfies the properties of a vector space.

1. First, let's check if S contains the zero vector. If it does not contain the zero vector, then it cannot be a subspace. The elements from S are of the form $\begin{bmatrix} x \\ mx \end{bmatrix}$, thus if we choose $x = 0$, then we get $\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in S$. So, S contains the zero vector. This means that S can be a subspace space of $\mathbb{R}^2$.

2. Let's verify vector scaling. Scaling the element $\begin{bmatrix} x \\ mx \end{bmatrix} \in S$ by a scalar c we get,

$$c \begin{bmatrix} x \\ mx \end{bmatrix} = \begin{bmatrix} cx \\ cmx \end{bmatrix} = \begin{bmatrix} cx \\ m\,(cx) \end{bmatrix} = \begin{bmatrix} y \\ my \end{bmatrix}, \quad \text{where } y = cx \in \mathbb{R}$$

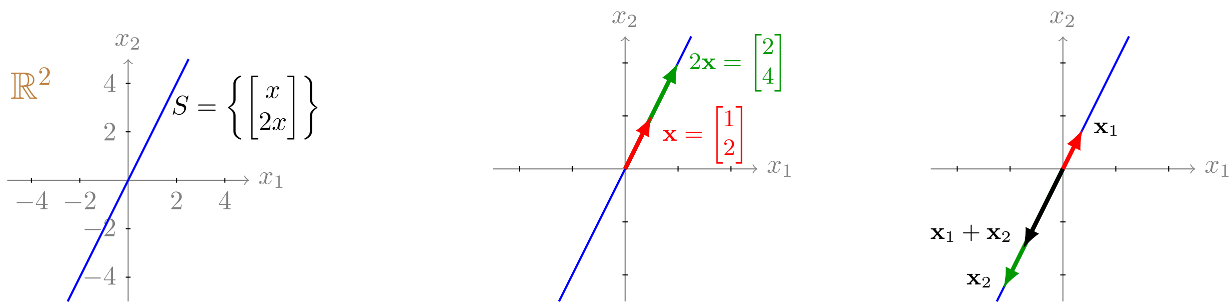
This means that $c \begin{bmatrix} x \\ mx \end{bmatrix}$ belongs to S, this the set S is closed under scalar multiplication. This still means that S can be a subspace of $\mathbb{R}^2$.

3. Let's verify vector addition. Adding two elements $\begin{bmatrix} x_1 \\ mx_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ mx_2 \end{bmatrix} \in S$ we get,

$$\begin{bmatrix} x_1 \\ mx_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ mx_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ mx_1 + mx_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ my_1 \end{bmatrix}, \quad \text{where } y_1 = x_1 + x_2 \in \mathbb{R}$$

This means that $\begin{bmatrix} y_1 \\ my_1 \end{bmatrix}$ belongs to S, this the set S is closed under vector addition. This means that S is a subspace of $\mathbb{R}^2$.

Since, the subset S is closed under both vector addition and scalar multiplication, it is a subspace of $\mathbb{R}^2$.



(a) A subspace of $\mathbb{R}^2$. **(b)** Vector scaling. **(c)** Vector addition.

Figure 1.5: Example of a subspace of $\mathbb{R}$. (a) Shows the set of all points in $\mathbb{R}^2$ corresponding to the subset $S = \left\{ \begin{bmatrix} x \\ 2x \end{bmatrix} \right\} \subset \mathbb{R}^2$. (b) Shows that the set S is closed under scalar multiplication. Take any vecotr from the line, and scale it and it remains on that blue line. (c) Shows that S is closed under vector addition. If we take any two vectors from the blue line and add them, the resulting vector remains in the blue line.

Example 1.5. A straight line not through the origin. Consider the set of all points in $\mathbb{R}^2$ of the following form:

$$S = \left\{ \mathbf{x} \,:\, x = \begin{bmatrix} x \\ mx + c \end{bmatrix} \in \mathbb{R}^2,\, m, c \in \mathbb{R} \right\}$$

This is shown in the Figure 1.6a.

How do we verfiy this is a subspace of $\mathbb{R}^2$? The definition above shows that any x in S comes from $\mathbb{R}^2$, which means its a subset of $\mathbb{R}^2$. Figure 1.6a shows the set S for $m = -\frac{1}{2}$ and $c = 1$. How do we verify if S is a subspace of $\mathbb{R}^2$? We need to now verify that S satisfies the properties of a vector space.

1. First, let's check if S contains the zero vector. If it does not contain the zero vector, then it cannot be a subspace. The elements from S are of the form $\begin{bmatrix} x \\ mx + c \end{bmatrix}$, thus if we choose $x = 0$, then we get $\begin{bmatrix} 0 \\ c \end{bmatrix} \in S$. So, S does not contain the zero vector, which implies that S is not a subspace space of $\mathbb{R}^2$. We need not check the other two conditions; but we will test them just to see which of these two fails.

2. Scaling the element $\begin{bmatrix} x \\ mx + c \end{bmatrix} \in S$ by a scalar d we get,

$$d \begin{bmatrix} x \\ mx + c \end{bmatrix} = \begin{bmatrix} dx \\ dmx + dc \end{bmatrix} = \begin{bmatrix} dx \\ m\,(dx) + dc \end{bmatrix} \neq \begin{bmatrix} y \\ my + c \end{bmatrix}, \quad \text{where } y = dx \in \mathbb{R}$$

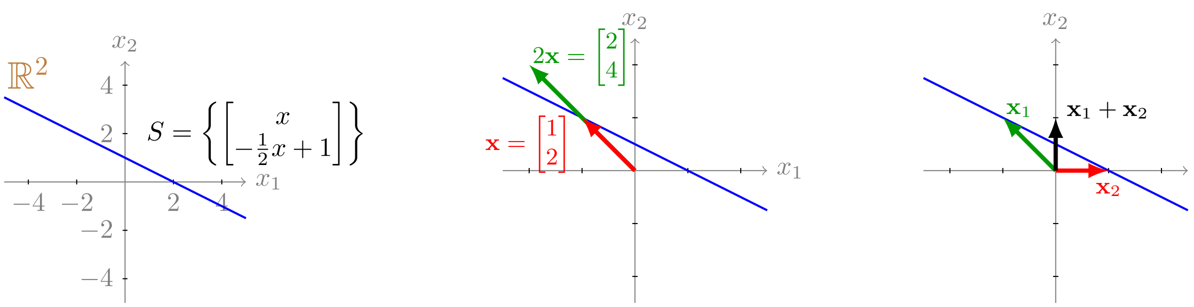
   This means that $d \begin{bmatrix} x \\ mx + c \end{bmatrix} \notin S$. Thus, the set S is closed under scalar multiplication. Another confirmation that it is not a subspace. This can be seen in Figure 1.6b, which shows that when we choose an elment from $\mathbf{x}$ (red arrow) from S (blue line), the scaled version of this vector leaves the set S, i.e., the tip of the green arrow does not stay on the blue line.

3. Let's verify vector addition. Adding two elements $\begin{bmatrix} x_1 \\ mx_1 + c \end{bmatrix}, \begin{bmatrix} x_2 \\ mx_2 + c \end{bmatrix} \in S$ we get,

$$\begin{bmatrix} x_1 \\ mx_1 + c \end{bmatrix} + \begin{bmatrix} x_2 \\ mx_2 + c \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ mx_1 + mx_2 + 2c \end{bmatrix} \neq \begin{bmatrix} y_1 \\ my_1 + c \end{bmatrix}, \quad \text{where } y_1 = x_1 + x_2 \in \mathbb{R}$$

This means that $\begin{bmatrix} x_1 \\ mx_1 + c \end{bmatrix} + \begin{bmatrix} x_2 \\ mx_2 + c \end{bmatrix} \notin S$. Thus the set S is closed under vector addition. We see this geometrically in Figure 1.6c, where the sum of two vectors in S does not stay in the set S. Even though the tips of the green and red arrow are on the blue line, the tip of the black arrow is not on the blue line.

Since, the subset S is closed under both vector addition and scalar multiplication, it is a subspace of $\mathbb{R}^2$.



(a) Not a subspace of $\mathbb{R}^2$. **(b)** Vector scaling. **(c)** Vector addition.

Figure 1.6: Example of a subspace of $\mathbb{R}$. (a) Shows the set of all points in $\mathbb{R}^2$ corresponding to the subset $S = \left\{ \begin{bmatrix} x \\ 2x \end{bmatrix} \right\} \subset \mathbb{R}^2$. (b) Shows that the set S is closed under scalar multiplication. Take any vecotr from the line, and scale it and it remains on that blue line. (c) Shows that S is closed under vector addition. If we take any two vectors from the blue line and add them, the resulting vector remains in the blue line.

Example 1.6. A parabola through the origin. Consider the set of all points in $\mathbb{R}^2$ of the following form:

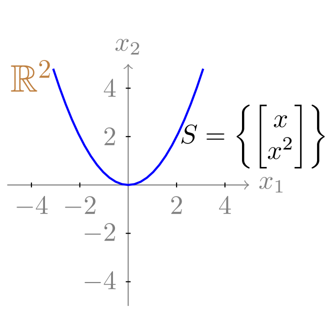
$$S = \left\{ \mathbf{x} \,:\, x = \begin{bmatrix} x \\ \frac{1}{2}x^2 \end{bmatrix} \in \mathbb{R}^2,\, m, c \in \mathbb{R} \right\}$$

This is not a subspace of $\mathbb{R}^2$. This is geometrically depicted in Figure 1.7a, Figure 1.7b and Figure 1.7c. You are encouraged to verify this algebraically by checking the properties of a vector space.

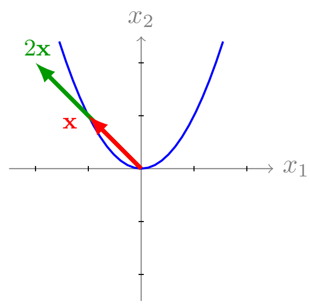
1.7 Linear combinations and others

Linear combination is an *algebraic operation* performed on a set of vectors. We can combine the two fundamental operations on vectors into a single operation called the *linear combination* of a set of vectors. Given a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m \in \mathbb{R}^n$ and scalars $c_1, c_2, \ldots, c_n \in \mathbb{R}$, the linear combination of the vectors is given by:

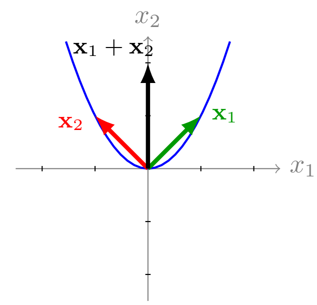
$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_m \in \mathbb{R}^n \tag{1.1}$$



(a) Not a subspace of $\mathbb{R}^2$.



(b) Vector scaling.



(c) Vector addition.

Figure 1.7: Example of a subspace of $\mathbb{R}$. (a) Shows the set of all points in $\mathbb{R}^2$ corresponding to the subset $S = \left\{ \begin{bmatrix} x \\ 2x \end{bmatrix} \right\} \subset \mathbb{R}^2$. (b) Shows that the set S is closed under scalar multiplication. Take any vecotr from the line, and scale it and it remains on that blue line. (c) Shows that S is closed under vector addition. If we take any two vectors from the blue line and add them, the resulting vector remains in the blue line.

Notice that the linear cominations of single vector $\mathbf{v}_1$ are simply different scaled versions of the vector $c_1\mathbf{v}_1$. Linear cominations are the bread-and-butter of linear algebra and we will encounter them again and again. An informal way to think of a linear combination of a set of vectors as process of mixing the set of vectors together with the corresponding scalar c_i determining the amount of a vector in the mixture. There are other types of combinations of vectors, which we will not discuss further in this book.

- **Affine combination:** $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_m, \quad \sum_{i=1}^{m} c_i = 1$
- **Convex combinations:** $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_m, \quad c_i \geq 0, \ \sum_{i=1}^{m} c_i = 1$
- **Conic combinations:** $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_m, \quad c_i \geq 0$

1.8 Linear independece of a set of vectors

Linear independece is a *property* of a set of vectors; a set of vector is either linearly independent or linearly dependent. The concept of linear independence is easy to understand the but the algebraic condition for independence can seem a bit unintuitive. A set of vectors is said to be linearly independent if no vector in the set can be expressed as a linear combination of the other vectors in the set.

More formally, a set of vectors $V = \{\mathbf{v}_i\}_{i=1}^{m}$ is said to be linearly independent if and only if the only way to produce the zero vector $\mathbf{0}$ through the linear combination of the set V is by setting all the scalars to zero, i.e.,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_m = \mathbf{0} \quad \text{if and only if} \quad c_1 = c_2 = \cdots = c_m = 0 \tag{1.2}$$

To understand this better, let's assume that the set V is linear dependent and let's assume that the vector $\mathbf{v}_m$ can be represented as the linear combination of the vectors $\mathbf{v}_1, \mathbf{v} - 2, \cdots \mathbf{v}_{m-1}$. This means that there exist a set of scalar $\alpha_i, \ 1 \leq i \leq m-1$, such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_{m-1}\mathbf{v}_{m-1} = \mathbf{v}_m$$

Multiplying both sides by a scalar $c_m \neq 0$ we get,

$$\begin{aligned} c_m\alpha_1\mathbf{v}_1 + c_m\alpha_2\mathbf{v}_2 + \cdots + c_m\alpha_{m-1}\mathbf{v}_{m-1} &= c_m\mathbf{v}_m \\ \implies c_m\alpha_1\mathbf{v}_1 + c_m\alpha_2\mathbf{v}_2 + \cdots + c_m\alpha_{m-1}\mathbf{v}_{m-1} - c_m\mathbf{v}_m &= \mathbf{0} \end{aligned}$$

This implies that there exist a set of scalar $c_i = c_m\alpha_i,\ 1 \leq i \leq m-1$, and c_m such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m = \mathbf{0}$, where not all c_i are zero. So when a set is linearly dependent, then there are scalars c_i, not all zero, such that the linear combination of the vectors from V with these scalars produces the zero vector.

Now, let's assume that the set V is linearly independent, that is no vector in the set V can be expressed as a linear combination of other vectors in that set. And let's assume that there are scalars c_i, not all zero, such that the linear combination of the vectors from V with these scalars produces the zero vector, i.e.,

$$\begin{aligned} c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m &= \mathbf{0} \\ \implies \frac{c_1}{c_m}\mathbf{v}_1 + \frac{c_2}{c_m}\mathbf{v}_2 + \cdots + \frac{c_{m-1}}{c_m}\mathbf{v}_{m-1} &= \mathbf{v}_m,\ c_m \neq 0 \end{aligned}$$

But this a contradiction because we have just expressed $\mathbf{v}_m$ is expressed as a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_{m-1}$.

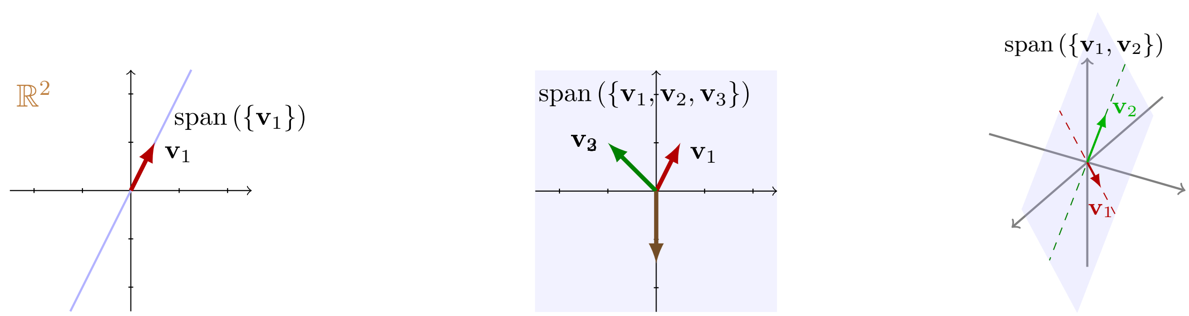
1.9 Span of a set of vectors

So, linear combinations of a set of vectors $V = \{\mathbf{v}_i\}_{i=1}^m$ ($\mathbf{v}_i \in \mathbb{R}^n$) is a way of generating new vectors not in that set. All we need to do is choose a random set of real numbers $\{c_i\}_{i=1}^m$, and "mix" the vectors $\mathbf{v}_i$ from the set using these as weights. Clearly there are infinite number of vectors we could generate through this process, and we can put them all together in a set. And this set has a name – the *span* of the set V. The span of a set of vectors $V = \{\mathbf{v}_i\}_{i=1}^m$ is denoted by span (V) and is defined as:

$$\text{span}\,(V) = \{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_m\ :\ c_i \in \mathbb{R}\} \subseteq \mathbb{R}^n \tag{1.3}$$

Its clear that this will be a subset of $\mathbb{R}^n$, but it turns out it is also a subspace of $\mathbb{R}^n$. Why? Can you verify this fact algebraically? (*Hint*: Just follow the steps in Examples 1.4-1.6).

Geometrically, this means that the span (V) will be a flat surface in $\mathbb{R}^n$. Which means that the linear combination operation generates vectors that lie on a flat surface spanned by the vectors emmployed in the linear combination.



(a) Not a subspace of $\mathbb{R}^2$. **(b)** Vector scaling. **(c)** Vector addition.

Figure 1.8: Span of a set of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

1.10 How big is a vector?

The size of a vector is an extension of the idea of the magnitude of a real number. The magnitude of a real number $a \in \mathbb{R}$ tells us how big the number is irrespective of its sign:

$$|a| = \begin{cases} a, & a \geq 0 \\ -a, & a < 0 \end{cases} \tag{1.4}$$

The "magnitude" or size of an element of a vector space (such as $\mathbb{R}^n$) is called the *norm* of the vector. The norm of a vector is a function that maps a vector to a non-negative real number, and satisfies the following properties:

- **Non-negativity:** For any vector $\mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{x}\| \geq 0$.
- **Definiteness:** The norm of a vector is zero if and only if the vector is the zero vector, i.e., $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
- **Homogeneity:** Scaling a vector by a scalar c, scales the norm of the vector by $|c|$. For any vector $\mathbf{x} \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$, $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$.
- **Triangle inequality:** For any vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

According to this definition, the magnitude of real numbers (Eq. 1.4) is a norm of the vector $\mathbb{R}$. The most common norm of a vector is the *Euclidean norm* or the *2-norm* of a vector. The Euclidean norm of a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is defined as:

$$\|\mathbf{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{1.5}$$

We are well-versed with this as the length of a vector in $\mathbb{R}^2$ and $\mathbb{R}^3$. The subscript 2 in Eq. 1.5 is used to indicate that it is the 2-norm, which is a special case of a general class of norms in $\mathbb{R}^n$ – the *p-norm*. The *p-norm* is defined as the following:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \ p \in \mathbb{Z}, \ p \geq 1 \tag{1.6}$$

Apart from the *2-norm*, the *1-norm* and the ∞*-norm* are also commonly used norms, which are defined as the following:

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|\mathbf{x}\|_\infty = \max_i |x_i| \tag{1.7}$$

The *1-norm* is the sum of the absolute value of the elements of the vector, and the ∞*-norm* is the maximum of the absolute value of the elements of the vector. The *1-norm* is also called the *Manhattan norm* or the *Taxicab norm* because it measures the distance between two points in a city if you can only travel along the grid of streets.

Problem 1.1. Why does the ∞-norm measure have this weird looking definition compared to the other p-norms?

Solution. Consider the vector $\mathbf{x} \in \mathbb{R}^n$, and $x_{max} = \max_{0 \leq i \leq n} |x_i|$; let's also assume that the j^{th} element of $\mathbf{x}$ has the maximum absolute value, i.e. $x_{max} = |x_j|$. The p-norm is defined as the following:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} = x_{max} \left(1 + \sum_{\substack{1 \leq i \leq n \\ i \neq j}} \left| \frac{x_i}{x_j} \right|^p \right)^{1/p} = x_{max} \left(N \right)^{1/p}$$

where, N is a real number between 1 and n, because $|\frac{x_i}{x_j}| \leq 1$ (why?). Now, if we increase the value of p to infinity, then the term $\lim_{p \rightarrow infty} (N)^{1/p} = 1$. Thus, we have $\|\mathbf{x}\|_\infty = \lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = x_{max} = \max_i |x_i|$. $\square$

1.10.1 Geometry of the p-norms

In the case of real numbers, the set of all numbers with a magnitude of 1 is the set $\{-1, 1\}$. We can plot these points in the real line as below.

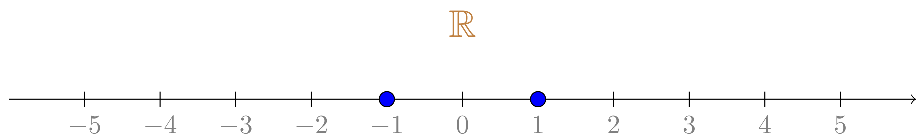
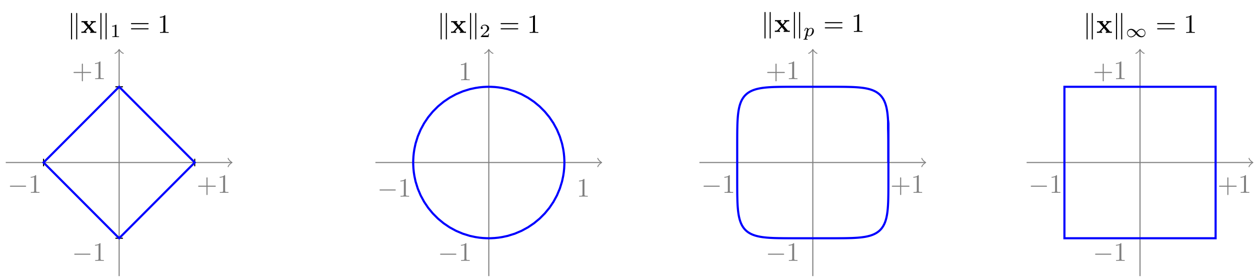


Figure 1.9: The set of all real numbers with magnitude 1. This set contains two numbers $\{-1, 1\}$.

In $\mathbb{R}^2$, the set of all vectors from $\mathbb{R}^2$ with a *2-norm* of 1 is the unit circle. The following figure shows the set of all points in $\mathbb{R}^2$ with unit 1, 2, p, and ∞ norm.



(a) 1-norm **(b)** 2-norm **(c)** p-norm **(d)** ∞-norm

Figure 1.10: Locus of all points with unit 1, 2, p, and ∞ norms in $\mathbb{R}^2$.

Problem 1.2. Can you explain why the different norms have these shapes?

Problem 1.3. Can you write a Python program to generate the above plots for different values of $p = 1, 2, 3, 10$ and ∞?

Problem 1.4. Can describe what these 1, 2, p and ∞ norms will look like in $\mathbb{R}^3$?

1.11 How similar are two vectors?

The idea of how similar two or more vectors are is an important topic in data analysis, in particular in classification problems in machine learning. Vectors that are "similar" somehow belong to the same "category" or "class", while vectors that are "dissimilar" belong to different categories or classes. There are various ways to measure the similarity between two vectors. We will look at two methods in this section where similarity is mesasured by computing the distance between two vectors or by computing the angle between two vectors.

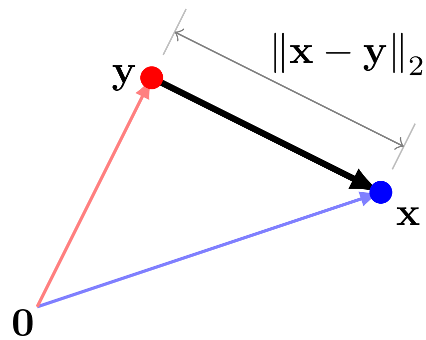
1.11.1 Distance between two vectors

The logic here is that similar vectors correspond to points that are close together, while disimilar vectors are father away. We can make use of the norm to compute the distance between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

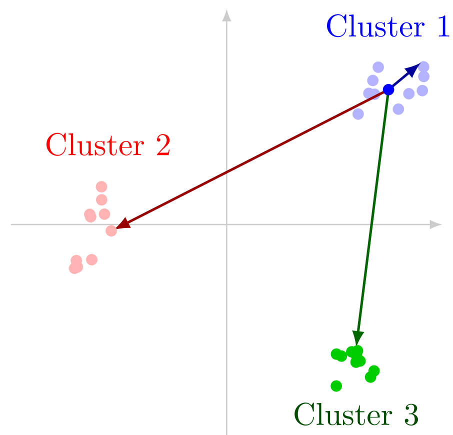
Since the the difference between these two vectors $\mathbf{x}-\mathbf{y}$ is also another vector, we can compute the distance between vectors $\mathbf{x}$ and $\mathbf{y}$ as the norm of the vector $\mathbf{x}-\mathbf{y}$ (Figure 1.11a).

$$\text{Distance between } \mathbf{x} \text{ and } \mathbf{y} = d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|_p$$

We could use any of the p norms to compute this or come-up with a new norm depending on the application we are dealing with. Take look at the clusters of points shown in Figure 1.11b, we would agree that the different colored points each form a cluster, since the points of the same color are closer to each other than points from another color.



(a) Distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^2$. This figures depicts the 2-norm, but any p-norm or valid norm function could be used to quantify the distance between two vectors or points.



(b) Distance between two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^2$. This figures depicts the 2-norm, but any p-norm or valid norm function could be used to quantify the distance between two vectors or points. could be used to quantify the distance between

Figure 1.11

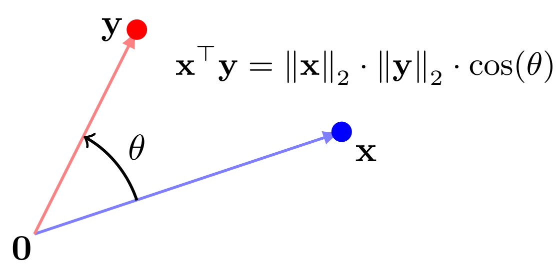
1.11.2 Angle between two vectors

This approach is based on the idea that the direction of the vector repesenting a point contains information about the point. Thus, vectors that point in a similar direction could be considered similar. But how do we measure the angle between two vectors in $\mathbb{R}^n$? This is where the concept of the *standard inner product* (or the dot product from vectors from high school math and physics). The standard inner product of two vectors $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ is defined as:

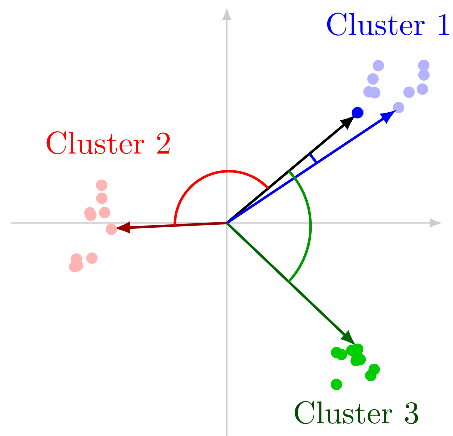
$$\mathbf{x}^\top\mathbf{y} = \sum_{i=1}^{n} x_i y_i$$

The superscript '$\top$' represents the transpose operation. We will not worry about what it means in the next chapter. The standard inner product takes in two vectors $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ and returns a scalar value $\mathbb{R}$; it can be both positive and negative. We compute it by simply taking the two vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, and multiply the two of them element-wise $x_i y_i$, $1 \le i \le n$ and add the n products together $\sum_{i=1}^{n} x_i y_i$ to obtain the inner product.

The standard inner product of two vectors $\mathbf{x}$ and $\mathbf{y}$ is related to the cosine of the angle θ between the



(a) Angle between two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^2$. The standard inner product provides a measure of the cosine of the angle between the two vectoes.



(b) The relative angle between the points of the same colors is smaller than that of points of different colors.

Figure 1.12

two vectors, and the 2-norm of the two vectors.

$$\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i = \|\mathbf{x}\|_2 \cdot \|\mathbf{y}\|_2 \cdot \cos(\theta)$$

The angle θ between the two vectors $\mathbf{x}$ and $\mathbf{y}$ can be computed as:

$$\theta = \cos^{-1}\left(\frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}\right), \quad \|\mathbf{x}\|_2 \neq 0, \|\mathbf{y}\|_2 \neq 0$$

If the 2-norms of the $\mathbf{x}$ and $\mathbf{y}$, then $\mathbf{x}^\top \mathbf{y}$ is simply the cosine of the angle between the vectors.

1.12 Standard and other inner products

$\mathbf{x}^\top \mathbf{y}$ is the standard inner product, which of course means there are non-standard inner products. But before we look at generalizing the concept of an inner product, let's look at some properties of the standard inner product.

- **Connection to the 2-norm**. The standard inner product of a vector $\mathbf{x}$ with itself is the square of the 2-norm of the vector, i.e., $\mathbf{x}^\top \mathbf{x} = \|\mathbf{x}\|_2^2 = \sum_{i=1}^n x_i^2$.
- **Cauchy-Bunyakovski-Schwartz Inequality**:

$$\mathbf{x}^\top \mathbf{y} \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \tag{1.8}$$

The concept of an inner product is a general one. An inner product $\langle \cdot, \cdot \rangle$ is a function that maps two vectors from $\mathbb{R}^n$ to a scalar value, and satisfies the following properties:

- **Positive definiteness:** $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, and $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
- **Symmetry:** For any vectors $\mathbf{x}, \mathbf{y}$, $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.
- **Linearity:** For any vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$, and any scalars $\alpha, \beta \in \mathbb{R}$, $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{y} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$.

We will come across other inner products in due course, but we will stick to the standard inner product for most problem in $\mathbb{R}^n$ in this course.

Problem 1.5. Consider the vector space $\mathbb{R}^n$. Is the the following a valid inner product of $\mathbb{R}^n$?

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} w_1 x_i y_i,\ \ w_i \in \mathbb{R},\ \ w_i > 0$$

Part II

Optimization

Part III

Probability and Statistics

Part IV

Least Squares