Linear Algebra

Riddhiman

July 2023

1 Preface

Self study Linear Algebra notes. Source is *Vector Calculus, Linear Algebra, and Differential Geometry* by John and Barbara Hubbard. These notes will start from Section 1.3. Notes for Sections 1.1 and 1.2 will be added later.

2 Matrix Multiplication as Linear Transformation

Definition 2.1 (Linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$): A *linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ is a mapping such that for all scalars a and for all $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^n$,

$$T(\vec{\mathbf{v}} + \vec{\mathbf{w}}) = T(\vec{\mathbf{v}}) + T(\vec{\mathbf{w}}) \text{ and } T(a\vec{\mathbf{v}}) = aT(\vec{\mathbf{v}}) \tag{2.1}$$

This above was the definition of a linear transformation. The following theorem will relate matrices to linear transformation.

Theorem 2.1 (Matrices and linear transformation)

1. Any $m \times n$ matrix A defines a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by matrix multiplication:
$$T(\vec{\mathbf{v}}) = A\vec{\mathbf{v}}. \tag{2.2}$$

2. Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by multiplication by the $m \times n$ matrix $[T]$:
$$T(\vec{\mathbf{v}}) = [T]\vec{\mathbf{v}}, \tag{2.3}$$
where the i^{th} column of $[T]$ is $T(\vec{\mathbf{e}}_i)$.

Proof to part 2 of Theorem 2.1. Given that our domain is $\mathbb{R}^n$ we can write any vector $\vec{\mathbf{v}} \in \mathbb{R}^n$ as

$$\vec{\mathbf{v}} = u_1\vec{\mathbf{e}}_1 + u_2\vec{\mathbf{e}}_2 + \ldots + u_n\vec{\mathbf{e}}_n = \sum_{i=1}^{n} u_i\vec{\mathbf{e}}_i$$

Thus,

$$\begin{aligned} T(\vec{\mathbf{v}}) &= T\left(\sum_{i=1}^{n} u_i\vec{\mathbf{e}}_i\right) \\ &= \sum_{i=1}^{n} u_i T(\vec{\mathbf{e}}_i) \end{aligned}$$

Also note that since every i^{th} column of $[T]$ is $T(\vec{\mathbf{e}}_i)$. When $[T]$ is multiplied by $\vec{\mathbf{v}}$, the i^{th} of the resultant column vector can be denoted as $u_1 T(\vec{\mathbf{e}}_1) + \ldots + u_n T(\vec{\mathbf{e}}_n)$ which is just $T(\vec{\mathbf{v}})$. $\square$

2.1 Finding the matrix of a linear transformation

By Theorem 2.1 is it clear that to find the matrix say $[T]$ associated with the linear transformation T we have to find what T does to the basis vectors $\vec{\mathbf{e}}_i$.

For example, consider a straight line passing through the origin of $\mathbb{R}^2$. Also consider two arbitrary vectors $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$. We wish to find the resultant vectors when $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ are reflected about the line. Using the figure below it is easy to see that the act of reflecting vectors about a straight line passing through the origin is a linear transformation. Let this transformation be called T.

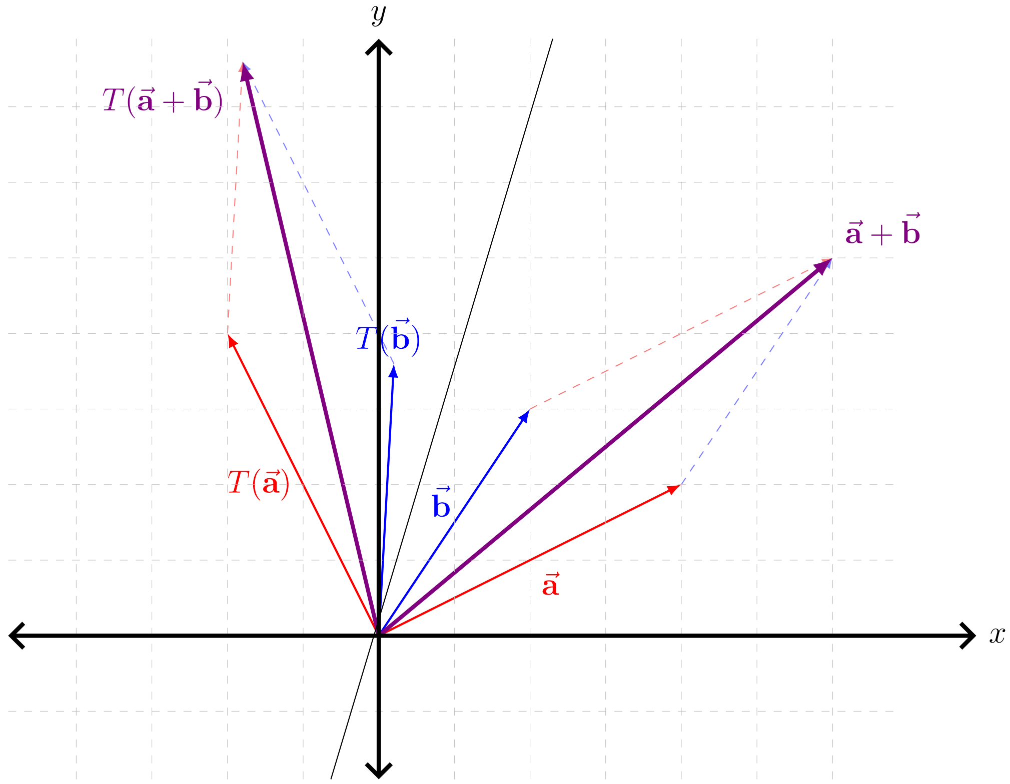


Figure 2.1: Reflection is linear

Recall that the first column of $[T]$ is $T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)$, and thus the second column is $T\left(\begin{bmatrix}0\\1\end{bmatrix}\right)$. Let the line make an angle θ with the x-axis. Then the matrix comes out to be

$$[T] = \begin{bmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{bmatrix}$$

Now just multiply this matrix by any vector to get the reflected vector. This is how matrices behave – they encode linear transformations.

Another example of a linear transformation is the projection matrix. This matrix projects given vectors onto a certain line passing through the origin. If said line is making an angle θ with the x-axis then the matrix is

$$\begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}$$

In both the cases above, how the where the basis vectors land after applying the transformation was found by basic trigonometry.

2.2 Geometry interpretation of linear transformation

A linear transformation may be applied to entire subsets of $\mathbb{R}^n$ instead of just individual vectors. Here are some transformations, their matrix representations, and how each point in $\mathbb{R}^n$ behaves when transformed.

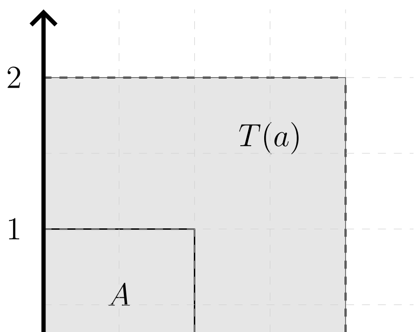
Example 2.1 (Identity transformation): The identity transformation : $\mathbb{R}^n \to \mathbb{R}^n$ is represented by I_n. Applying this transformation to a subset of $\mathbb{R}^n$ leaves it unchanged.

Example 2.2 (Scaling transformation): This transformation T enlarges everything by a factor of a and is given by $\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$.

Example 2.3 (Stretching transformation): These transformations are of the form $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. It will stretch the unit square into a rectangle.

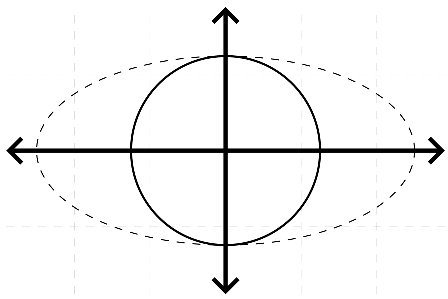
Example 2.4 (Rotation transformation): This is a transformation R which rotates the subset by an angle θ in the counterclockwise around the origin. This is given by

$$[R] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



(a) The scaling transformation given by $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ turns the unit square into the square with side length 2

.



(b) The linear transformation given by $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ stretches the unit circle into an ellipse

Theorem 2.2 (Composition corresponds to matrix multiplication)

Suppose $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^l$ are linear transformations given by the matrices $[S]$ and $[T]$ respectively. Then the $[T \circ S]$ is linear and

$$[T \circ S] = [T][S] \tag{2.4}$$

Proof. First we will prove that $T \circ S$ is indeed a linear transformation. Hence, consider the following

$$\begin{aligned}(T \circ S)(a\vec{\mathbf{v}} + b\vec{\mathbf{w}}) &= T(S(a\vec{\mathbf{v}} + b\vec{\mathbf{w}})) = T(aS(\vec{\mathbf{v}}) + bS(\vec{\mathbf{w}})) \\ &= T(aS(\vec{\mathbf{v}})) + T(bS(\vec{\mathbf{w}})) \\ &= aT(S(\vec{\mathbf{v}})) + T(S(\vec{\mathbf{w}})) \\ &= a(T \circ S)(\vec{\mathbf{v}}) + b(T \circ S)(\vec{\mathbf{w}}).\end{aligned}$$

This shows that $T \circ S$ is a linear transformation has a corresponding matrix $[T \circ S]$. Now we have to prove Equation 2.4. For this we will use the following two facts:

1. $A\vec{\mathbf{e}}_i$ is the i^{th} column of A.
2. The i^{th} column of AB is $A\vec{\mathbf{b}}_i$ where $\vec{\mathbf{b}}_i$ is the i^{th} column of B.

Now,

$$[T \circ S]\vec{\mathbf{e}}_i = (T \circ S)(\vec{\mathbf{e}}_i) = T(S(\vec{\mathbf{e}}_i)) = T([S]\vec{\mathbf{e}}_i) = [T]([S]\vec{\mathbf{e}}_i). \tag{2.5}$$

Notice that by fact (1), $[T \circ S]\vec{\mathbf{e}}_i$ is the i^{th} column of $[T \circ S]$ and that $[S]\vec{\mathbf{e}}_i$ is the i^{th} column of $[S]$, and thus by fact (2), $[T]([S]\vec{\mathbf{e}}_i)$ becomes the i^{th} column of $[T][S]$. Since the columns of $[T \circ S]$ and $[T][S]$ are equivalent, they are the same matrix, which proves Equation 2.4. □

Given below is an example desmostrating composition of linear transformations.

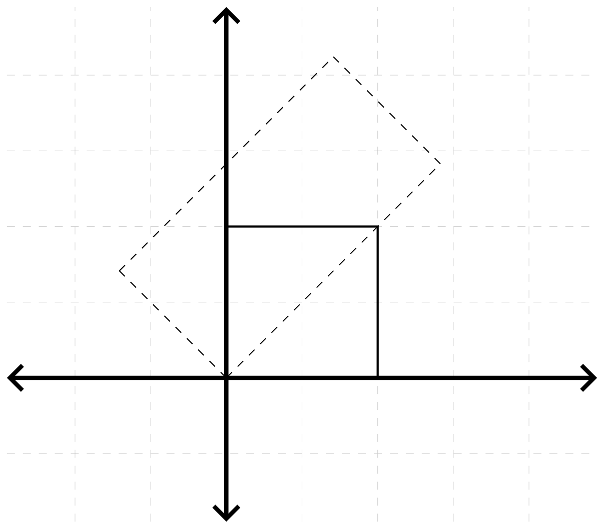


Figure 2.3: Result of applying $\begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ to the unit square.

2.3 Invertibility of matrices and linear transformations

Proposition 2.3

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible if and only if the $m \times n$ matrix $[T]$ is invertible. If it is invertible, then

$$[T^{-1}] = [T]^{-1} \tag{2.6}$$

Before we prove this lets first break it down a bit. This proposition has essentially two statements which we have to prove,

- A linear transformation T is invertible if $[T]$ is invertible.
- $[T]$ is invertible if the linear transformation T is invertible and $[T^{-1}] = [T]^{-1}$.

Proof. We prove the statements one by one.

Assume that $[T]$ is invertible. To prove that $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible we just have to prove that T is both one-one and onto. Consider a vector $\vec{\mathbf{y}} \in \mathbb{R}^m$ and a transformation $S : \mathbb{R}^m \to \mathbb{R}^n$ such that $[S] = [T]^{-1}$. Thus we can write

$$\vec{\mathbf{y}} = ([T][S])\vec{\mathbf{y}} = T(S(\vec{\mathbf{y}})),$$

This is enough to prove that T is onto, since $\vec{\mathbf{y}}$ can be any vector in $\mathbb{R}^m$ and it always has a solution $S(\vec{\mathbf{y}})$ in $\mathbb{R}^n$. Next we prove that T is one-one. Assume two vectors $\vec{\mathbf{x}}_1, \vec{\mathbf{x}}_1 \in \mathbb{R}^n$ and that $T(\vec{\mathbf{x_1}}) = T(\vec{\mathbf{x}}_2)$. Thus,

$$\vec{\mathbf{x}}_1 = ([S][T])\vec{\mathbf{x}}_1 = S(T(\vec{\mathbf{x}}_1)) = S(T(\vec{\mathbf{x}}_2)) = ([S][T])\vec{\mathbf{x}}_2 = \vec{\mathbf{x}}_2.$$

Thus T is one-one. Hence, if $[T]$ is invertible then T is also invertible. But aren't we assuming that S is injective here?

Now we will prove the second statement. Assume that $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible. We will first prove that $T^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ is also linear (and hence has a matrix $[T^{-1}]$ associated with it). This is proved by the following computation. Let $\vec{\mathbf{y}}_1, \vec{\mathbf{y}}_2 \in \mathbb{R}^m$, then

$$\begin{aligned} T(aT^{-1}(\vec{\mathbf{y}}_1) + bT^{-1}(\vec{\mathbf{y}}_2)) &= aT(T^{-1}(\vec{\mathbf{y}}_1)) + bT(T^{-1}(\vec{\mathbf{y}}_2)) = a\vec{\mathbf{y}}_1 + b\vec{\mathbf{y}}_2 \\ &= T \circ T^{-1}(a\vec{\mathbf{y}}_1 + b\vec{\mathbf{y}}_2). \end{aligned}$$

Since, T is one-one, we can conclude that $aT^{-1}(\vec{\mathbf{y}}_1) + bT^{-1}(\vec{\mathbf{y}}_2) = T^{-1}(a\vec{\mathbf{y}}_1 + b\vec{\mathbf{y}}_2)$ proving that T^{-1} is linear. Now we are left to prove that $[T^{-1}] = [T]^{-1}$. □

3 The geometry of $\mathbb{R}^n$

3.1 The dot product

Just everyone's favourite dot product. Also known as the *standard inner product.*

Definition 3.1 (Dot product): The *dot product* $\vec{\mathbf{x}} \cdot \vec{\mathbf{y}}$ of two vectors $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathbb{R}^n$ is

$$\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} := x_1 y_1 + x_2 y_2 + \ldots + x_n y_n. \tag{3.1}$$

The dot product is commutative, distributive and can be denoted as matrix multiplication:

$$\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = \vec{\mathbf{x}}^\top \vec{\mathbf{y}} = \vec{\mathbf{y}}^\top \vec{\mathbf{x}}.$$

Similarly, matrix multiplication can be viewed in form of dot products.

Now we define length in terms of dot product.

Definition 3.2 (Length of a vector): The *length* $|\vec{\mathbf{x}}|$ of a vector $\vec{\mathbf{x}} \in \mathbb{R}^n$ is

$$|\vec{\mathbf{x}}| := \sqrt{\vec{\mathbf{x}} \cdot \vec{\mathbf{x}}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \tag{3.2}$$

3.2 Geometric interpretation

Proposition 3.1 (Geometric interpretation of dot product).

Let $\vec{\mathbf{x}}$, $\vec{\mathbf{y}}$ be vectors in $\mathbb{R}^2$ or in $\mathbb{R}^3$, and let α be the angle between them. Then

$$\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = |\vec{\mathbf{x}}||\vec{\mathbf{y}}| \cos\alpha. \tag{3.3}$$

Proof. To prove this we will use the cosine law which states that for a triangle with sides a, b, c and angle γ between sides a and b, the following relation is true

$$c^2 = a^2 + b^2 - 2ab\cos\gamma.$$

In the figure below consider the triangle made by vectors $\vec{\mathbf{x}}$, $\vec{\mathbf{y}}$ and $\vec{\mathbf{x}} - \vec{\mathbf{y}}$ and angle α between $\vec{\mathbf{x}}$ and $\vec{\mathbf{y}}$. Then by cosine law we have,

$$|\vec{\mathbf{x}} - \vec{\mathbf{y}}|^2 = |\vec{\mathbf{x}}|^2 + |\vec{\mathbf{y}}|^2 - 2|\vec{\mathbf{x}}||\vec{\mathbf{y}}|\cos\alpha \tag{3.4}$$

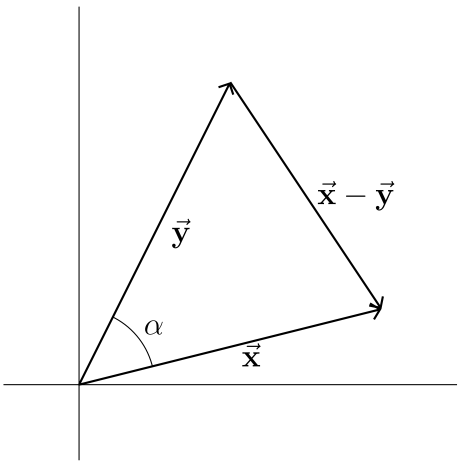


Figure 3.1: Figure for cosine law

But note that $|\vec{\mathbf{x}} - \vec{\mathbf{y}}|^2 = (\vec{\mathbf{x}} - \vec{\mathbf{y}}) \cdot (\vec{\mathbf{x}} - \vec{\mathbf{y}})$. Thus we can rewrite Equation 3.4 as,

$$\begin{aligned} |\vec{\mathbf{x}} - \vec{\mathbf{y}}|^2 &= (\vec{\mathbf{x}} - \vec{\mathbf{y}}) \cdot (\vec{\mathbf{x}} - \vec{\mathbf{y}}) = ((\vec{\mathbf{x}} - \vec{\mathbf{y}}) \cdot \vec{\mathbf{x}}) - ((\vec{\mathbf{x}} - \vec{\mathbf{y}}) \cdot \vec{\mathbf{y}}) && (3.5) \\ &= (\vec{\mathbf{x}} \cdot \vec{\mathbf{x}}) - (\vec{\mathbf{y}} \cdot \vec{\mathbf{x}}) - (\vec{\mathbf{x}} \cdot \vec{\mathbf{y}}) + (\vec{\mathbf{y}} \cdot \vec{\mathbf{y}}) && (3.6) \\ &= |\vec{\mathbf{x}}|^2 + |\vec{\mathbf{y}}|^2 - 2\vec{\mathbf{x}} \cdot \vec{\mathbf{y}}. && (3.7) \end{aligned}$$

Comparing the two equation we get $\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} = |\vec{\mathbf{x}}||\vec{\mathbf{y}}|\cos\alpha$.

$\square$

Corollary 3.1 (The dot product in terms of projections): If $\vec{\mathbf{a}}$ and $\vec{\mathbf{b}}$ are two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, then $\vec{\mathbf{a}} \cdot \vec{\mathbf{b}}$ is the product of $|\vec{\mathbf{b}}|$ and the signed length of $\vec{\mathbf{a}}$ onto the line spanned by $\vec{\mathbf{b}}$. The signed length of the projection is positive if it points in the direction of $\vec{\mathbf{b}}$; it is negative if it points in the opposite direction.

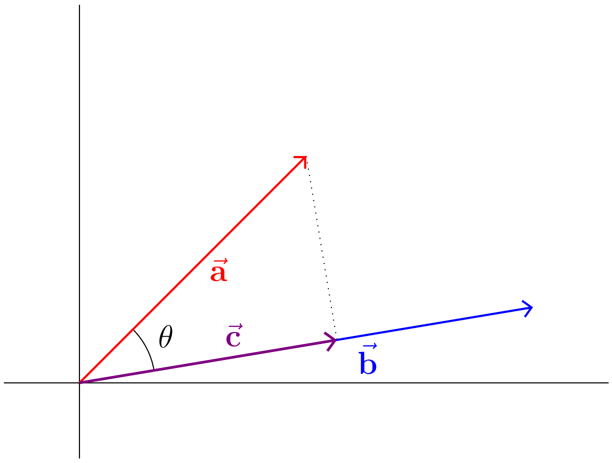


Figure 3.2: Projecting a vector

So above, $\vec{\mathbf{a}} \cdot \vec{\mathbf{b}} = |\vec{\mathbf{a}}||\vec{\mathbf{b}}| \cos\theta = |\vec{\mathbf{b}}||\vec{\mathbf{c}}|$.

3.3 Defining angles between vectors in $\mathbb{R}^n$

Next, we want to use Equation 3.3 to define angles in $\mathbb{R}^n$, where we can't invoke elementary geometry when $n > 3$. Thus, we define

$$\alpha = \arccos \frac{\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}}{|\vec{\mathbf{v}}||\vec{\mathbf{w}}|} \tag{3.8}$$

But we haven't proven that

$$-1 \leq \frac{\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}}{|\vec{\mathbf{v}}||\vec{\mathbf{w}}|} \leq 1. \tag{3.9}$$

This can be proven by the *Schwarz's inequality.*

Theorem 3.2: Schwarz's inequality

For any vectors $\vec{\mathbf{v}}, \vec{\mathbf{w}} \in \mathbb{R}^n$, we have

$$|\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}| \leq |\vec{\mathbf{v}}||\vec{\mathbf{w}}|. \tag{3.10}$$

The two sides are equal if and only if $\vec{\mathbf{v}}$ or $\vec{\mathbf{w}}$ is a multiple of the other.

Proof. If either $\vec{\mathbf{v}}$ or $\vec{\mathbf{w}}$ is $\vec{\mathbf{0}}$ then the statement is obvious. Hence, suppose that both are non-zero. Now consider the function $|\vec{\mathbf{v}} + t\vec{\mathbf{w}}|^2$. Expanding this we get,

$$|\vec{\mathbf{v}} + t\vec{\mathbf{w}}|^2 = |\vec{\mathbf{v}}|^2 + 2t(\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}) + t^2|\vec{\mathbf{w}}|^2, \tag{3.11}$$

which is nothing but a quadratic in t. Since we know that $|\vec{\mathbf{v}} + t\vec{\mathbf{w}}|^2 \geq 0$, we must have the determinant, $4(\vec{\mathbf{v}} \cdot \vec{\mathbf{w}})^2 - 4|\vec{\mathbf{w}}|^2|\vec{\mathbf{v}}|^2 \leq 0$ which gives us the required inequality.

Now we want to prove that equality holds when (say) $\vec{\mathbf{w}} = t\vec{\mathbf{v}}$. Then $|\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}| = |t||\vec{\mathbf{v}}|^2 = (|\vec{\mathbf{v}}|)(|t||\vec{\mathbf{v}}|) = |\vec{\mathbf{v}}||\vec{\mathbf{w}}|$. □

Now we can easily define angles between two vectors,

Definition 3.3 (The angle between two vectors): The *angle* between two vectors $\vec{\mathbf{v}}$ and $\vec{\mathbf{w}}$ in $\mathbb{R}^n$ is that angle α satisfying $0 \leq \alpha \leq \pi$ such that

$$\cos \alpha = \frac{\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}}{|\vec{\mathbf{v}}||\vec{\mathbf{w}}|}. \tag{3.12}$$

An important consequence of the Schwarz's inequality is the triangle inequality.

Theorem 3.3: The triangle inequality

For vectors $\vec{\mathbf{x}}, \vec{\mathbf{y}} \in \mathbb{R}^n$,

$$|\vec{\mathbf{x}} + \vec{\mathbf{y}}| \leq |\vec{\mathbf{x}}| + |\vec{\mathbf{y}}|. \tag{3.13}$$

Proof. This can be proved by the following computation

$$|\vec{\mathbf{x}} + \vec{\mathbf{y}}|^2 = |\vec{\mathbf{x}}|^2 + 2\vec{\mathbf{x}} \cdot \vec{\mathbf{y}} + |\vec{\mathbf{y}}|^2 \leq |\vec{\mathbf{x}}|^2 + 2|\vec{\mathbf{x}}||\vec{\mathbf{y}}| + |\vec{\mathbf{y}}|^2 = (|\vec{\mathbf{x}}| + |\vec{\mathbf{y}}|)^2 \tag{3.14}$$

which implies $|\vec{\mathbf{x}} + \vec{\mathbf{y}}| \leq |\vec{\mathbf{x}}| + |\vec{\mathbf{y}}|$. □

3.4 Measuring matrices

For reasons unclear to me right now, it turns out we need a way to measure the "length" of a matrix. For this we take the $m \times n$ entires of a matrix and imagine it as a point in $\mathbb{R}^m n$ and get the distance as usual:

Definition 3.4 (The length of a matrix): If A is an $n \times m$ matrix, its length $|A|$ is the square root of the sum of squares of all its entries:

$$|A|^2 \coloneqq \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j}^2. \tag{3.15}$$

3.5 Lengths and matrix multiplication

Proposition 3.1 (Lengths of products)**.** *Let A be an $n \times m$ matrix, B an $m \times k$ matrix, and $\vec{\mathbf{a}}$ a vector in $\mathbb{R}^m$. Then*

$$|A\vec{\mathbf{b}}| \leq |A||\vec{\mathbf{b}}|, \tag{3.16}$$
$$|AB| \leq |A||B|. \tag{3.17}$$