EE951: Linear Algebra **eMasters**

Chapter 8: Norms

Author: Ketan Rajawat

Disclaimer: *These notes are taken from various sources, sometimes without proper citation. Do not distribute outside the class or upload on the Internet.*

Contents

8.1 Introduction

The Euclidean norm, also known as the 2-norm or the Euclidean length, measures the "length" or magnitude of a vector in Euclidean space. For a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$, the Euclidean norm, denoted as $\|\mathbf{x}\|$ or $\|\mathbf{x}\|_2$, is defined as:

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

In other words, it is the square root of the sum of the squares of the vector's components. The Euclidean norm provides a measure of the "size" or "magnitude" of a vector. It represents the distance from the origin to the point in Euclidean space represented by the vector. Geometrically, the Euclidean norm corresponds to the length of the vector when it is interpreted as a line segment.

Intuitively, the Euclidean norm captures the notion of distance or magnitude in a familiar way. For example, in two-dimensional space ($\mathbb{R}^2$), the Euclidean norm of a vector (x, y) corresponds to the length of the line segment connecting the origin $(0, 0)$ to the point (x, y), which we recognize as the familiar Pythagorean theorem.

For example, consider the vector $\mathbf{v} = (3, 4)$ in $\mathbb{R}^2$. The Euclidean norm of $\mathbf{v}$ is calculated as:

$$\|\mathbf{v}\| = \sqrt{3^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5$$

So, the Euclidean norm of $\mathbf{v}$ is 5, indicating that the length of the line segment connecting the origin to the point (3, 4) is 5 units.

The Euclidean norm is named after the Greek mathematician Euclid of Alexandria, who lived around 300 BCE, though he himself did not define or use the concept. The formalization of the Euclidean norm can be attributed to the French mathematician and philosopher René Descartes (1596-1650) in his work “La Géométrie,” published in 1637.

8.2 Other Norms

It is possible to define other norms as well, and examples include:

1. *Manhattan norm* (also known as the 1-norm or taxicab norm): The Manhattan norm of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is defined as the sum of the absolute values of its components:

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|$$

   The Manhattan norm measures the distance between two points in a grid-like fashion. It corresponds to the distance that can be traveled by moving only horizontally and vertically (like a taxicab) to reach from one point to another.

2. *Maximum norm* (also known as the infinity norm or Chebyshev norm): The maximum norm of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is defined as the maximum absolute value of its components:

$$\|\mathbf{x}\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$$

   The maximum norm represents the largest absolute value among all the components of the vector. The corresponding distance $\|\mathbf{x} - \mathbf{y}\|_\infty$ measures the “chebyshev distance” or the maximum difference between any two corresponding components of vectors $\mathbf{x}$ and $\mathbf{y}$.

8.2.1 Matrix Norm

The matrix norm, often denoted as $\|\cdot\|$, is a measure of the “size” or “magnitude” of a matrix. It is a function that maps a matrix to a non-negative scalar value. One commonly used matrix norm is the Frobenius norm, which is defined for an $m \times n$ matrix $\mathbf{A}$ as:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2}$$

In other words, the Frobenius norm of a matrix is the square root of the sum of the squares of its individual entries.

The Frobenius norm of $\mathbf{A}$ is equal to the Euclidean norm of the vector obtained by concatenating the column (or row) vectors:

$$\|\mathbf{A}\|_F = \|\mathrm{vec}(\mathbf{A})\|_2 \tag{8.1}$$

Here, $\|\cdot\|_2$ represents the Euclidean norm or 2-norm of a vector.

Similarly, it is easy to show that $\|\mathbf{A}\| = \left\|\mathbf{A}^\top\right\|$ for square matrices. Further, we have that

$$\|\mathbf{A}\|_F^2 = \|\mathbf{a}_1\|^2 + \|\mathbf{a}_2\|^2 + \ldots + \|\mathbf{a}_n\|^2 = \left\|\mathbf{a}_{1\cdot}^\top\right\|^2 + \left\|\mathbf{a}_{2\cdot}^\top\right\|^2 + \ldots + \left\|\mathbf{a}_{m\cdot}^\top\right\|^2$$

where $\mathbf{a}_i$ are the columns and $\mathbf{a}_{j\cdot}$ are the rows of the matrix $\mathbf{A}$. Here, since the rows are $1 \times n$ matrices, we have transposed them into column vectors so as to apply the norm. The transpose operation may be skipped if the norm is suitably defined for row vectors.

8.3 Properties of norm

When considering inner product spaces, we can define the norm as

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \tag{8.2}$$

for any $\mathbf{v} \in \mathbb{V}$. Examples include the Euclidean and Frobenius norms:

$$\|\mathbf{v}\|_2 = \sqrt{\mathbf{v}^\top \mathbf{v}} \qquad\qquad \|\mathbf{A}\|_F = \sqrt{\operatorname{tr}\left(\mathbf{A}^\top \mathbf{A}\right)} \tag{8.3}$$

Remarkably however, the norm can even be defined for spaces without an inner product. Both Manhattan and Chebyshev norms for instance, cannot be written as (8.2) for any valid definition of an inner product.

In general, a norm is any real-valued function that satisfies the following four properties:

1. *Non-negativity:* For any vector $\mathbf{x}$, the norm $\|\mathbf{x}\|$ is non-negative, meaning $\|\mathbf{x}\| \geq 0$.
2. *Homogeneity:* For any vector $\mathbf{x}$ and scalar c, the norm of the scalar multiple $c\mathbf{x}$ is equal to the absolute value of c multiplied by the norm of $\mathbf{x}$. Mathematically, $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$.
3. *Definiteness:* The norm of a non-zero vector is always positive. In other words, if $\mathbf{x}$ is a non-zero vector, then $\|\mathbf{x}\| > 0$. The norm is equal to zero ($\|\mathbf{x}\| = 0$) if and only if the vector is the zero vector ($\mathbf{x} = \mathbf{0}$). Mathematically:
$$\|\mathbf{x}\| = 0 \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{0} \tag{8.4}$$
4. *Triangle Inequality:* For any vectors $\mathbf{x}$ and $\mathbf{y}$, the norm of the sum of the vectors is less than or equal to the sum of their individual norms. Mathematically, $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

The combination of the first and third properties, which state that the norm is always positive, unless the vector itself is zero, is called *positive definiteness.*

Let us establish the first three properties for the Euclidean norm, i.e., $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$.

1. Since $\|\mathbf{x}\|^2$ is sum of squares of numbers, it is non-negative by definition. Its square-root is also non-negative by definition.
2. The homogeneity property follows from the linearity property of inner product. Specifically, it holds that
$$\begin{aligned} \|\alpha \mathbf{v}\| &= \sqrt{(\alpha \mathbf{v})^\top (\alpha \mathbf{v})} && (8.5) \\ &= \sqrt{\alpha^2 \mathbf{v}^\top \mathbf{v}} && (8.6) \\ &= |\alpha| \sqrt{\mathbf{v}^\top \mathbf{v}} && (8.7) \\ &= |\alpha| \, \|\mathbf{v}\| . && (8.8) \end{aligned}$$

3. To establish this, first assume that $\mathbf{x} = \mathbf{0}$, then it follows that $\|\mathbf{x}\| = 0$. For the second step, assume that $\|\mathbf{x}\| = 0$. Then it follows that $x_1^2 + x_2^2 + \ldots + x_n^2 = 0$. We have a sum of non-negative terms being zero, implying that each term must be zero, i.e., $x_1 = x_2 = \ldots = x_n = 0$.

8.4 Norm of sum

Let us derive the formula for the norm of the sum of two vectors. Consider two vectors $\mathbf{u}$ and $\mathbf{v}$. The norm of their sum, $\|\mathbf{u} + \mathbf{v}\|$, can be calculated as follows:

$$\|\mathbf{u} + \mathbf{v}\| = \sqrt{\langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle}$$

Expanding the inner product using the distributive property (**??**), we have:

$$\|\mathbf{u} + \mathbf{v}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle}$$

Using the commutativity property of the inner product ($\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$) and the fact that the inner product is a real scalar ($\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle^*$), we can simplify the expression:

$$\|\mathbf{u} + \mathbf{v}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle}$$

The inner product of a vector with itself gives us the norm squared:

$$\langle \mathbf{u}, \mathbf{u} \rangle = \|\mathbf{u}\|^2 \quad \text{and} \quad \langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$$

Substituting back into the equation, we get:

$$\|\mathbf{u} + \mathbf{v}\| = \sqrt{\|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2}. \tag{8.9}$$

We have the following theorem called the Pythagoras Theorem.

Theorem 8.1. For orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$, it holds that

$$\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 = \|\mathbf{u} + \mathbf{v}\|_2^2 \tag{8.10}$$

Theorem 8.2. For orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$, it holds that

$$\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2 = \|\mathbf{u} + \mathbf{v}\|_2^2 \tag{8.11}$$

Proof: The proof follows from substituting the orthogonality condition $\mathbf{u}^\mathsf{T}\mathbf{v} = 0$ in the expression for the norm of the sum of vectors $\mathbf{u}$ and $\mathbf{v}$. ■

8.5 Orthogonal Decomposition

The goal of orthogonal decomposition is to write a vector $\mathbf{u}$ as a scalar multiple of a given $\mathbf{v} \neq \mathbf{0}$ and another vector $\mathbf{w}$ that is orthogonal to $\mathbf{v}$. If $\mathbf{u}$ and $\mathbf{v}$ are displacement vectors, then the decomposition can be visualized in Fig. 8.1.

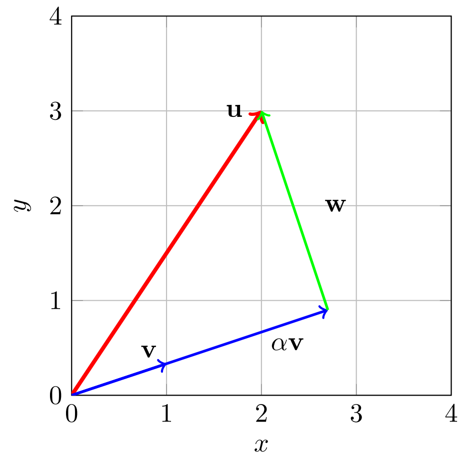

Figure 8.1: Orthogonal decomposition of $\mathbf{u}$ and $\alpha\mathbf{v}$ and $\mathbf{w}$.

To decompose $\mathbf{u}$ in this way we need to add and subtract $\alpha\mathbf{v}$ for some unknown α, so that

$$\mathbf{u} = \alpha\mathbf{v} + (\mathbf{u} - \alpha\mathbf{v}) \tag{8.12}$$

so that $\mathbf{w} = \mathbf{u} - \alpha\mathbf{v}$. We should now see if we can choose α so that $\mathbf{v}$ is orthogonal to $\mathbf{w}$, i.e.,

$$\mathbf{v}^\mathsf{T}\mathbf{w} = \mathbf{v}^\mathsf{T}(\mathbf{u} - \alpha\mathbf{v}) = 0 \tag{8.13}$$

$$\Rightarrow \alpha = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\mathbf{v}^\mathsf{T}\mathbf{v}} = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2}. \tag{8.14}$$

The orthogonal decomposition therefore becomes:

$$\mathbf{u} = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v} + \left(\mathbf{u} - \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}\right) \tag{8.15}$$

Example 8.1. Consider the example in Fig. 8.1. We have that

$$\mathbf{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad \mathbf{v} = \begin{bmatrix} 1 \\ 1/3 \end{bmatrix} \tag{8.16}$$

so that

$$\alpha = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2} = 2.7 \tag{8.17}$$

The orthogonal decomposition is given by

$$\mathbf{u} = 2.7\mathbf{v} + (\mathbf{u} - 2.7\mathbf{v}) = \begin{bmatrix} 2.7 \\ 0.9 \end{bmatrix} + \begin{bmatrix} -0.7 \\ 2.1 \end{bmatrix} \tag{8.18}$$

where the summands can be verified to be orthogonal.

8.6 Cauchy Schwarz and Triangle Inequalities

We look at an important inequality that allows us to define angle between arbitrary vectors.

Lemma 8.3. Suppose that $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$. Then it holds that

$$|\mathbf{u}^\mathsf{T}\mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \tag{8.19}$$

with the equality if and only if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$.

Proof: Let us assume, without loss of generality, that $\mathbf{v} \neq \mathbf{0}$ (otherwise the inequality trivially holds). The orthogonal decomposition of $\mathbf{u}$, as obtained in (8.15) is given by

$$\mathbf{u} = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v} + \mathbf{w}. \tag{8.20}$$

Note that since $\mathbf{v}$ and $\mathbf{w}$ are orthogonal, the Pythagoras theorem implies that

$$\|\mathbf{u}\|^2 = \left\| \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v} \right\|^2 + \|\mathbf{w}\|^2 \tag{8.21}$$

$$= \frac{|\mathbf{u}^\mathsf{T}\mathbf{v}|^2}{\|\mathbf{v}\|^2} + \|\mathbf{w}\|^2 \tag{8.22}$$

$$\geq \frac{|\mathbf{u}^\mathsf{T}\mathbf{v}|^2}{\|\mathbf{v}\|^2} \tag{8.23}$$

where the last inequality holds since $\|\mathbf{w}\|^2 \geq 0$. Re-arranging, we obtain the required inequality.

It can also be seen that the Cauchy-Schwarz (CS) inequality is an equality only when $\mathbf{w} = \mathbf{0}$, which would happen only if $\mathbf{u}$ is along $\mathbf{v}$. Indeed, comparing with (8.15),

$$\mathbf{w} = \mathbf{0} \Leftrightarrow \mathbf{u} = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v}. \tag{8.24}$$

■

An alternative way or writing the CS inequality is

$$-\|\mathbf{u}\| \, \|\mathbf{v}\| \leq \mathbf{u}^\mathsf{T}\mathbf{v} \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \tag{8.25}$$

which follows from the definition of the absolute value function.

The triangle inequality introduced earlier also follows from the CS inequality.

Lemma 8.4. Suppose that $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^n$. Then it holds that

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \tag{8.26}$$

with the equality if and only if $\mathbf{u}$ is a non-negative scalar multiple of $\mathbf{v}$.

Proof: We use the norm of sum result from earlier and the CS inequality to obtain

$$\|\mathbf{u} + \mathbf{v}\|^2 \overset{(8.9)}{=} \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u}^\mathsf{T}\mathbf{v} \tag{8.27}$$

$$\overset{(8.25)}{\leq} \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\| \, \|\mathbf{v}\| \tag{8.28}$$

$$= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \tag{8.29}$$

Taking square roots on both sides and observing that norms are non-negative, we obtain the required result. It can also be seen that the triangle inequality becomes an equality if

$$\mathbf{u}^\mathsf{T}\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{8.30}$$

which holds when $\mathbf{u}$ is a non-negative scalar multiple of $\mathbf{v}$. ■

8.6.1 Angle between vectors

The Cauchy-Schwarz inequality allows us to define the angle θ between non-zero vectors $\mathbf{u}$ and $\mathbf{v}$ as

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}. \tag{8.31}$$

This definition aligns with the Cauchy-Schwarz Inequality, where the absolute value of the inner product is bounded by the product of the magnitudes of the vectors.

Special cases of angles occur when $\theta = 0$ for parallel vectors and $\theta = \pi/2$ or $3\pi/2$ for orthogonal vectors.

8.7 Norm of block vectors

Consider a block vector $\mathbf{x}$ composed of m sub-vectors or blocks:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_m \end{bmatrix} \tag{8.32}$$

where each $\mathbf{x}_i$ represents a sub-vector. We can use the individual norms to calculate the norm of the block vector $\mathbf{x}$ using the formula:

$$\|\mathbf{x}\| = \sqrt{\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 + \ldots + \|\mathbf{x}_m\|^2}.$$

This calculation involves taking the square root of the sum of the squares of the individual norms. An implication of this formula is:

$$\left\| \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right\| = \left\| \begin{bmatrix} \|\mathbf{x}_1\| \\ \|\mathbf{x}_2\| \end{bmatrix} \right\| \tag{8.33}$$

where we observe that if $\mathbf{x}_1 \in \mathbb{R}^{m_1}$ and $\mathbf{x}_2 \in \mathbb{R}^{m_2}$, then the left-hand side involves calculating the norm of a $(m_1 + m_2)$-vector while the right-hand side involves calculating the norm of a 2-vector with scalar entries $\|\mathbf{x}_1\|$ and $\|\mathbf{x}_2\|$.

8.7.1 Orthonormal Vectors

Recall that two vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^n$ are said to be orthogonal if $\mathbf{u}^\mathsf{T}\mathbf{v} = 0$. Further, from the definition of orthogonality (a) $\mathbf{0}$ is orthogonal to every vector since $\mathbf{0}^\mathsf{T}\mathbf{v} = 0$ and (b) $\mathbf{0}$ is the only vector that is orthogonal to itself. Indeed, if there exists a $\mathbf{v}$ such that $\mathbf{v}^\mathsf{T}\mathbf{v} = 0$, then it follows from the definiteness property of norm that $\mathbf{v} = \mathbf{0}$.

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_m\}$ is said to be *orthonormal* if they are pairwise orthogonal and have unit norm, i.e.,

$$\mathbf{v}_i^\mathsf{T} \mathbf{v}_j = 0 \qquad i \neq j \tag{8.34}$$

$$\|\mathbf{v}_i\| = 1 \qquad 1 \leq i \leq m \tag{8.35}$$

Compactly, the condition can be written as

$$\mathbf{v}_i^\mathsf{T} \mathbf{v}_j = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases} \tag{8.36}$$

8.8 Complexity of calculating inner products and norms

Recall that the inner product operation entails multiplying each pair of entries, which requires n flops, and subsequently adding the resulting n terms, which requires $n-1$ flops, and hence the total flop count of inner product is $2n-1$. Finally, norm calculation requires an inner product operation followed by a square root operation, which could take 6 flops on typical CPU architectures, and hence the norm calculation requires $2n+5$ flops.

Computing the inner product of two $n \times n$ matrices involves multiplying each corresponding element and summing the products. For each element, we perform one multiplication, and since there are n^2 elements in total, and summing these elements requires $n^2 - 1$ additions, the flop count for the matrix inner product is $2n^2 - 1$ floating-point operations.

The Frobenius norm of an $n \times n$ matrix is calculated by taking the square root of the sum of the squares of its elements. Computing the square of each element involves n^2 multiplications, and summing these squared elements requires $n^2 - 1$ additions. Finally, taking the square root of the sum requires approximately 6 flops on typical CPU architectures. Hence, the flop count for matrix Frobenius norm is approximately $2n^2 + 5$.

Table 8.1: Flop Counts for operations on n-vectors

Operation	**Flop count (dense)**
Vector Inner Product	$2n$
Vector Norm	$2n$
Matrix Inner Product	$2n^2$
Matrix Frobenius Norm	$2n^2$