

UNIVERSITY OF CALIFORNIA, BERKELEY

LITERALLY EVERYTHING I KNOW ABOUT

Linear Algebra

Warren Alphonso

A very reductionist summary of LINEAR ALGEBRA AND ITS APPLICATIONS by Lay, Lay, and McDonald, as well as LINEAR ALGEBRA DONE WRONG by Treil.

June 28, 2019

Contents

1	Inner Product Spaces	1
1.1	Inner Product	1
1.2	Orthogonality	4
1.3	Orthogonal Projection and Gram-Schmidt Orthogonalization	5

Chapter 1

Inner Product Spaces

Keep in mind that theory for inner product space is only developed for \mathbb{R} and \mathbb{C} , so \mathbb{F} will always denote one of those two fields in this chapter.

1.1 Inner Product

Definition 1.1.1. We define the **norm** of a vector to be the generalization of *length*. That is, the norm of a vector $x \in \mathbb{R}^n$ is

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

For any complex number $z = x + iy$, we can write $|z|^2 = x^2 + y^2 = z\bar{z}$, where \bar{z} denotes the complex conjugate of z . So for any z in a complex field \mathbb{C}^n , we can write

$$z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} x_1 + iy_1 \\ \vdots \\ x_n + iy_n \end{bmatrix}$$

so it is natural to define the norm $\|z\|$ as

$$\|z\|^2 = \sum_{k=1}^n (x_k^2 + y_k^2) = \sum_{k=1}^n |z_k|^2$$

Definition 1.1.2. The **inner product** of two vectors $x, y \in \mathbb{R}^n$ is

$$(x, y) = x_1 y_1 + \cdots + x_n y_n = x^T y = y^T x$$

This yields another definition for the **norm**:

$$\|x\| = \sqrt{(x, x)}$$

For complex fields, we need a definition of inner product such that $\|z\|^2 = (z, z)$. One definition that is consistent with this requirement will be our definition for the **standard inner product in \mathbb{C}^n** ,

$$(z, w) = z_1 \bar{w}_1 + \cdots + z_n \bar{w}_n$$

To simplify this, we will define the **Hermitian adjoint**, or simply **adjoint** A^* , by $A^* = \bar{A}^T$.

Using this, we can write

$$(z, w) = w^* z$$

The inner products we defined for \mathbb{R}^n and \mathbb{C}^n have the following properties:

1. Symmetry: $(x, y) = \overline{(y, x)}$
2. Linearity: $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$
3. Non-negativity: $(x, x) \geq 0$
4. Non-degeneracy: $(x, x) = 0$ if and only if $x = 0$

Note that properties 1 and 2 imply that

$$(x, \alpha y + \beta z) = \overline{(\alpha y + \beta z, x)} = \overline{\alpha(x, y) + \beta(x, z)} = \overline{\alpha}(x, y) + \overline{\beta}(x, z)$$

Lemma 1.1.1. *Let x be a vector in V . Then $x = 0$ if and only if*

$$(x, y) = 0 \quad \forall y \in V$$

Proof. Since $(0, y) = 0$, we need to only show that $x = 0$ if $(x, y) = 0$. Subbing in $y = x$, we get $(x, x) = 0$ and property 3 asserts that $x = 0$.

Lemma 1.1.2. *Let x, y be vectors in V . Then $x = y$ if and only if*

$$(x, z) = (y, z) \quad \forall z \in V$$

Proof. Using the above lemma, if we set $(x - y, z) = 0 \quad \forall z \in V$, then it follows that $x = y$ and $(x, z) = (y, z)$.

Theorem 1.1.3. *Suppose two operators $X, Y : A \rightarrow B$ satisfy*

$$(Ax, y) = (Bx, y) \quad \forall x \in X, \forall y \in Y$$

Then $A = B$.

Proof. Using the previous lemma, we can fix x and take all $y \in Y$, which means $Ax = Bx$. Since this is true for all x , A and B are the same operator. ■

Theorem 1.1.4 (Cauchy-Schwartz Inequality).

$$|(x, y)| \leq \|x\| \cdot \|y\|$$

Proof. If x or y is 0, then the proof is trivial. Assuming neither is 0, we will prove both the real and complex cases. But first consider only the real case:

$$0 \leq \|x - ty\|^2 = (x - ty, x - ty) = \|x\|^2 - 2t(x, y) + t^2\|y\|^2$$

Taking the derivative with respect to t and setting it to 0 gives us $t = \frac{(x, y)}{\|y\|^2}$. We will use this same t value for the following proof of the real and complex cases:

$$\begin{aligned}
0 \leq \|x - ty\|^2 &= (x - ty, x - ty) \\
&= (x, x - ty) - t(y, x - ty) \\
&= \|x\|^2 - \bar{t}(x, y) - t(y, x) + |t|^2 \|y\|^2
\end{aligned}$$

Using property 1 of inner products, we have

$$t = \frac{(x, y)}{\|y\|^2} = \frac{\overline{(y, x)}}{\|y\|^2}$$

Subbing in t , we get

$$0 \leq \|x\|^2 - \frac{|(xy)|^2}{\|y\|^2}$$

which completes the proof. ■

Theorem 1.1.5 (Triangle Inequality).

$$\|x, y\| \leq \|x\| + \|y\|$$

Proof.

$$\begin{aligned}
\|x + y\|^2 &= (x + y, x + y) = \|x\|^2 + \|y\|^2 + (x, y) + (y, x) \\
&\leq \|x\|^2 + \|y\|^2 + 2|(x, y)| \\
&\leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$
■

Theorem 1.1.6. The following **polarization identities** allow us to construct the inner product from the norm:

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

$$(x, y) = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2)$$

For $x, y \in \mathbb{C}^n$,

$$(x, y) = \frac{1}{4} (\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2)$$

Proof. For the real case,

$$\begin{aligned}
\|x + y\|^2 - \|x - y\|^2 &= (x + y, x + y) - (x - y, x - y) \\
&= \|x\|^2 + \|y\|^2 + 2(x, y) - \|x\|^2 - \|y\|^2 + 2(x, y) \\
&= 4(x, y)
\end{aligned}$$

For the complex case,

$$\begin{aligned}
\sum_{k=0}^3 i^k \|x + i^k y\|^2 &= \sum_{k=0}^3 i^k (x + i^k y, x + i^k y) \\
&= \sum_{k=0}^3 i^k (\|x\|^2 + \|y\|^2 + (x, i^k y) + (i^k y, x)) \\
&= \sum_{k=0}^3 (i^k \|x\|^2 + i^k \|y\|^2 + (x, y) + (i^{2k} y, x)) \\
&= 4(x, y)
\end{aligned}$$

where the last step follows from

$$\sum_{k=0}^3 i^k = \sum_{k=0}^3 i^{2k} = 0$$

■

Theorem 1.1.7 (Parallelogram Identity). *Another important property of the norm is the parallelogram identity. For vectors u and v :*

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2)$$

Proof. The theorem follows easily from the fact that the sum of the diagonals of a parallelogram equal the sum of all four sides. ■

To review, we have so far proved the following properties about the norm $\|u\|$:

1. Homogeneity: $\|\alpha u\| = |\alpha| \cdot \|u\|$
2. Triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$
3. Non-negativity: $\|u\| \geq 0$
4. Non-degeneracy: $\|u\| = 0$ if and only if $u = 0$

In a vector space V , if we assign to each vector u a number $\|u\|$ that satisfies these 4 properties, we can say that the space V is a **normed space**.

1.2 Orthogonality

Definition 1.2.1. Two vectors u and v are **orthogonal**, denoted $u \perp v$, if and only if $(u, v) = 0$

Theorem 1.2.1. *If $u \perp v$, then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

Proof.

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2 + (u, v) + (v, u) = \|u\|^2 + \|v\|^2$$

Since $(u, v) = (v, u) = 0$ because of orthogonality. ■

Definition 1.2.2. A vector u is **orthogonal to vector space V** if u is orthogonal to all vectors in V .

Theorem 1.2.2. *Let V be spanned by v_1, \dots, v_n . Then $u \perp V$ if and only if*

$$u \perp v_k \quad \forall k = 1, \dots, n$$

Proof. Proving “only if” is trivial by the definition of $u \perp V$. Proving “if” comes easily after noticing that any vector can be rewritten as a linear combination of the basis vectors, so if u is perpendicular to all the basis vectors, then it is perpendicular to any other vector in V . ■

Definition 1.2.3. A set of vectors v_1, \dots, v_n are orthogonal if any two vectors in the set are orthogonal to each other. If $\|v_k\| = 1$ for all k , we call the set orthonormal.

Lemma 1.2.3 (Generalized Pythagorean Theorem). *Let v_1, \dots, v_n be an orthogonal system. Then*

$$\left\| \sum_{k=1}^n a_k v_k \right\|^2 = \sum_{k=1}^n |a_k|^2 \|v_k\|^2$$

Proof.

$$\left\| \sum_{k=1}^n a_k v_k \right\|^2 = \left(\sum_{k=1}^n a_k v_k, \sum_{j=1}^n a_j v_j \right) = \sum_{k=1}^n \sum_{j=1}^n a_k \overline{a_j} (v_k, v_j)$$

Since the set is orthogonal, (v_k, v_j) is only nonzero when $k = j$, so

$$= \sum_{k=1}^n |a_k|^2 \|v_k\|^2$$

Definition 1.2.4. An orthogonal set of vectors that is also a basis is called an **orthogonal basis**.

Typically, to find coordinates of a vector in a basis, we need to solve a system of equations. For orthogonal bases, it is much simpler. Suppose v_1, \dots, v_n is an orthogonal basis and let

$$x = \alpha_1 v_1 + \dots + \alpha_n v_n$$

Taking the inner product with v_1 yields

$$(x, v_1) = \left(\sum_{j=1}^n \alpha_j v_j, v_1 \right) = \alpha_1 (v_1, v_1) = \alpha_1 \|v_1\|^2$$

Thus, to find any coordinate α_k of a vector x in orthogonal basis v_1, \dots, v_n :

$$\alpha_k = \frac{(x, v_k)}{\|v_k\|^2}$$

This is a simple example of abstract orthogonal Fourier decomposition – simple because classical Fourier decomposition deals with infinite orthonormal systems.

1.3 Orthogonal Projection and Gram-Schmidt Orthogonalization

Definition 1.3.1. The **orthogonal projection** of a vector v onto the subspace E is the vector $w := P_E v$ such that $w \in E$ and $v - w \perp E$.

Theorem 1.3.1. *The orthogonal projection $w = P_E v$ minimizes the distance from v to E . In other words,*

$$\|v - w\| \leq \|v - x\| \quad \forall x \in E$$

Additionally, if for some $x \in E$

$$\|v - w\| = \|v - x\|$$

then $x = w$.

Proof. Let $y = w - x \in E$. Then

$$v - x = v - w + w - x = v - w + y$$

Since $v - w \perp E$, we know $y \perp v - w$. By the Pythagorean Theorem,

$$\|v - x\|^2 = \|v - w\|^2 + \|y\|^2 \geq \|v - w\|^2$$

To finish the proof, note that equality only arises when $y = 0$, ie when $x = w$. ■

There is a formula for finding an orthogonal projection if we know an orthogonal basis in E . Let v_1, \dots, v_n be an orthogonal basis in E . Then the projection $P_E v$ of a vector v is

$$P_E v = \sum_{k=1}^n a_k v_k \quad \text{where} \quad a_k = \frac{(v, v_k)}{\|v_k\|^2}$$

In other words,

$$P_E v = \sum_{k=1}^n \frac{(v, v_k)}{\|v_k\|^2} v_k$$

This is great if we have an orthogonal basis, but if even if we only have a basis in E , we can use the following algorithm to find an orthogonal basis.

Theorem 1.3.2 (Gram-Schmidt Orthogonalization Algorithm). *Suppose we have linearly independent system x_1, \dots, x_n . The Gram-Schmidt algorithm constructs from this an orthogonal system v_1, \dots, v_n such that*

$$\text{span}(x_1, \dots, x_n) = \text{span}(v_1, \dots, v_n)$$

Additionally, for all $r \leq n$

$$\text{span}(x_1, \dots, x_r) = \text{span}(v_1, \dots, v_r)$$

The algorithm is as follows:

1. Define $v_1 := x_1$.

Define $E_1 := \text{span}(v_1) = \text{span}(x_1)$.

2. Define $v_2 := x_2 - P_{E_1} x_2 = x_2 - \frac{(x_2, v_1)}{\|v_1\|^2} v_1$.

Define $E_2 := \text{span}(v_1, v_2) = \text{span}(x_1, x_2)$.

3. Define $v_3 := x_3 - P_{E_2} x_3 = x_3 - \frac{(x_3, v_1)}{\|v_1\|^2} v_1 - \frac{(x_3, v_2)}{\|v_2\|^2} v_2$.

Define $E_3 := \text{span}(v_1, v_2, v_3) = \text{span}(x_1, x_2, x_3)$.

4. Continue until we have n vectors and $\text{span}(v_1, \dots, v_n) = \text{span}(x_1, \dots, x_n)$. The formula for vector v_{r+1} given v_1, \dots, v_r is

$$v_{r+1} := x_{r+1} - P_{E_r} x_{r+1} = x_{r+1} - \sum_{k=1}^r \frac{(x_{r+1}, v_k)}{\|v_k\|^2} v_k$$

Note that at each step, we are adding in x_{r+1} which means the resulting vector will not exist in E_r .

Proof. At each step, we add in x_{r+1} and then subtract its projection the subspace spanned by x_1, \dots, x_r , meaning each additional vector is orthogonal to the ones previously defined. Since we set $v_1 = x_1$, we have proved the algorithm by induction. ■

Since multiplication by a scalar does not change orthogonality, we can multiply vectors v_k returned by Gram-Schmidt by any non-zero numbers. One use case is to normalize the orthogonal vectors by dividing by their norms $\|v_k\|$ to yield an orthonormal system.

Definition 1.3.2. For a subspace E , its **orthogonal complement** E^\perp is the set of all vectors orthogonal to E . Since at least 0 is orthogonal to E , E^\perp is always a subspace.

By the definition of orthogonal projection, any vector in an inner product space V has a unique representation of the form

$$v = v_1 + v_2 \quad v_1 \in E, v_2 \in E^\perp$$

This statement is usually written as $V = E \oplus E^\perp$.

Theorem 1.3.3. For subspace E of V ,

$$(E^\perp)^\perp = E$$

Proof. We will show $E \subseteq (E^\perp)^\perp$ and $(E^\perp)^\perp \subseteq E$.

Let $u \in E$. Then $(u, v) = 0$ for all $v \in E^\perp$. Since u is orthogonal to every vector $v \in E^\perp$, then $u \in (E^\perp)^\perp$ so $E \subseteq (E^\perp)^\perp$.

Now let $u \in (E^\perp)^\perp$. Since $V = E \oplus E^\perp$, we can write $u = v + w$, where $v \in E$ and $w \in E^\perp$. This means that $u - v = w \in E^\perp$. Since we know $E \subseteq (E^\perp)^\perp$, we have $u \in (E^\perp)^\perp$ and $v \in (E^\perp)^\perp$, which means $u - v \in (E^\perp)^\perp$. Therefore, $u - v \in E^\perp \cap (E^\perp)^\perp$. Since the only vector that is orthogonal to itself is 0, $u = v$, and because $v \in E$, $(E^\perp)^\perp \subseteq E$. ■

1.4 Least Square Solution