3 Lecture 3

3.1 Norms

We are now going to define a new class of operators on a linear space that are going to allow us to think about a new set of relations between the elements of that space. First, we are going to define the concept of a **norm.**

Definition 8 *A norm on a linear space* V *is a function* $n : V \to \mathbb{R}$ *that satisfies three properties*

1. $n(x) \geq 0 \ \forall \ x \in V$, *with* $n(x) = 0$ *if and only if* $x = \emptyset$
2. $n(\lambda x) = |\lambda| n(x) \ \forall \ \lambda \in \mathbb{R}, \ x \in V$
3. $n(x + y) \leq n(x) + n(y) \ \forall \ x, y \in V$

Intuitively, a norm is a notion of the 'length' of a vector in a linear space, or the distance of that object from the zero object. This concept becomes clearer with the definition of the Euclidian norm on $\mathbb{R}^n$.

Definition 9 *The Euclidian norm of a vector* $x \in \mathbb{R}^n$ *is represented by* $||x||$ *and defined as*

$$||x|| = \sqrt{\sum_{i=1}^{n} x_i^2}$$

For homework you wil check that the Euclidian norm is in fact a norm.

If you are paying attention, it should be clear that there is some relationship between metrics and norms, in fact the Euclidian metric looks very like the Euclidian norm! The relationship between metrics and norms is not, in fact, isomorphic. While every norm defines a metric such that $d(x, y) = n(x - y)$ it is not the case that every metric defined on a linear space generates a norm in a similar way - that is, it is not the case that for any metric, the function $\bar{n}(x) = d(x, \varnothing)$ is a norm. Consider for example the discrete metric defined on a linear space. then, for $x \neq \emptyset$ and

$r > 1$

$$
\begin{aligned}
&= |r|.\bar{n}(x) \\
&= |r|d(x, \emptyset) \\
&= r \\
&\neq 1 \\
&= d(rx, \emptyset) \\
&= \bar{n}(rx)
\end{aligned}
$$

Thus violating one of the properties of the norm. In fact, we can put properties on a metric so it will generate a norm. This is summed up by the following theorem

Theorem 5 *Let V be a linear space. Then for any norm $||.||$ on L, the function $d(x, y) = ||x - y||$ is a metric. If $d(x, y)$ is a metric, then the function $||x|| = d(x, \emptyset)$ is a norm if and only if*

1. *$d(x, y) = d(x + a, y + a)$ (translation invariance)*
2. *$d(\alpha x, \alpha y) = |\alpha| d(x, y)$ (homogeneity)*

Proof. *Beyond the scope of this course (See OK J.1.4)* ■

Of course, there are lots of other examples of useful norms. One example of a commonly used class on norms are the $\ell - p$ norms. On $\mathbb{R}^n$, $\ell - p$ norms are defined as follows

Definition 10 *On $\mathbb{R}^n$, for any $p \in [1, \infty)$, the $\ell - p$ norm is defined as*

$$||x||_p = \left(\sum_{i=1}^{n} |x_i^p| \right)^{\frac{1}{p}}$$

$\ell - \infty$ is defined as

$$||x||_\infty = \max \{|x_1|, |x_2|, .., |x_n|\}$$

Note that, on $\ell - 2$, this is the Euclidian norm. It is pretty easy to show that $\ell - 1$ and $\ell - \infty$ are norms, It is also fun to figure out what the unit circle looks like under these different norms.

For other values of p, it is somewhat harder to prove that they are norms, but follows easily from Minkowski's Inequality 1 (which we came across in the real analysis bit of the course, but we repeat here)

Theorem 6 *(Minkowski's Inequality 1): For any $x, y \in \mathbb{R}^n$ and $1 \leq p < \infty$*

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{\frac{1}{p}}$$

Proof. *See real analysis lecture notes.* ■

$\ell - p$ norms (and other versions of Minkowski's inequality) can in fact be applied to a much wider class of linear spaces - such as real valued sequences and functions, and form an important area of study in functional and convex analysis. Hopefully we will come back to them if we have time.

Next, we define the concept of an inner product.

Definition 11 *For a linear space V a function $< .,. >: V \times V \rightarrow \mathbb{R}$ is called an inner product if and only if it satisfies the following*

1. $< x, x > \geq 0$ *with* $< x, x >= 0$ *if and only if* $x = \emptyset$
2. $< x, y >=< y, x > \forall \; x, y \in V$
3. $< ax + by, z >= a < x, z > +b < y, z > \forall \; a, b \in \mathbb{R}, \; x, y, z \in V$

As with the definition of addition we used when defining a linear space, there is a 'natural' way of defining an inner product in $\mathbb{R}^n$ called the "dot product".

Definition 12 *The "dot product" of two vectors in $\mathbb{R}^n$ is*

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

Notice, that if we use the Euclidian norm on $\mathbb{R}^n$, then there is a relationship between the dot product and the norm

Remark 4 *If we use $||.||$ to indicate the Euclidian norm on $\mathbb{R}^n$, then*

$$||x||^2 = x.x$$

Proof.

$$\begin{aligned} ||x||^2 &= \sum_{i=1}^{n} x_i^2 \\ &= \sum_{i=1}^{n} x_i x_i \\ &= x.x \end{aligned}$$

■

Thus we could use the dot product to define the Euclidian norm. Can all norms be generated from some inner product in this way? It turns out that the answer is no - only norms that satisfy a condition called the parallelogram property[3] can be generated from some norm. . For example, of the $l-p$ norms on $\mathbb{R}^n$, only $l-2$ can be generated by a norm in this way. However, as we will be spending much of our time working with the euclidian norm and dot product, this is a handy result to know. It also implies another useful results.

Lemma 1 *For $x, y \in \mathbb{R}^n$, $||x+y||^2 = ||x||^2 + ||y||^2 + 2 < x, y >$*

Proof. *(In fact, this is true for any real valued inner product space with associated norm, but we will do it for $\mathbb{R}^n$). First note that*

$$\begin{aligned} ||x+y||^2 &= << x+y >< x+y >> \\ &= < x, x > + < y, y > +2 < x, y > \\ &= ||x||^2 + ||y||^2 + 2 < x, y > \end{aligned}$$

■

To gain insight into what an inner product does, we can show that in $\mathbb{R}^2$ the dot product is related to the angle between two vectors

[3] $2||x||^2 + 2||y||^2 = ||x+y||^2 + ||x-y||^2$

Theorem 7 *In* $\mathbb{R}^2$

$$| < x, y > | = ||x||.||y|| \cos\theta$$

where θ *is the angle between* x *and* y

Proof. *Assume* $\theta \leq \frac{\pi}{2}$. *Set* r *such that* $\Delta(0, x, ry)$ *is a right angle triangle. Then, by definition*

$$\cos\theta = \frac{||ry||}{||x||}$$

Now, using Pythagoras

$$\begin{aligned} ||x||^2 &= ||ry||^2 + ||x - ry||^2 \\ &= r^2||y||^2 + ||x||^2 - 2r < x, y > + r^2||y||^2 \\ &\Rightarrow r||y||^2 =< x, y > \\ &\Rightarrow | < x, y > | = ||x||.||y|| \cos\theta \end{aligned}$$

■

In fact, in more general inner product vector spaces, this is how the concept of an angle is defined

Definition 13 *If* $(V, < ., . >)$ *is an inner product linear space then, for any two vectors* x *and* y *the angle* θ *between* x *and* y *is defined by*

$$\frac{< x, y >}{||x||.||y||} = \cos\theta$$

The above result tells us that in $\mathbb{R}^2$, the dot product of two vectors has to be less than the product of the Euclidian norm of those two vectors. It turns out that this is a more general property.

Theorem 8 ***(Cauchy-Schwartz inequality)*** *For any* $x, y \in \mathbb{R}^n$

$$| < x, y > | \leq ||x||.||y||$$

Proof. *For any* $r \in \mathbb{R}$, $||x - ry|| \geq 0$ *by definition. This implies that*

$$\begin{aligned} 0 &\leq ||x - ry||^2 \\ &< (x - ry), (x - ry) > \\ &= ||x||^2 - 2r < x, y > + r^2||y||^2 \end{aligned}$$

If $y = 0$ we are done (as $|<x,0>| = 0$), so assume not, and set

$$r = \frac{<x,y>}{||y||^2}$$

Which implies that

$$\begin{aligned} 0 &\leq ||x||^2 - \frac{<x,y>^2}{||y||^2} \\ | &< x,y>| \leq ||x||.||y|| \end{aligned}$$

■

We now define the concept of orthogonality.

Definition 14 *2 vectors $x, y \in V$ are orthogonal if $<x,y>= 0$*

Definition 15 *A collection of vectors $\{x_1, .., x_k\} \subset V$ is orthogonal if $<x_i, x_j>= 0 \ \forall \ i \neq j$ and is orthonormal if $||x^i|| = 1 \ \forall \ i$*

Note that, from this definition, it is obvious that the zero vector is orthogonal to each vector.

We define an **orthonormal basis** of an inner product space as a basis of that space whereby each of the elements are orthonormal. It turns out that, using the Gram Schmidt process, any collection of linearly independent vectors can be 'orthonormalized', and so every inner product space has an orthonormal basis.

From theorem 7 it is clear that in $\mathbb{R}^2$, vectors are orthogonal if there is a 90 degree angle between them. Also, it should be clear that you can have at most two (non zero) orthogonal vectors in $\mathbb{R}^2$. This gives a hint to the following theorem

Theorem 9 *An orthogonal set that does not include the zero element is linearly independent.*

Proof. *Exercise* ■

We are now going to define the orthogonal complement of a set. This is just the set of objects that is orthogonal to every object in that set.

Definition 16 *If $V, <.,.>$ is an inner product space, the orthogonal complement of a subset $S \subset V$ is defined as*

$$S^{\perp} = \{x \in V | < x, y >= 0 \ \forall \ y \in S\}$$

For homework, you will prove the following three properties of the orthonormal set.

Theorem 10 *Properties of the orthogonal compliment of $S \subset \mathbb{R}^n$*

1. *$S^{\perp}$ is a linear subspace*
2. *If S is a linear subspace with basis $\{s^1, .., s^k\}$, then*

$$S^{\perp} = \{x \in \mathbb{R}^n | < x, s^i >= 0 \ \forall \ i \in \{1, .., k\}\}$$

3. *If S is a linear subspace, then $S \cap S^{\perp} = \{0\}$*

Proof. *Exercise* ■

One particularly interesting property of a linear subspace and its orthogonal compliment of a set is that it allows us a unique decomposition of any vector into the sum of two other vectors: one in the original set and one in its compliment

Theorem 11 *Let S be a linear subspace of $\mathbb{R}^n$. Then each $x \in V$ admits a* ***unique*** *decomposition of the form*

$$x = s_x + s_x^{\perp}$$

where $s_x \in S$ and $s_x^{\perp} \in S^{\perp}$. We call s_x the ***orthogonal projection*** *of x onto S*

Proof. *Let $\{s^1, ..., s^k\}$ be an orthonormal basis for S. Define*

$$s_x = \sum_{i=1}^{k} < x, s^i > s^i$$

and $s_x^{\perp} = x - s_x$.

Now we need to show that $s_x^\perp \in S^\perp$. By a result that you will prove for homework, it is enough to show that $< s^j, s_x^\perp >= 0 \ \forall \ j$.

$$\begin{aligned} < \ & s^j, s_x^\perp >=< s^j, x - s_x > \\ = \ & < s^j, x > - < s^j, s_x > \\ = \ & < s^j, x > - \sum_{i=1}^{k} < x, s^i >< s^i, s^j > \\ = \ & < s^j, x > - < x, s^j >< s^j, s^j > \\ = \ & 0 \end{aligned}$$

For uniqueness, say that $x = s_x + s_x^\perp = t_x + t_x^\perp$. Then

$$s_x - t_x = s_x^\perp - t_x^\perp$$

But $s_x - t_x \in S$ and $s_x^\perp - t_x^\perp \in S^\perp$ (because both S and $S^\perp$ are linear subspaces), and because the only intersection between S and $S^\perp$ is 0, this implies that

$$s_x - t_x = s_x^\perp - t_x^\perp = 0$$

so $s_x = t_x$ and $s_x^\perp = t_x^\perp$ ■

One reason that the orthogonal projection is interesting is that the element s_x is the *closest* element to x in S in the following sense:

Theorem 12 *Let s_x be the orthogonal projection of a vector x onto a linear subspace S. Then*

$$||s_x - x|| \leq ||s - x|| \ \forall \ s \in X$$

Proof. *Take any $s \in S$, and note that*

$$\begin{aligned} ||s - x||^2 \ &= \ ||(s - s_x) + (s_x - x)||^2 \\ ||(s - s_x)||^2 + ||(s_x - x)||^2 + 2 \ &< \ (s - s_x), (s_x - x) > \end{aligned}$$

but

$$< (s - s_x), (s_x - x) >= 0$$

as $(s - s_x) \in S$ *and* $(s_x - x) = s_x^{\perp} \in S^{\perp}$. *Thus*

$$||s-x||^2 = ||(s-s_x)||^2 + ||(s_x - x)||^2 \geq ||(s_x - x)||^2$$

■