LECTURE TWELVE OF MATH TWO-FOURTY FIVE BY GUEST LECTURE JIAHUI HUANG JUNE NINTH TWENTY-TWENTY SIX

1. Last time

You have seen all these things you can do with inner product. You can do a projection of a vector, you can do all kinds of inequalities like Cauchy-Schwartz and geometric formulas like parallelogram law that you have seen in high school. But don't forget, we didn't make everything so abstract just to work with pictures over $\mathbb{R}$ again. We can also do $\mathbb{C}$, or $\mathbb{F}_p$. But the pictures can still be very useful, as they give intuitions to lots of our useful identities.

Before we move on to the next topic, let us cover another geometric formula.

Proposition 1.1 (Polarization Identity)**.** *Let $\langle \cdot, \cdot \rangle$ be an inner product on V over $\mathbb{R}$. Then*

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

Proof. Just expand $\|x+y\|^2 = \langle x+y, x+y \rangle$ and $\|x-y\|^2 = \langle x-y, x-y \rangle$, and take their difference, then you will get $4\langle x, y \rangle$. □

Geometrically, this is saying the dot product is equal to one quarter of the difference of the lengths of diagonals of your parallelogram, squared. This does not sound geometric at all? Parallelograms are geometric, aren't they?

Remember when we had parallelogram law? It says if you have a norm defined $\|x\| = \sqrt{\langle x, x \rangle}$, then it satisfies an identity called parallelogram law. Since polarization identity has something to do with parallelograms, maybe the it also has something to do with the law. Indeed, we can prove the converse of the law using the identity.

Theorem 1.2. *We have the following.*

1. *(Parallelogram Law) If we start with a inner product on V and defined a norm by $\|x\| = \sqrt{\langle x, x \rangle}$, then it satisfies the parallelogram law.*
2. *(Jordan-von Neumann's theorem) If we start with a norm $\|\cdot\|$ on V that satisfies the parallelogram law, then there exists an inner product on V such that $\|x\| = \sqrt{\langle x, x \rangle}$.*

Proof. We proved (1) last time. Let us prove (2). To get the inner product you want, just define

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

Then you should check why this is really the inner product that we want. This is your homework. □

There is also a polarization identity and Jordan-von Neumann's theorem over $\mathbb{C}$, Their proofs are similar, and you should look them up. By the way, Jordan-von Neumann is two names, not someone whose first name is Jordan-von.

2. Orthonormal basis

Have you heard of when two things are orthogonal? That means perpendicular. More precisely, ortho means perpendicular. If you make 3D models, you might have heard orthographic projections. If you have two cars hitting each other in a perpendicular fashion, you call it an orthocar.

We have some properties of vectors that we would consider very useful, for reasons explained in the next example. These properties are as follows.

Definition 2.1. let V be an inner product space.

1. A set S of vectors is **orthogonal** if for any $x, y \in S$, we have $\langle x, y \rangle = 0$.

(2) A vector x is called a **unit vector** if $\|x\| = 1$.
(3) If $x \in V$ is non-zero, its **normalization** is the unit vector

$$\hat{x} := \frac{x}{\|x\|}.$$

(4) If S is orthogonal and all of its vectors are normalized, then S is **orthonormal**.

Remark 2.2. The definition of unit vector only makes sense over $\mathbb{R}$ or $\mathbb{C}$, because otherwise you can not divide by the norm, which is a real number. ⋄

Example 2.3.

(1) A common example of orthogonal set is the standard basis B_{std} in $\mathbb{R}^n$. In fact, the standard basis is orthonormal because it consists of unit vectors. Suppose we had a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then we quickly note that its B_{std}-component in front of e_i is exactly $x \cdot e_i$.
(2) You already knew that? But did you know if I give you any orthonormal basis $B_{\text{ortho}} = \{b_1, \ldots, b_n\}$ and a vector v, then the B-component in front of b_i is exactly $\langle v, b_i \rangle$? How easy! Therefore orthonormal basis are super easy to work with. Imagine if I just gave you a random basis B and ask you to compute the B-components, that would be so much computations.

⋄

Question 2.4. *Show that if S is an orthogonal set of non-zero vectors, then it is linearly independent.*

Therefore orthonormal sets are automatically linearly independent, because it consists of normalized vectors which are necessarily non-zero. Hence the size of an orthonormal set is at most $\dim V$, and if it is equal to $\dim V$, then it must be an orthonormal basis.

Example 2.5. The set

$$\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

is orthogonal. It is an orthogonal basis of $\mathbb{R}^2$. It is not orthonormal. How do we get an orthonormal basis from it? Just normalize its vectors and get

$$\left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}.$$

This is an orthonormal basis. ⋄

Theorem 2.6. *If $B = \{x_1, \ldots, x_n\}$ is an orthogonal basis, then*

$$B' := \{\hat{x}_1, \ldots, \hat{x}_n\}$$

is an orthonormal basis. It follows that

(1) For every $x \in V$, we have

$$x = \sum_{i=1}^{n} \langle x, \hat{x}_i \rangle \hat{x}_i, \quad \textit{in particular} \quad [x]_{B'} = \begin{pmatrix} \langle x, \hat{x}_1 \rangle \\ \vdots \\ \langle x, \hat{x}_n \rangle \end{pmatrix}.$$

(2) For every $x \in V$, we have

$$x = \sum_{i=1}^{n} \frac{\langle x, x_i \rangle}{\langle x_i, x_i \rangle} x_i, \quad \textit{in particular} \quad [x]_{B'} = \begin{pmatrix} \frac{\langle x, x_1 \rangle}{\langle x_1, x_1 \rangle} \\ \vdots \\ \langle x, \hat{x}_n \rangle \end{pmatrix}.$$

Proof. (1) is just Example 2.3(2). To prove (2), we simply substitute the definition $\hat{x}_i = x_i / \|x_i\|$. □

Look at how ugly (2) is in the above theorem. All those denominators. This tells you why orthonormal is better than orthogonal. But orthogonal is okay, at least we still have a formula, and the ugly fraction

$$\frac{\langle x, x_i \rangle}{\langle x_i, x_i \rangle}$$

can be written more compactly as $\text{proj}_{x_i}(x)$. [[♣♣♣ Draw a picture illustrating how $\text{proj}_{x_i}(x) x_i$ sum to x.]]

3. Gram-Schmidt Process

Now we have seen that if we start with an orthogonal basis, then normalizing it gives us the good good orthonormal basis which we crave. But what if we don't have an orthogonal basis to start with? Are we doomed? Don't worry, let me teach you how to get an orthogonal basis from an arbitrary basis.

Theorem 3.1 (Gram-Schmidt Process)**.** *Let V be an inner product space and let $S = \{v_1, \dots, v_k\}$ be any linearly independent set. Define w_i recursively as follows*

- *Begin with*
$$w_1 = v_1.$$
- *We want to take w_2 to be a vector perpendicular to w_1 within the plane* $\operatorname{span}\{w_1, v_2\}$*. So we set*
$$w_2 = v_2 - \operatorname{proj}_{w_1}(v_2).$$
- *We want w_3 to be perpendicular to w_1, w_2 within the space* $\operatorname{span}\{w_1, w_2, v_3\}$*. So we set*
$$w_3 = v_3 - \operatorname{proj}_{w_1}(v_3) - \operatorname{proj}_{w_2}(v_3).$$
[[♣♣♣ Insert pictures for how these work]]
- *Repeat recursively ...*
- *(General rule) For each $i \geq 2$, we want w_i to be perpendicular to $w_1, \dots, w_{i-1}$ within the space* $\operatorname{span}\{w_1, \dots, w_{i-1}, v_i\}$*. So we set*
$$w_i = v_i - \operatorname{proj}_{w_1}(v_i) - \cdots - \operatorname{proj}_{w_{i-1}}(v_i).$$

Then

(1) The Gram-Schmidt process preserves span in the following sense
$$\operatorname{span}\{w_1, \dots, w_k\} = \operatorname{span}\{v_1, \dots, v_k\}.$$
(2) The set $T = \{w_1, \dots, w_k\}$ is orthogonal.

Corollary 3.2. *If we begin with a basis of V, then apply Gram-Schmidt process, we get an orthogonal basis of V. If we further normalize it, we get an orthonormal basis. Hence every inner product space has an orthonormal basis.*

Proof of Theorem. First, we need to check that each w_i is non-zero inductively, because otherwise a term like $\operatorname{proj}_{w_i}(v_j)$ is not well-defined. Let us prove (1) inductively, during which we see well-definedness comes as a bonus.

If $k = 1$, then we are done. Now suppose $k \geq 2$ and the Gram-Schmidt process of $v_1, \dots, v_{k-1}$ gives us $w_1, \dots, w_{k-1}$ that satisfy (1). Suppose for a contradiction that $w_i = 0$, then
$$v_k = \sum_{i=1}^{k} \operatorname{proj}_{w_i}(v_k).$$
But recall that $\operatorname{proj}_{w_i}(v_k) \in \operatorname{span}\{w_i\}$, which means $v_k \in \operatorname{span}\{w_1, \dots, w_{k-1}\} = \operatorname{span}\{v_1, \dots, v_{k-1}\}$. This is a contradiction because S is linearly independent. On the other hand, note that our definition of w_k gives us
$$w_k \in \operatorname{span}\{w_1, \dots, w_{k-1}, v_k\} = \operatorname{span}\{v_1, \dots, v_k\}$$
and
$$v_k \in \operatorname{span}\{w_1, \dots, w_k\}.$$
This proves $\operatorname{span}\{v_1, \dots, v_k\} = \operatorname{span}\{w_1, \dots, w_k\}$.

For part (2), we simply need to check inductively that $\langle w_k, w_i \rangle = 0$ for $i < k$ by expanding the definition of w_k. This is left as an exercise. □

Question 3.3. *Suppose $\{v_1, \dots, v_{k-1}\}$ is linearly independent, but $\{v_1, \dots, v_k\}$ is linearly dependent. We know we can not apply Gram-Schmidt on linearly dependent sets, but what if we did it anyway? Show that "the Gram-Schmidt process detects linear dependencies" by checking what happens to w_k if you apply the process anyway.*

Everything we have done so far assumes V is finite dimensional. Can you guess what happens to Gram-Schmidt if V is infinite dimensional?