18.700 - LINEAR ALGEBRA, DAY 15
ORTHONORMAL BASES AND GRAM-SCHMIDT
ORTHOGONAL COMPLEMENTS, MINIMIZATION

SAM SCHIAVONE

CONTENTS

I. PRE-CLASS PLANNING

I.1. Goals for lesson.

(1) Students will learn how to construct an orthonormal basis using Gram-Schmidt.
(2) Students will learn how to compute the orthogonal projection of a vector onto a subspace.
(3) Students will learn properties of orthogonal complements.

I.2. Methods of assessment.

(1) Student responses to questions posed during lecture
(2) Student responses to worksheet

I.3. Materials to bring.
(1) Laptop + adapter (2) Worksheets (3) Chalk

II. Lesson Plan

(0:00)

II.1. **Last time.**

- Showed that a linear operator T is diagonalizable iff $\text{minpoly}(T)$ splits into degree 1 factors and has no repeated roots.
- Efficiently computed powers of a linear operator using diagonalization.
- Reviewed properties of inner product and norm for $\mathbb{R}^n$ and $\mathbb{C}^n$.
- Gave definition of an abstract inner product space.

II.2. **6A: Inner products and norms, cont.**

Definition 1. An *inner product* on V is a function

$$\begin{aligned} \langle \cdot, \cdot \rangle : V \times V &\to \mathbb{F} \\ (u, v) &\mapsto \langle u, v \rangle \end{aligned}$$

with the following properties. For all $u, v, w \in V$ and $\lambda \in \mathbb{F}$, we have...

(1) Positivity. $\langle v, v \rangle \geq 0$.
(2) Definiteness. $\langle v, v \rangle = 0$ iff $v = 0$.
(3) Additivity in first component. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$.
(4) Homogeneity in first component. $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$.
(5) Conjugate symmetry. $\langle u, v \rangle = \overline{\langle v, u \rangle}$.

Definition 2. An *inner product space* is a vector space equipped with an inner product.

For the rest of the lecture, let V and W be inner product spaces over $\mathbb{F}$.

Proposition 3. *Suppose $u, v, w \in V$ and $\lambda \in \mathbb{F}$.*

(i) $\langle 0, v \rangle = 0$ *and* $\langle v, 0 \rangle = 0$.
(ii) The function $v \mapsto \langle \cdot, v \rangle$, *i.e.,*

$$\begin{aligned} V &\to \mathbb{F} \\ x &\mapsto \langle x, v \rangle \end{aligned}$$

is linear.
(iii) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$.
(iv) $\langle u, \lambda, v \rangle = \overline{\lambda} \langle u, v \rangle$.

Proof sketch. For part (iii):

$$\langle u, v + w \rangle = \overline{\langle v + w, u \rangle} = \overline{\langle v, u \rangle + \langle w, u \rangle} = \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} = \langle u, v \rangle + \langle u, w \rangle$$

The other parts: exercise. □

Definition 4. Given $v \in V$, the *norm* of v is

$$\|v\| := \sqrt{\langle v, v \rangle}.$$

Proposition 5. *Given $v \in V$ and $\lambda \in \mathbb{F}$,*

(i) $\|v\| = 0$ *iff* $v = 0$; *and*
(ii) $\|\lambda v\| = |\lambda| \|v\|$.

Proof. Exercise. □

Definition 6. Vectors $v, w \in V$ are *orthogonal* if $\langle u, v\rangle = 0$. This is denoted $u \perp v$.

Remark 7. Since $\langle u, v\rangle = 0$ iff $\langle v, u\rangle = 0$, the orthogonality relation is symmetric.

Lemma 8. *Given $u, v \in \mathbb{R}^2$, then*

$$\langle u, v\rangle = \|u\|\|v\|\cos(\theta)$$

where θ is the angle between u and v.

Definition 9. Given $u, v \in V$, we define the *angle between u and v* to be

$$\angle(u, v) := \arccos\left(\frac{\langle u, v\rangle}{\|u\|\|v\|}\right).$$

Remark 10. You will show in an exercise that this definition makes sense.

Lemma 11.

- 0 *is orthogonal to every $v \in V$.*
- 0 *is the only vector in V that is orthogonal to itself.*

Proof. Exercise. □

Theorem 12 (Pythagorean theorem)**.** *If $u, v \in V$ are orthogonal, then*

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

Proof. Exercise. □

Proposition 13 (Cauchy-Schwarz inequality)**.** *Given $u, v \in V$, then*

$$|\langle u, v\rangle| \leq \|u\|\|v\|.$$

Moreover, we have an equality in the above iff u and v are scalar multiples of each other.

Proof. Exercise. □

Proposition 14 (Triangle Inequality)**.** *Given $u, v \in V$, then*

$$\|u + v\| \leq \|u\| + \|v\|.$$

Proposition 15 (Parallelogram identity)**.** *Given $u, v \in V$, then*

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

[Draw picture: $u + v$ and $u - v$ are diagonals of parallelogram.]

Proof. Exercise. □

II.3. **Orthonormal bases and Gram-Schmidt.** Bases of orthogonal vectors, all having length 1, have some very convenient properties. We will see that any basis can be transformed into an orthonormal basis.

Definition 16. A list $e_1, \ldots, e_m$ of vectors is *orthogonal* if $\langle e_i, e_j\rangle = 0$ for all $i \neq j$. It is *orthonormal* it is orthogonal and $\|e_i\| = 1$ for all i.

In other words $e_1, \ldots, e_n$ is orthonormal iff

$$\langle e_i, e_j\rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Example 17.

- The standard basis of $\mathbb{F}^n$ is an orthonormal list.
- The list
$$\frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{2}}(-1,1,0), \frac{1}{\sqrt{6}}(1,1,-2)$$
is orthonormal.

Proposition 18. *Every orthonormal list is linearly independent.*

Proof. Suppose $e_1, \ldots, e_m \in V$ is an orthonormal list. Suppose
$$a_1 e_1 + \cdots + a_m e_m = 0$$
for some $a_1, \ldots, a_m \in \mathbb{F}$. Then
$$0 = \langle 0, e_1 \rangle = \langle a_1 e_1 + \cdots + a_m e_m, e_1 \rangle = a_1 \langle e_1, e_1 \rangle + \cdots + a_m \cancelto{0}{\langle e_m, e_1 \rangle} = a_1$$
so $a_1 = 0$. Similarly applying $\langle \cdot, e_i \rangle$, we find $a_i = 0$ for each i. $\square$

Definition 19. An *orthonormal basis* of V is an orthonormal list in V that is also a basis of V.

In general, given a basis $v_1, \ldots, v_n$ of V and a vector $u \in V$, it can be time-consuming to compute the scalars $a_1, \ldots, a_n \in \mathbb{F}$ realizing u as a linear combination of $v_1, \ldots, v_n$, i.e., such that
$$u = a_1 v_1 + \cdots + a_n v_n \,.$$
However, if this basis is orthonormal, it is easy to compute these a_i.

Proposition 20. *Suppose $e_1, \ldots, e_m$ is an orthonormal basis of V and $u, v \in V$. Then*

(i) $v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$

(ii) $\langle u, v \rangle = \langle u, e_1 \rangle \overline{\langle v, e_1 \rangle} + \cdots + \langle u, e_n \rangle \overline{\langle v, e_n \rangle}$

The following procedure describes how to transform a basis into an orthonormal basis.

Theorem 21 (Gram-Schmidt procedure)**.** *Suppose $v_1, \ldots, v_n$ is a linearly independent list. Let $f_1 := v_1$, and for $k = 2, \ldots, m$, define f_k recursively by*
$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1} \qquad (*)$$
For each k, let $e_k := \dfrac{f_k}{\|f_k\|}$. Then $e_1, \ldots, e_m$ is an orthonormal list in V such that
$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k)$$
for all $k = 1, \ldots, m$.

Proof. By induction on k. Base case: $k = 1$. Then
$$\|e_1\| = \left\| \frac{f_1}{\|f_1\|} \right\| = \frac{\|f_1\|}{\|f_1\|} = 1 \,.$$
Since e_1 is a nonzero multiple of v_1, then $\operatorname{span}(e_1) = \operatorname{span}(v_1)$.

<u>Inductive step</u>: Assume $k \geq 2$ and the result holds for $k-1$, so the list $e_1, \ldots, e_{k-1}$ defined by $(*)$ is orthonormal and

$$\operatorname{span}(e_1, \ldots, e_{k-1}) = \operatorname{span}(v_1, \ldots, v_{k-1}).$$

Since $v_1, \ldots, v_k$ are linearly independent, then

$$v_k \notin \operatorname{span}(v_1, \ldots, v_{k-1}) = \operatorname{span}(f_1, \ldots, f_{k-1}) = \operatorname{span}(e_1, \ldots, e_{k-1})$$

Thus $f_k \neq 0$, so $\|f_k\| \neq 0$. Then

$$\|e_k\| = \left\| \frac{f_k}{\|f_k\|} \right\| = \frac{\|f_k\|}{\|f_k\|} = 1.$$

Given $j \in \{1, \ldots, k-1\}$, then

$$\begin{aligned} \langle e_k, e_j \rangle &= \frac{1}{\|f_k\| \|f_j\|} \langle f_k, f_j \rangle \\ &= \frac{1}{\|f_k\| \|f_j\|} \left\langle v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_j \rangle}{\|f_j\|^2} f_j - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}, f_j \right\rangle. \end{aligned}$$

Since $f_1, \ldots, f_{k-1}$ are orthogonal, this becomes

$$\begin{aligned} &\frac{1}{\|f_k\| \|f_j\|} \left(\langle v_k, f_j \rangle - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} \cancelto{0}{\langle f_1, f_j \rangle} - \cdots - \frac{\langle v_k, f_j \rangle}{\|f_j\|^2} \langle f_j, f_j \rangle - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} \cancelto{0}{\langle f_{k-1}, f_j \rangle} \right) \\ &\quad = \frac{1}{\|f_k\| \|f_j\|} \left(\langle v_k, f_j \rangle - \frac{\langle v_k, f_j \rangle}{\|f_j\|^2} \|f_j\|^2 \right) = \frac{1}{\|f_k\| \|f_j\|} \left(\langle v_k, f_j \rangle - \langle v_k, f_j \rangle \right) = 0. \end{aligned}$$

By solving for v_k in $(*)$, we see that

$$v_k \in \operatorname{span}(f_1, \ldots, f_k) = \operatorname{span}(e_1, \ldots, e_k)$$

and combining this with the inductive hypothesis yields

$$\operatorname{span}(v_1, \ldots, v_k) \subseteq \operatorname{span}(e_1, \ldots, e_k). \tag{$\dagger$}$$

Both $v_1, \ldots, v_k$ and $e_1, \ldots, e_k$ are linearly independent—the v_i by hypothesis, and the e_i because they are orthonormal—so both subspaces have dimension k, hence we have equality in (†). □

We can now add the adjective "orthonormal" to many results about bases of vector spaces.

Proposition 22. *Every finite-dimensional inner product space V has an orthonormal basis.*

Proof. By a previous result, V has a basis $\mathcal{B}$. Apply Gram-Schmidt to $\mathcal{B}$: this produces an orthonormal, hence linearly independent, list of $\dim(V)$ vectors. By another previous result, then this is a basis of V. □

Proposition 23. *Suppose that V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

Proof sketch. Letting L be such a list, then by a previous result, we can extend L to a basis $\mathcal{B}$ of V. Now apply Gram-Schmidt. □

[Skip next two results if necessary.]

Proposition 24. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then there exists an orthonormal basis $\mathcal{E}$ of V such that $[T]_{\mathcal{E}}$ is upper triangular iff* $\operatorname{minpoly}(T)$ *splits into degree* 1 *factors.*

Proof. Exercise. □

Corollary 25. *Suppose V is a finite-dimensional $\mathbb{C}$-inner product space and $T \in \mathcal{L}(V)$. Then there exists an orthonormal basis $\mathcal{E}$ of V such that $[T]_{\mathcal{E}}$ is upper triangular.*

II.4. **6C Orthogonal complements and minimization.**

Definition 26. Given a subset $S \subseteq V$, the *orthogonal complement of S* is

$$S^{\perp} := \{v \in V : \langle u, v \rangle = 0 \ \forall u \in S\} = \{v \in V : v \perp u \ \forall u \in S\}\,.$$

I.e., the set of all vectors that are orthogonal to every vector in S.

Example 27.

- Let $V = \mathbb{R}^3$ and
$$S := \left\{ \begin{pmatrix} 2 \\ -3 \\ 7 \end{pmatrix} \right\}.$$
Then
$$S^{\perp} = \{(x, y, z) \in \mathbb{R}^3 : 2x - 3y + 7z = 0\}\,.$$
[Draw picture.]
- Let $V = \mathbb{R}^3$ and
$$S := \{(x, y, z) \in \mathbb{R}^3 : 2x - 3y + 7z = 0\}\,.$$
Then
$$S^{\perp} = \operatorname{span}\begin{pmatrix} 2 \\ -3 \\ 7 \end{pmatrix} = \left\{ \begin{pmatrix} 2 \\ -3 \\ 7 \end{pmatrix} t : t \in \mathbb{R} \right\}.$$

Proposition 28.

(a) If S is a subset of V, then $S^{\perp}$ is a subspace of V.
(b) [Ask students.] $\{0\}^{\perp} = V$.
(c) [Ask students.] $V^{\perp} = \{0\}$
(d) If S is a subset of V, then $S \cap S^{\perp} = \{0\}$.
(e) If S_1 and S_2 are subsets of V with $S_1 \subseteq S_2$, then $S_1^{\perp} \supseteq S_2^{\perp}$.

Proof. Exercise. □

Part (d) of the above proposition hints at the following result.

Proposition 29. *Suppose U is a finite-dimensional subspace of V. Then*

$$V = U \oplus U^{\perp}\,.$$

Proof. Since U and $U^\perp$ are subspaces, then $0 \in U$ and $0 \in U^\perp$, so $U \cap U^\perp = \{0\}$ by part (d) of the previous result. Thus $U + U^\perp$ is direct.

It remains to show that $V = U + U^\perp$. Certainly $V \supseteq U + U^\perp$, so it suffices to show that $V \subseteq U + U^\perp$. [Ask students.] Suppose $v \in V$. By a previous result, there exists an orthonormal basis $e_1, \dots, e_m$ of U. Let

$$u := \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$
$$w := v - u.$$

Then $v = u + w$ and $u \in U$. <u>Goal:</u> $w \in U^\perp$. [Ask students how to show this.] For each $k \in \{1, \dots, m\}$, we have

$$\langle w, e_k \rangle = \left\langle v - \sum_{i=1}^{m} \langle v, e_i \rangle e_i, e_k \right\rangle = \langle v, e_k \rangle - \sum_i \langle v, e_i \rangle \overbrace{\langle e_i, e_k \rangle}^{=0 \text{ for } i \neq k} = \langle v, e_k \rangle - \langle v, e_k \rangle .$$

Thus w is orthogonal to $e_1, \dots, e_m$, so w is orthogonal to every vector in $\operatorname{span}(e_1, \dots, e_m) = U$. Thus $w \in U^\perp$. □

Corollary 30. *Suppose V is finite-dimensional and U is a subspace of V. Then*

$$\dim(U^\perp) = \dim(V) - \dim(U) .$$

Proposition 31. *Suppose U is a finite-dimensional subspace of V. Then*

$$(U^\perp)^\perp = U .$$

Proof. ($\supseteq$): Exercise.

($\subseteq$): Suppose $v \in (U^\perp)^\perp$. By a previous result, we can write $v = u + w$ where $u \in U$ and $w \in U^\perp$. <u>Goal</u>: $w = 0$. From the first part, we have $u \in U \subseteq (U^\perp)^\perp$, so

$$w = v - u \in (U^\perp)^\perp .$$

But then $w \in U^\perp \cap (U^\perp)^\perp = \{0\}$, so $w = 0$ and $v = u \in U$. □

Corollary 32. *With the same hypotheses as above,*

$$U^\perp = \{0\} \iff U = V .$$

Proof. Exercise. □

Definition 33 (Orthogonal projection)**.** Suppose U is a finite-dimensional subspace of V. For each $v \in V$, we write write $v = u + w$ where $u \in U$ and $w \in U^\perp$. The *orthogonal projection of v onto U* is $\operatorname{proj}_U(v) := u$. This defines a linear map $\operatorname{proj}_U \in \mathcal{L}(V)$.

Since $V = U \oplus U^\perp$, then the expression $v = u + w$ above is unique, so the map proj_U is well-defined.

Proposition 34. *Suppose U is a finite-dimensional subpsace of V. Then*

(i) $\operatorname{proj}_U \in \mathcal{L}(V)$;
(ii) $\operatorname{proj}_U |_U = I_U$, *i.e.,* $\operatorname{proj}_U(u) = u$ *for all* $u \in U$;
(iii) $\operatorname{proj}_U |_{U^\perp} = 0$, *i.e.,* $\operatorname{proj}_U(w) = 0$ *for all* $w \in U^\perp$;
(iv) *[Ask students]* $\operatorname{img}(\operatorname{proj}_U) = U$;

(v) *[Ask students]* $\ker(\text{proj}_U) = U^\perp$;

(vi) $v - \text{proj}_U(v) \in U^\perp$ *for all* $v \in V$;

(vii) $\text{proj}_U^2 = \text{proj}_U$;

(viii) $\|\text{proj}_U(v)\| \leq \|v\|$ *for all* $v \in V$;

(ix) *if* $e_1, \ldots, e_m$ *is an orthonormal basis of* U, *then*

$$\text{proj}_U(v) = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m \,.$$

Proof. Exercise. □

Remark 35. Property (ix) gives us a formula to compute an orthogonal projection, given an orthonormal basis for the subspace.

Proposition 36 (Minimizing distance to a subspace)**.** *Suppose* U *is a finite-dimensional subpsace of* V *and* $v \in V$. *Then*

$$\|v - \text{proj}_U(v)\| \leq \|v - u\|$$

for all $u \in U$, *with equality iff* $u = \text{proj}_U(v)$.

Proof. Given $u \in U$, then $\text{proj}_U(v) - u \in U$. By orthogonal decomposition, $v - \text{proj}_U(v) \in U^\perp$. Since

$$v - u = (v - \text{proj}_U(v)) + (\text{proj}_U(v) - u)$$

and these last two are orthogonal, then

$$\|v - u\|^2 = \|v - \text{proj}_U(v)\|^2 + \overbrace{\|\text{proj}_U(v) - u\|^2}^{\geq 0} \geq \|v - \text{proj}_U(v)\|^2 \,.$$

Taking square roots yields the result. □

In calculus, you were sometimes faced with the following problem. Suppose L is a line through the origin in $\mathbb{R}^2$ and P is a point not lying on the line L. What is the distance from P to L, i.e., what is the point on L closest to P? [Draw picture.]

The answer uses the ideas of orthogonal projection and orthogonal decomposition. Let u be the vector from the origin to P, and let v be a vector in the direction of L. [Continue picture.] Then $L = \text{span}(v)$ and $\frac{1}{\|v\|}v$ is an orthonormal basis for L. By the proposition, then

$$\text{proj}_L(u) = \left\langle u, \frac{1}{\|v\|}v \right\rangle \frac{1}{\|v\|}v = \frac{1}{\|v\|^2}\langle u, v \rangle v$$

is the point on L that is closest to P.

II.5. **Worksheet.**