18.700 - LINEAR ALGEBRA, DAY 14
INNER PRODUCTS AND NORMS,
ORTHONORMAL BASES AND GRAM-SCHMIDT

SAM SCHIAVONE

Contents

I. Pre-class Planning

I.1. Goals for lesson.

(1) Students will compute powers of a linear operator using diagonalization.

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

II. LESSON PLAN

(0:00)

II.1. **Last time.**

- Defined eigenspaces.
- Proved criteria for diagonalizability.

II.2. **Worksheet.**

Theorem 1. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T is diagonalizable iff* $\operatorname{minpoly}(T)$ *splits into degree* 1 *factors and has no repeated roots, i.e., iff there exist distinct* $\lambda_1, \dots, \lambda_m \in \mathbb{F}$ *such that* $\operatorname{minpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_m)$.

Proof. ($\Rightarrow$): Assume T is diagonalizable. Then there is a basis of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then for each j, there exists λ_k such that $(T - \lambda_k I)v_j = 0$. Thus

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)v_j = 0$$

for each j, so $\operatorname{minpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_m)$.

($\Leftarrow$): Assume the $\operatorname{minpoly}(T) = (z - \lambda_1) \cdots (z - \lambda_m)$ where $\lambda_1, \dots, \lambda_m \in \mathbb{F}$ are distinct. By strong induction on m.

Base case: $m = 1$. Then $T - \lambda_1 I = 0$, so $T = \lambda_1 I$, which is diagonalizable.

Inductive step: Assume $m \geq 2$ and the result holds for all $k < m$. By a previous result, $U := \operatorname{img}(T - \lambda_m I)$ is T-invariant, so we can restrict T to this subspace. Given $u \in \operatorname{img}(T - \lambda_m I)$, then [ask students] $u = (T - \lambda_m I)(v)$ for some $v \in V$. Then

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)(T - \lambda_m I)(v) = 0\,, \qquad (\ddagger)$$

so this polynomial kills $T|_U$. Thus $(z - \lambda_1) \cdots (z - \lambda_{m-1})$ is a polynomial multiple of $\operatorname{minpoly}(T|_U)$. By the inductive hypothesis, then $\operatorname{img}(T - \lambda_m I)$ has a basis consisting of eigenvectors of $T|_U$, and hence of T.

Claim: $\operatorname{img}(T - \lambda_m I) + \ker(T - \lambda_m I)$ is direct. Given $u \in \operatorname{img}(T - \lambda_m I) \cap \ker(T - \lambda_m I)$, then $(T - \lambda_m I)(u) = 0 \iff T(u) = \lambda_m u$. Since $u \in \operatorname{img}(T - \lambda_m I)$, then

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (\lambda_m - \lambda_1) \cdots (\lambda_m - \lambda_{m-1})u$$

by ($\ddagger$). Since the λ_i are distinct, this implies that $u = 0$. Thus $\operatorname{img}(T - \lambda_m I) \cap \ker(T - \lambda_m I) = \{0\}$.

Since the sum is direct, then

$$\dim(\operatorname{img}(T - \lambda_m I) \oplus \ker(T - \lambda_m I)) = \dim(\operatorname{img}(T - \lambda_m I)) + \dim(\ker(T - \lambda_m I)) = \dim(V)$$

by Rank-Nullity. Thus $V = \operatorname{img}(T - \lambda_m I) \oplus \ker(T - \lambda_m I)$.

We already saw that $\operatorname{img}(T - \lambda_m I)$ has a basis of eigenvectors of T. Since $\ker(T - \lambda_m I)$ is exatly the λ_m-eigenspace of T, taking a basis of $\ker(T - \lambda_m I)$ and concatenating it with the basis of $\operatorname{img}(T - \lambda_m I)$ yields a basis of V of eigenvectors of T. Thus T is diagonalizable. □

Corollary 2. *Suppose $T \in \mathcal{L}(V)$ is diagonalizable and U is a T-invariant subspace of V. Then $T|_U$ is diagonalizable.*

Proof. Since T is diagonalizable, then $\operatorname{minpoly}(T)$ has no repeated roots. By a previous results, $\operatorname{minpoly}(T)$ is a polynomial multiple of $\operatorname{minpoly}(T|_U)$, so $\operatorname{minpoly}(T|_U)$ also has no repeated roots. □

II.2.1. *Gershgorin disc theorem.* [Cut if necessary.]

Definition 3. Suppose $T \in \mathcal{L}(V)$ and $\mathcal{B} := (v_1, \ldots, v_n)$ is a basis of V. Let $A = [T]_{\mathcal{B}}$. A *Gershgorin disc* of T with respect to $\mathcal{B}$ is a set of the form

$$\left\{ z \in \mathbb{F} : |z - A_{j,j}| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}| \right\}$$

where $j \in \{1, \ldots, n\}$.

Remark 4.

- T has n Gershgorin discs, one for each $j = 1, \ldots, n$.
- For $\mathbb{F} = \mathbb{R}$, the j^{th} disc is a closed interval centered at $A_{j,j}$, with radius the sum of the absolute values of all the entries in the j^{th} row.
- For $\mathbb{F} = \mathbb{R}$, the j^{th} disc is a closed disc centered at $A_{j,j}$. [Draw picture.]

Theorem 5 (Gershgorin Disc Theorem)**.** *With notation as above, each eigenvalue of T is contained in some Gershgorin disc of T with respect to $\mathcal{B}$.*

Proof. Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of T with corresponding eigenvector w. Then we can uniquely write

$$w = c_1 v_1 + \cdots + c_n v_n$$

for some $c_1, \ldots, c_n \in \mathbb{F}$. Let $A = [T]_{\mathcal{B}}$. Applying T to both sides, then

$$\lambda c_1 v_1 + \cdots + \lambda c_n v_n = \lambda w = T(w) = T\left(\sum_{k=1}^{n} c_k v_k \right) = \sum_{k=1}^{n} c_k T(v_k) \,. \tag{6}$$

Now

$$[T(v_k)]_{\mathcal{B}} = [T]_{\mathcal{B}} [v_k]_{\mathcal{B}} = A e_k = \begin{pmatrix} A_{1,k} \\ \vdots \\ A_{n,k} \end{pmatrix} ,$$

so

$$T(v_k) = A_{1,k} v_1 + \cdots + A_{n,k} v_n = \sum_{j=1}^{n} A_{j,k} v_j \,.$$

Substituting this into (6), we have

$$\sum_{k=1}^{n} c_k T(v_k) = \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} A_{j,k} c_k \right) v_j \,.$$

Equating the coefficients of v_j from the above and (6), then

$$\lambda_j = \sum_{k=1}^{n} A_{j,k} c_k \tag{\dagger}$$

for all $j = 1, \ldots, n$. Now let $m \in \{1, \ldots, n\}$ be such that

$$|c_m| = \max\{|c_1|, \ldots, |c_n|\} \,.$$

Taking $j = m$ in (†) and subtracting $A_{m,m}c_m$ from both sides, we have [start in middle]

$$(\lambda - A_{m,m})c_m = \lambda c_m - A_{m,m}c_m = \left(\sum_{k=1}^{n} A_{m,k}c_k\right) - A_{m,m}c_m = \sum_{\substack{k=1\\k\neq m}}^{n} A_{m,k}c_k \,.$$

Then

$$|\lambda - A_{m,m}| = \left|\sum_{\substack{k=1\\k\neq m}}^{n} A_{m,k}\frac{c_k}{c_m}\right| \leq \sum_{\substack{k=1\\k\neq m}}^{n} |A_{m,k}\left|\frac{c_k}{c_m}\right| \leq \sum_{\substack{k=1\\k\neq m}}^{n} |A_{m,k}|$$

by the triangle inequality. Thus λ is in the m^{th} Gershgorin disc. □

II.3. **6A: Inner products and norms.** We defined abstract vector spaces by considering the algebraic properties of addition and scalar multiplication of vectors in $\mathbb{R}^n$, and then abstracting these properties. But vectors in $\mathbb{R}^n$ also have geometric properties, namely length and angle. We now abstract these geometric properties and define inner product spaces.

Throughout this section, let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$, and let V and W be $\mathbb{F}$-vector spaces.

- Inner product on $\mathbb{R}^n$.

  Definition 7. Given $x, y \in \mathbb{R}^n$, their *dot product* or *inner product*, $x \cdot y$, is defined by

  $$x \cdot y := x_1y_1 + \cdots + x_ny_n \,,$$

  where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

  We define the *norm* or *length* of x to be

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

  Note that

  $$\|x\|^2 = x_1^2 + \cdots + x_n^2 = \langle x, x\rangle \,.$$

- Inner product on $\mathbb{C}^n$. Given $\lambda \in \mathbb{C}$ with $\lambda = a + bi$ with $a, b \in \mathbb{R}$, then
  - $|\lambda| = \sqrt{a^2 + b^2}$. [Draw picture.]
  - The *conjugate* of λ is

    $$\overline{\lambda} := a - bi \,.$$

  Note that $\lambda\,\overline{\lambda} = |\lambda|^2$.

  Definition 8. Given $z, w \in \mathbb{C}^n$, their *inner product*, $z \cdot w$, is defined by

  $$z \cdot w := z_1\overline{w_1} + \cdots + z_n\overline{w_n} \,,$$

  where $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$.

  We define the *norm* or *length* of x to be

  $$\|z\| := \sqrt{|z_1|^2 + \cdots + |z_n|^2} \,.$$

  Note that

  $$\|z\|^2 = z_1^2 + \cdots + z_n^2 = \langle z, z\rangle \,.$$

Definition 9. 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. $\langle u,v\rangle = \overline{\langle v,u\rangle}$.

Remark 10.

- Over $\mathbb{R}$, the last condition becomes symmetry: $\langle u,v\rangle = \langle v,u\rangle$.
- Other authors choose to require additivity and homogeneity in the second component.

Definition 11. An *inner product space* is a vector space equipped with an inner product.

Example 12.

- Usual Euclidean inner product on $\mathbb{F}^n$:

$$\langle (w_1,\ldots,w_n),(z_1,\ldots,z_n)\rangle = w_1\overline{z_1} + \cdots + w_n\overline{z_n}\,.$$

- Weighted Euclidean inner product on $\mathbb{F}^n$: given positive real numbers $c_1,\ldots,c_n \in \mathbb{R}$, define

$$\langle (w_1,\ldots,w_n),(z_1,\ldots,z_n)\rangle = c_1w_1\overline{z_1} + \cdots + c_nw_n\overline{z_n}\,.$$

  Then $\langle\cdot,\cdot\rangle$ is an inner product on V: since the c_i are positive, we get positivity.
- Let V be the vector space of continuous functions $[-1,1] \to \mathbb{R}$. Given $f, g \in V$, define

$$\langle f,g\rangle := \int_{-1}^{1} f(x)g(x)\,dx\,.$$

  Exercise: check that this defines an inner product on V. (Positivity and definiteness are the only conditions that are tricky to check.)

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

Proposition 13. *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 14. Given $v \in V$, the *norm* of v is

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

Proposition 15. *Given $v \in V$ and $\lambda \in \mathbb{F}$,*

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

Proof. Exercise. □

Definition 16. Vectors $v, w \in V$ are *orthogonal* if $\langle u, v\rangle = 0$. This is denoted $u \perp v$.

Remark 17. Since $\langle u, v\rangle = 0$ iff $\langle v, u\rangle = 0$, the orthogonality relation is symmetric.

Lemma 18. *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 19. 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 20. You will show in an exercise that this definition makes sense.

Lemma 21.
- *0 is orthogonal to every $v \in V$.*
- *0 is the only vector in V that is orthogonal to itself.*

Proof. Exercise. □

Theorem 22 (Pythagorean theorem). *If $u, v \in V$ are orthogonal, then*

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

Proof. Exercise. □

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.] The point on L lying closest to P will be perpendicular to L. In other words, we want to find a vector w perpendicular to L that ends at P.

In order to find the distance, which is the length of w, we proceed as follows. Letting c be an undetermined scalar, we can write

$$u = cv + (u - cv)$$

as a sum of a component parallel to v, and one perpendicular to v. Since $u - cv$ is orthogonal to v, then

$$0 = \langle u - cv, v\rangle = \langle u, v\rangle - c\langle v, v\rangle = \langle u, v\rangle - c\|v\|^2 .$$

Solving for c, we find

$$c = \frac{\langle u, v, \rangle}{\|v\|^2} .$$

This proof carries over to abstract inner product spaces.

Proposition 23. *Suppose $u, v \in V$, with $v \neq 0$, and let $c := \frac{\langle u, v, \rangle}{\|v\|^2}$. Letting $w = u - \frac{\langle u, v, \rangle}{\|v\|^2} v$, then*

$$u = cv + w \text{ and } \langle w, v\rangle = 0 .$$

Proposition 24 (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 a scalar multiples of each other.

Proof. Exercise. □

Proposition 25 (Triangle Inequality)**.** *Given $u, v \in V$, then*

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

Proposition 26 (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.4. **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 27. A list $e_1, \dots, e_m$ of vectors is *orthonormal* if $\|e_i\| = 1$ for all i, and $\langle e_i, e_j\rangle = 0$ for all $i \neq j$.

In other words $e_1, \dots, e_n$ 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 28.

- 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 29. *Every orthonormal list is linearly independent.*

Proof. Suppose $e_1, \dots, e_m \in V$ is an orthonormal list. Suppose

$$a_1 e_1 + \cdots + a_m e_m = 0$$

for some $a_1, \dots, a_m \in \mathbb{F}$. 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 30. An *orthonormal basis* of V is an orthonormal list in V that is also a basis of V.

In general, given a basis $v_1, \dots, v_n$ of V and a vector $u \in V$, it can be time-consuming to compute the scalars $a_1, \dots, a_n \in \mathbb{F}$ realizing u as a linear combination of $v_1, \dots, v_n$, 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 31. *Suppose $e_1, \dots, e_m$ 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 32 (Gram-Schmidt procedure). *Suppose $v_1, \dots, v_n$ is a linearly independent list. Let $f_1 := v_1$, and for $k = 2, \dots, m$, 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 $e_1, \dots, e_m$ is an orthonorma list in V such that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$$

for all $k = 1, \dots, m$.

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 $\operatorname{span}(e_1) = \operatorname{span}(v_1)$.

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

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

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

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

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

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

$\square$