Part I: Inner product spaces

1 Inner products

Main reference: Axler §1.B

- The *field* $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. Elements of $\mathbb{F}$ are called *numbers* or *scalars*. A *complex number* $z \in \mathbb{C}$ is a number of the form $z = x + iy$ where $x, y \in \mathbb{R}$. In $\mathbb{C}$ we have usual algebraic properties of multiplication and addition, plus the property that $i^2 = -1$, so $(1+2i)(3+4i) = 3 + 4i + 6i + 8i^2 = -5 + 10i$.
- Why $\mathbb{C}$? Cutting a long story short... Want to count: $\mathbb{N}$. Want to subtract: $\mathbb{N} \rightsquigarrow \mathbb{Z}$. Want to divide: $\mathbb{Z} \rightsquigarrow \mathbb{Q}$. Want intermediate value theorem: $\mathbb{Q} \rightsquigarrow \mathbb{R}$. Want polynomials to have roots: $\mathbb{R} \rightsquigarrow \mathbb{C}$.
- A *vector space over the field* $\mathbb{F}$ is a set V together with the operations of addition and scalar multiplication satisfying certain properties.

Main reference: Axler §6.A

- The dot product can be seen as a generalization of multiplication on $\mathbb{R}$. The inner product can be seen as a generalization of the usual dot product on $\mathbb{R}^n$.
- An *inner product* on a vector space V is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$ satisfying:
  1. For every $v \in V$, $\langle v, v \rangle \geqslant 0$ (in particular $\langle v, v \rangle \in \mathbb{R}$);
  2. $\langle v, v \rangle = 0$ if and only if $v = \mathbf{0}$;
  3. For every $u, v, w \in V$ and $\alpha \in \mathbb{F}$, $\langle u + \alpha v, w \rangle = \langle u, w \rangle + \alpha \langle v, w \rangle$;
  4. For every $u, v \in V$, $\langle v, u \rangle = \overline{\langle u, v \rangle}$.

  Examples:

  – The dot product $\langle x, y \rangle = x \cdot y$ on $\mathbb{R}^n$ is the simplest case.
  – For positive real numbers $c_1, \dots, c_n$, we can define $\langle x, y \rangle = \sum_j c_j x_j y_j$.
  – On $\mathbb{C}^n$, we can define $\langle z, w \rangle = \sum_j z_j \overline{w_j}$ or $\langle z, w \rangle = \sum_j c_j z_j \overline{w_j}$.
  – On $\mathcal{P}(\mathbb{C})$, we can define $\langle p, q \rangle = \int_0^\infty e^{-t} p(t) \overline{q(t)} \, \mathrm{d}t$.
- An *inner product space* is a vector space endowed with an inner product.
- $\langle u, v + \alpha w \rangle = \langle u, v \rangle + \overline{\alpha} \langle u, w \rangle$, for every $u, v, w \in V$ and for every $\alpha \in \mathbb{F}$.
- Two vectors $u, v \in V$ are *orthogonal* if $\langle u, v \rangle = 0$. We denote it $u \perp v$.
- $\mathbf{0}$ is the only vector orthogonal to itself.

2 Inner products and norms

Main reference: Axler §6.A

Terminology. In these lecture notes, "proof" means just the main idea of the proof. The complete proof is the one written on the whiteboard or in the textbook.

- On an inner product space V, the *norm* of $v \in V$ is defined by $\|v\| = \sqrt{\langle v, v\rangle}$.
- Properties of a norm:
  1. For every $v \in V$, $\|v\| \geqslant 0$
  2. $\|v\| = 0$ if and only if $v = \mathbf{0}$
  3. For every $u, v \in V$, $\|u + v\| \leqslant \|u\| + \|v\|$
  4. For every $v \in V$ and every $\alpha \in \mathbb{F}$, $\|\alpha v\| = |\alpha| \cdot \|v\|$

  Proof. Triangle inequality is postponed. The other three immediate from $\langle \cdot, \cdot \rangle$.
- *Pythagorean Theorem* : If $u \perp v$, then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.

  Proof. Just expand $\langle u + v, u + v\rangle$ and group the cross term.

  Counter-example: $u = (1, 1 + i), v = (2 - i, -1 - i) \in \mathbb{C}^2$. By carefully examining the proof, we can see why the inner product is in $i\,\mathbb{R}$ for such examples.
- *Orthogonal decomposition:* For $u, v \in V$, $v \neq \mathbf{0}$, we have

  $$u = \lambda v + w, \qquad \text{with } \langle w, v\rangle = 0,$$

  where the scalar λ and vector w are given by $\lambda = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ and $w = u - \lambda v$.

  Proof. Assume there is such a decomposition and compute $\langle u, v\rangle$.
- *Cauchy-Schwarz Inequality.* For every $u, v \in V$, $|\langle u, v\rangle| \leqslant \|u\| \cdot \|v\|$.
  Equality holds if and only if $v = \mathbf{0}$ or if $u = \lambda v$ for some $\lambda \in \mathbb{F}$.

  Proof. Apply Pythagorean Theorem to the orthogonal decomposition of u onto v.

  Example: for continuous functions $f, g : [0, 1] \to \mathbb{R}$,

  $$\int_0^1 f(t)g(t)\,\mathrm{d}t \leqslant \sqrt{\Big(\int_0^1 f(t)^2\,\mathrm{d}t\Big)\Big(\int_0^1 g(t)^2\,\mathrm{d}t\Big)}.$$
- *Triangle Inequality.* For every $u, v \in V$, $\|u + v\| \leqslant \|u\| + \|v\|$.
  Equality holds if and only if $v = \mathbf{0}$ or $u = \lambda v$ for some $\lambda \geqslant 0$.

  Proof. Expand $\|u + v\|^2$ and use Cauchy-Schwarz Inequality.

  Example: for continuous functions $f, g : [0, 1] \to \mathbb{C}$,

  $$\sqrt{\int_0^1 |f(x) + g(x)|^2\,\mathrm{d}t} \leqslant \sqrt{\int_0^1 |f(x)|^2\,\mathrm{d}t} + \sqrt{\int_0^1 |g(x)|^2\,\mathrm{d}t}.$$

3 Orthonormal bases

Main reference: Axler §6.B

- A family of vectors $e_1, \ldots, e_m \in V$ is *orthonormal* if, for every $1 \leqslant j, k \leqslant m$,

$$\langle e_j, e_k \rangle = \begin{cases} 1 \text{ if } j = k, \\ 0 \text{ if } j \neq k. \end{cases}$$

  Examples: $\{(\cos\theta, \sin\theta), (-\sin\theta, \cos\theta)\}$, $\{(0, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}), (1, 0, 0)\}$.

- ○ *Parseval's Identity.* Let $e_1, \ldots, e_m \in V$ be an *orthonormal* family of vectors. For every $a_1, \ldots, a_m \in \mathbb{F}$, we have

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2.$$

  Proof. Pythagorean Theorem.

- If $e_1, \ldots, e_m \in V$ are orthonormal then they are linearly independent.

  Proof. Use previous identity.

- An *orthonormal basis* is a basis consisting of orthonormal vectors.
  Every orthonormal family containing $\mathsf{dim}\, V$ vectors is an orthonormal basis.

- *Decomposition.* Let $e_1, \ldots, e_m$ be an orthonormal basis for V. For every $v \in V$,

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n,$$

  and $\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$.

  Proof. Direct computation and Parseval's identity.

- *The Gram-Schmidt procedure:* Let $v_1, \ldots, v_m \in V$ be linearly independent. Define $e_1 = \frac{1}{\|v_1\|} v_1$ and, for $j = 2, \ldots, m$,

$$\tilde{e}_j = v_j - \langle v_j, e_1 \rangle e_1 - \langle v_j, e_2 \rangle e_2 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1} \qquad e_j = \frac{1}{\|\tilde{e}_j\|} \tilde{e}_j.$$

  Then $e_1, \ldots e_m$ is orthonormal and $\mathsf{span}(v_1, \ldots, v_j) = \mathsf{span}(e_1, \ldots, e_j)$ for every j.

  Example: On $\mathcal{P}_2(\mathbb{R})$ with $\langle p, q \rangle = \int_{-1}^{1} p(t) q(t)\, \mathrm{d}t$, apply G-S to $t^2, 1, t$.

- ○ Let $e_1, \ldots, e_m$ denote the orthonormal family obtained by applying Gram-Schmidt to $v_1, \ldots, v_m$, and let $e'_1, \ldots, e'_m, e'_{m+1}$ denote the family obtained from $v_1, \ldots, v_m, v_{m+1}$. Then $e_1 = e'_1, \ldots, e_m = e'_m$.

  An analogous property would not hold for $v_{m+1}, v_1, \ldots, v_m$.

  That is, Gram-Schmidt procedure gives a consistent output if we add more vectors at the end of the list, but not at the beginning.

4 Schur's Theorem and Riesz Representation Theorem

Main reference: Axler §6.B (Uses results from Axler §5.B)

Notation. A " ∘ " indicates a point that it is not quite following the textbook.

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

  Proof. Apply Gram-Schmidt to an arbitrary basis.

- For **finite-dimensional** inner product spaces, every orthonormal family can be extended to an orthonormal basis.

  Proof. Extend to an arbitrary basis (add at the end) and apply Gram-Schmidt.

- If an operator T on a **finite-dimensional** inner product space has upper-triangular matrix with respect to some basis, then it has an upper-triangular matrix with respect to some orthonormal basis.

  Proof. Recall that $[T]_{v_1,\dots,v_m}$ is upper-triangular if and only if $\mathsf{span}(v_1,\dots,v_j)$ is invariant under T for every $j = 1,\dots,m$. Apply Gram-Schmidt and conclude.

- *Schur's Theorem*: Every operator on a **finite-dimensional** and **complex** inner product space has upper-triangular matrix with respect to some orthonormal basis.

  Proof. Recall that every operator on a finite-dimensional complex vector space is upper-triangular with respect to some basis. Apply previous result.

  Example: compare $T(x,y) = (-y,x)$ on $\mathbb{R}^2$ and $\mathbb{C}^2$.

- A *linear functional* φ on V is a linear map from V to $\mathbb{F}$, i.e. $\varphi \in \mathcal{L}(V,\mathbb{F})$.

  Example: $\varphi(z) = 2z_1 - iz_3 + 3z_3 + 5iz_1$ on $\mathbb{C}^3$.

  Remark. The above example can be written as $\varphi(z) = \langle z, (2-5i, 0, 3+i)\rangle$.

  Example: $\varphi(p) = \int_0^1 p(t)\cos \pi t \,\mathrm{d}t$ on $\mathcal{P}(\mathbb{C})$.

- *Riesz Representation Theorem.* Let V be a **finite-dimensional** inner product space and $\varphi \in \mathcal{L}(V,\mathbb{F})$. There is a unique $u \in V$ such that $\varphi(v) = \langle v, u\rangle \ \forall v$.

  Proof. Take an orthonormal basis, write $u = \sum_j \alpha_j e_j$, and expand $\langle e_j, u\rangle$. Show that this u has the desired property. For uniqueness, compute $\varphi(u - u')$.

  Remark. So we compute u, and the result does not depend on the choice of basis.

∘ As a remark, for $\varphi(p) = \int_0^1 p(t)\cos \pi t \,\mathrm{d}t$ on $\mathcal{P}(\mathbb{C})$, there is no $q \in \mathcal{P}(\mathbb{C})$ with this property. The proof is analytic, requiring Stone-Weierstrass Theorem, etc. This is Functional Analysis, which combines Linear Algebra, Analysis and Topology.

  However, if we restrict this functional φ to $\mathcal{P}_2(\mathbb{C})$, then we can find $q \in \mathcal{P}_2(\mathbb{C})$ such that $\varphi(p) = \langle p, q\rangle$ for every $p \in \mathcal{P}_2(\mathbb{C})$.

5 Orthogonal complement

Main reference: Axler §6.C

Notation. $C_{\mathbb{R}}[a,b] = \{f : [a,b] \to \mathbb{R} \text{ continuous}\}$, $C_{\mathbb{C}}[a,b] = \{f + ig : f, g \in C_{\mathbb{R}}[a,b]\}$.

- If U is a subset of V, then the *orthogonal complement* of U, denoted $U^\perp$, is the set of vectors in V that are orthogonal to every vector in U, that is

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

- Let V be an inner product space. Then

  1. For every subset U of V, $U^\perp$ is a subspace;
  2. $\{\mathbf{0}\}^\perp = V$ and $V^\perp = \{\mathbf{0}\}$;
  3. For every subset $U \subset V$, $U^\perp \cap U \subset \{\mathbf{0}\}$;
  4. For every subsets $U \subset W \subset V$, $W^\perp \subset U^\perp$.

  Proof. Proof. In each case, check the definitions.

- If V is a inner product space and $U \subset V$ is a **finite-dimensional** subspace then

$$V = U \oplus U^\perp.$$

  Proof. For $V = U + U^\perp$, take an orthonormal basis $e_1, \ldots, e_m$ of U, write v as a sum of its components $\langle v, e_j \rangle e_j$ in U plus the remainder w in $U^\perp$.
  Finally, $\oplus$ follows from $U \cap U^\perp = \{\mathbf{0}\}$.

- ○ What can go wrong?
  Let $V = C_{\mathbb{R}}[-1,1]$ with $\langle f, g \rangle = \int_{-1}^{1} f(t)g(t)\,\mathrm{d}t$.
  For $U = \{f \in V : f(0) = 0\}$, $U^\perp = \{\mathbf{0}\}$ and $U + U^\perp \neq V$.

- If U is a finite-dimensional subspace, then $U = (U^\perp)^\perp$.

  Proof. To show that $U \subset (U^\perp)^\perp$ check the definition. For $U \supset (U^\perp)^\perp$, let $v \in (U^\perp)^\perp$, and write $v = u + w$. Since $U \subset (U^\perp)^\perp$, $u \in (U^\perp)^\perp$, hence $(v - u) \in (U^\perp)^\perp$. On the other hand, $(v - u) = w \in U^\perp$, thus $v - u = \mathbf{0}$, hence $v \in U$.

- ○ With the previous counter-example, $(U^\perp)^\perp$ is bigger than U.

- If V is finite-dimensional and U is a subspace, then $\dim U^\perp = \dim V - \dim U$.

  Proof. Just recall that a sum is a direct sum if and only if dimensions add up.

6 Orthogonal projection and best approximation

Main reference: Axler §6.C

In this lecture, we assume that U is a finite-dimensional subspace of V.

- The *orthogonal projection* of V onto U is the map $P_U : V \to U$ defined as follows. For every $v \in V$, write v as

$$v = u + w \quad \text{with } u \in U \text{ and } w \in U^{\perp},$$

  and take $P_U v = u$. This is well-defined because $V = U \oplus U^{\perp}$.

- Properties of the projection operator.

  1. $P_U \in \mathcal{L}(V)$;
  2. $P_U u = u,\ \forall u \in U$ and $P_U w = \mathbf{0},\ \forall w \in U^{\perp}$;
  3. $\mathsf{range}\, P_U = U$ and $\mathsf{ker}\, P_U = U^{\perp}$;
  4. $(v - P_U v) \in U^{\perp}$;
  5. $(P_U)^2 = P_U$;
  6. $\|P_U v\| \leqslant \|v\|$ and equality holds if and only if $v \in U$.
  7. Given an orthonormal basis $e_1, \dots, e_n$ of U, we have, for every $v \in V$,

$$P_U v = \sum_{i=1}^{m} \langle v, e_i \rangle e_i.$$

  Proof. Check the definition of linear. Use the definition of P_U. Use previous property and $V = U \oplus U^{\perp}$. Use the definition of P_U. Use the second property. Pythagorean Theorem to $v = P_U v + (v - P_U v)$. Done in the proof of $V = U \oplus U^{\perp}$.

- *Best approximation.* For every $v \in V$ and every $u \in U$, we have that

$$\|v - P_U v\| \leqslant \|v - u\|,$$

  and equality holds if and only if $u = P_U v$.

  Proof. Apply Pythagorean Theorem to $v - u = (v - P_U v) + (P_U v - u)$.

- The assumption that U is finite-dimensional is of course very restrictive. But the theory is still very powerful under this assumption, see Example 6.58.

∘ Example: Find the affine function that best approximates $g(t) = \sqrt{t}$ in terms of mean squared difference on $[0, 1]$.

Part II: Operators on inner product spaces

We assume that all spaces are finite-dimensional.

7 The adjoint operator

Main reference: Axler §7.A

- Let V and W be two inner product spaces. The *adjoint* of a linear map $T \in \mathcal{L}(V, W)$ is the map T^* defined by $w \mapsto T^*w$ where T^*w is the unique vector in V such tat
$$\langle Tv, w \rangle_W = \langle v, T^*w \rangle_V$$
for every $v \in V$. The operator T^* is well-defined by Riesz Representation Theorem.

- Examples:
$T(x_1, x_2, x_3) = (-2x_2, 5x_3 - 3x_1)$
$Tv = \langle v, u \rangle_V \, x$ for fixed $u \in V$ and $x \in W$
$Tv = Av$ on $\mathbb{R}^4$, where $A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 4 & 5 & 6 & 7 \end{bmatrix}$

- Properties. On the appropriate spaces,

  1. $T^* \in \mathcal{L}(W, V)$
  2. $(S + T)^* = S^* + T^*$
  3. $(\lambda T)^* = \bar{\lambda} T^*$
  4. $(T^*)^* = T$
  5. $(ST)^* = T^* S^*$
  6. $I^* = I$

  Proof. Check the definitions.

- For every T, $\operatorname{ker} T^* = (\operatorname{range} T)^\perp$.
Proof. Use $V^\perp = \{\mathbf{0}\}$ and the definitions of ker, adjoint, $\perp$ and range.

- Corollaries: $\operatorname{range} T = (\operatorname{ker} T^*)^\perp$, $\operatorname{range} T^* = (\operatorname{ker} T)^\perp$, $\operatorname{ker} T = (\operatorname{range} T^*)^\perp$.
Proof. Take $\perp$. Replace T by T^*. Take $\perp$.

- For $T \in \mathcal{L}(V, W)$ and orthonormal bases $\mathcal{A}$ of V and $\mathcal{B}$ of W, we have
$$[T^*]_{\mathcal{A},\mathcal{B}} = ([T]_{\mathcal{B},\mathcal{A}})^*,$$
where A^* denotes the *conjugate transpose* of a matrix A.
Proof. Let $C = [T]_{\mathcal{B},\mathcal{A}}$ and $D = [T^*]_{\mathcal{A},\mathcal{B}}$. Then $c_{j,k} = \cdots = \langle Te_k, g_j \rangle$. Likewise, $d_{j,k} = \langle T^*g_k, e_j \rangle$. On the other hand, $\langle T^*g_k, e_j \rangle = \langle g_k, Te_j \rangle = \overline{\langle Te_j, g_k \rangle} = \overline{c_{k,j}}$.

8 Self-adjoint and normal operators

Main reference: Axler §7.A

- An operator $T \in \mathcal{L}(V)$ is *self-adjoint* if $T^* = T$.
- In some sense, taking the adjoint of a linear operator is the analogous of taking the complex conjugate of a number. In the same spirit, since a number being real is equivalent to being equal to its conjugate, the property of an operator being self-adjoint has many similarities with the property of a number being real.
- If $T \in \mathcal{L}(V)$ is self-adjoint, then all its eigenvalues are real.

  Proof. Show that $\lambda\|v\|^2 = \bar{\lambda}\|v\|^2$.
- Let V be a **complex** inner product space and let $T \in \mathcal{L}(V)$ be an operator. T is self-adjoint if and only if $\langle Tv, v\rangle \in \mathbb{R}$ for every $v \in V$.

  Proof. Check that $\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle (T - T^*)v, v\rangle$ and use following fact.

  * If $\langle Sv, v\rangle = 0$ for every $v \in V$, then $S = \mathbf{0}$.

  Proof. For $u, w \in V$, expanding $0 = \sum_k i^k\langle S(u + i^k w), u + i^k w\rangle = 4\langle Su, w\rangle$ gives $Su \in V^{\perp} = \{\mathbf{0}\}$. Example: Compare $S(x, y) = (-y, x)$ on $\mathbb{R}^2$ and $\mathbb{C}^2$.
- Let V be an inner product space (real or complex) and let $T \in \mathcal{L}(V)$ be a **self-adjoint** operator. If $\langle Tv, v\rangle = 0$ for every $v \in V$, then $T = \mathbf{0}$.

  Proof. For $\mathbb{F} = \mathbb{C}$ it is the previous theorem, so we can assume $\mathbb{F} = \mathbb{R}$. Expanding $0 = \langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle$ gives $4\langle Tu, w\rangle = 0$, hence $Tu \in V^{\perp} = \{\mathbf{0}\}$.
- An operator $T \in \mathcal{L}(V)$ is *normal* if $TT^* = T^*T$.

  Examples:
  $T(x, y) = (-2y, 2x)$ on $\mathbb{F}^2$. Check that $T^* = (2y, -2x)$, $TT^* = T^*T = (4x, 4y)$.
  $T(x, y) = (y, y)$. Check that $T^* = (0, x + y)$, $TT^* = (x + y, x + y)$, $T^*T = (0, 2y)$.
- An operator $T \in \mathcal{L}(V)$ is normal if and only if $\|Tv\| = \|T^*v\|$ for every $v \in V$.

  Proof. Write $S = T^*T - TT^*$. We have $S = S^*$, want to show that $S = \mathbf{0}$. Show that $\langle Sv, v\rangle = \|T^*v\|^2 - \|Tv\|^2$ and use previous theorem.
- If $T \in \mathcal{L}(V)$ is a normal operator with eigenvector $v \in V$ associated to some eigenvalue $\lambda \in \mathbb{F}$, then v is also an eigenvector of T^* associated to the eigenvalue $\bar{\lambda}$.

  Proof. Show that $T - \lambda I$ is also normal, whence $\|Tv - \lambda v\| = \|T^*v - \bar{\lambda}v\|$.
- If $T \in \mathcal{L}(V)$ is a normal operator then eigenvectors associated to distinct eigenvalues are orthogonal.

  Proof. By previous theorem, $Tu = \alpha u$ and $T^*v = \bar{\beta}v$. Show that $(\alpha - \beta)\langle u, v\rangle = 0$.

9 Spectral Theorem

Main reference: Axler §7.B

- Spectral Theorems. Why do we care? Diagonal operators? Orthogonal bases?
- Can you diagonalize $T(x, y) = (-2y, 2x)$ on $\mathbb{C}^2$? What about $\mathbb{R}^2$?
- Can you diagonalize $T(x, y) = (2x + y, x + 2y)$ on $\mathbb{C}^2$? What about $\mathbb{R}^2$?
- *Complex Spectral Theorem:* Let V be a **complex** inner product space and $T \in \mathcal{L}(V)$ be an operator. T is **normal** if and only if it is orthonormal diagonalizable.

  Proof. ($\Leftarrow$) We know that $[T^*] = [T]^*$, so $[T^*]$ is also diagonal with respect to the same orthonormal basis, thus $[T]^*$ and $[T]$ commute, hence T^* and T commute.
  ($\Rightarrow$) By Schur's Theorem, there is orthonormal basis $\mathcal{A} = \{e_1, \ldots, e_n\}$ such that $[T]_{\mathcal{A}}$ is upper-triangular. Using $[T^*] = [T]^*$, $\|Te_j\| = \|T^*e_j\|$ and Pythagorean Theorem, show recursively that off-diagonal entires of each row of $[T]$ are zero.

- *Real Spectral Theorem.* Let V be a **real** inner product space and let $T \in \mathcal{L}(V)$ be an operator. Then T is **self-adjoint** if and only if it is orthonormal diagonalizable.

  We prove the Real Spectral Theorem with three lemmas.

  An *irreducible quadratic factor* is a polynomial of the form $q(t) = t^2 + 2bt + c$ for real numbers $c > b^2$. Note that $q(t) > 0$ for every $t \in \mathbb{R}$.

  * If $T^* = T$ and $c > b^2 \geqslant 0$, then $S = T^2 + 2bT + cI$ is invertible.

  Proof. For $\|v\| = 1$, expand using $(T + bI)^* = T + bI$ to show $\langle Sv, v\rangle > 0$.

  * If $V \neq \{\mathbf{0}\}$ and $T^* = T$, then T has at least one eigenvalue.

  Proof. Assume $\mathbb{F} = \mathbb{R}$. Fix $v \neq \mathbf{0}$. Writing $n = \mathsf{dim}\, V$, the family $v, Tv, T^2v, \ldots, T^nv$ must be linearly dependent, whence $\sum_j a_jT^jv = \mathbf{0}$. Consider the non-zero polynomial $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, and recall that every polynomial over $\mathbb{R}$ can be factorized as a constant times several irreducible quadratic factors times several factors of the form $x - \lambda_k$ for $\lambda_k \in \mathbb{R}$. Since $p(T)v = \mathbf{0}$, at least one of the linear factors $T - \lambda_kI$ is not injective, and therefore λ_k is an eigenvalue. The case $\mathbb{F} = \mathbb{C}$ is similar but simpler, since polynomials over $\mathbb{C}$ factorize completely.

  * If $T^* = T$ and $U \subset V$ is an invariant subspace under T, then $U^\perp$ is invariant under T, $T_{|_U} \in \mathcal{L}(U)$ is self-adjoint and $T_{|_{U^\perp}} \in \mathcal{L}(U^\perp)$ is self-adjoint.

  Proof. For $T(U^\perp) \subset U^\perp$, check the definition using $T^* = T$. For $T_{|_U}$ self-adjoint, check the definition. For $T_{|_{U^\perp}}$ self-adjoint, switch U and $U^\perp$.

  Proof of Real Spectral Theorem: ($\Leftarrow$) Since $[T]$ is diagonal with respect to an orthonormal basis, we have that $[T^*] = [T]^* = [T]^t = [T]$, hence $T^* = T$.
  ($\Rightarrow$) Use the two lemmas above and induction on the dimension.

10 Positive operators and isometries

Main reference: Axler §7.C

- The idea is that every linear map can be "diagonalized" by two orthonormal bases. Whereas a diagonal operator only does some "stretching," having different bases reflects some "rotation" done on the top of that. By flipping elements of the second base, we can even restrict the stretching to positive real scalars. As we will see later, stretching is represented by a positive operator, and rotation by an isometry.

- An operator $T \in \mathcal{L}(V)$ is *positive* if $T^* = T$ and if $\langle Tv, v \rangle \geqslant 0$ for every $v \in V$.

  Examples: The identity operator is positive. An orthogonal projection P_U is positive. If $T^* = T$ and $c > b^2$ are real, then $T^2 + 2bT + cI$ is positive. (7.32)

  Remark: By the Spectral Theorem, a self-adjoint operator is positive if and only if all eigenvalues are non-negative.

- Let $T \in \mathcal{L}(V)$ be an operator. An operator $R \in \mathcal{L}(V)$ is a *square root* of T if $T = R^2$. (7.33)

  Example: $T(x, y, z) = (z, 0, 0)$ and $R(x, y, z) = (y, z, 0)$ on $\mathbb{C}^3$.

- If T is a positive operator, then T has a *unique* positive square root, denoted $\sqrt{T}$.

  Proof. Use the Spectral Theorem. For existence, take R having eigenvalues $\sqrt{\lambda_k}$ with same orthonormal eigenvectors as T. For uniqueness, let v be an eigenvector of T, so $Tv = \lambda v$. Let $\sqrt{\lambda_k}$ denote the eigenvalues of R for an orthonormal basis $(e_1, \ldots, e_n)$, show that $\sum_k a_k \lambda_k e_k = \sum_k a_k \lambda e_k$. Conclude that a_k with different eigenvalues are zero, hence $Rv = \sqrt{\lambda} v$. Since eigenvectors form a basis, this determines R. (7.36)

- An operator $S \in \mathcal{L}(V)$ is an *isometry* if $\|Sv\| = \|v\|$ for every $v \in V$.

◦ Let $S \in \mathcal{L}(V)$. The following are equivalent:

  1. S is an isometry;
  2. $\langle Su, Sv \rangle = \langle u, v \rangle$ for every $u, v \in V$;
  3. $Se_1, \ldots, Se_m$ is orthonormal for every orthonormal family $e_1, \ldots, e_m$;
  4. $Se_1, \ldots, Se_n$ is orthonormal for some orthonormal basis $e_1, \ldots, e_n$;
  5. S is invertible and $S^{-1} = S^*$;
  6. S^* is an isometry.

  Proof. First $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 6$. Use that $4\langle u, v \rangle = \sum_k i^k \|u + i^k v\|^2$. Orthonormality is preserved once inner product is preserved. Existence follows from Gram-Schmidt. Mapping a basis to a basis, S is invertible and, by linear combinations on a basis it preserves inner product, so $\langle u, S^{-1}v \rangle = \langle Su, SS^{-1}v \rangle$ for all u, v. $\langle S^*v, S^*v \rangle = \langle SS^*v, v \rangle = \|v\|^2$. For $6 \Rightarrow 1$, use $1 \Rightarrow 6$ and $T^{**} = T$. (7.42)

11 Polar Decomposition & Singular Value Decomposition

Main reference: Axler §7.D

Numbers refer to page or otherwise equation/theorem numbering in Axler.

- *Polar Decomposition.* Let T be an operator on V. There exist an isometry S and a positive operator P such that $T = SP$. This P can be given by $P = \sqrt{T^*T}$.

  Proof. Use $T^*T = P^*P$ to show $\|Pv\| = \|Tv\|$. Let $U = \operatorname{range} P$ and $W = \operatorname{range} T$. Define $S_1 \in \mathcal{L}(U, W)$ by $S_1 Pv = Tv$. Since S_1 preserves norm, it is injective. It follows from the definition that it is also surjective, so $\dim W = \dim U$. Define $S_2 \in \mathcal{L}(U^\perp, W^\perp)$ by $Te_j = g_j$ for some orthonormal bases. Define $S = S_1 P_U + S_2 P_{U^\perp}$. Since S_1 and S_2 preserve norm, by orthogonality so does S. (7.45)

- Let $T \in \mathcal{L}(V)$. The *singular values* of T is the list consisting of the eigenvalues of $\sqrt{T^*T}$, where an eigenvalue λ is counted $\dim\ker(\sqrt{T^*T} - \lambda I)$ times.

  Remark: Singular values are nonnegative. By the Spectral Theorem there are always $\dim V$ of them. (236)

  Example: With computations we check that $T(x, y, z, w) = (0, 3x, 2y, -3w)$ has singular values $(0, 2, 3, 3)$ but only 0 and -3 are eigenvalues. (7.49)

- *Singular Value Decomposition.* Let $T \in \mathcal{L}(V)$. There exist orthonormal bases $\mathcal{A}$ and $\mathcal{B}$ such that $[T]_{\mathcal{B},\mathcal{A}}$ has the singular values on the diagonal and zero elsewhere.

  Proof. Express $P = \sqrt{T^*T}$ in terms an orthonormal eigenbasis. Using Polar Decomposition, apply S to this expression, getting $Tv = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n$. (7.51)

Part III: Operators on complex vector spaces

We assume that $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and $\dim V = n \geqslant 1$.

12 Generalized eigenvectors

Main reference: Axler §8.A

- Let $T \in \mathcal{L}(V)$. We have $\{\mathbf{0}\} = \ker T^0 \subset \ker T^1 \subset \ker T^2 \subset \ker T^3 \subset \cdots$

  Proof. Just check the definition.

- If $\ker T^m = \ker T^{m+1}$, then $\ker T^m = \ker T^{m+1} = \ker T^{m+2} = \ker T^{m+3} = \cdots$

  Proof. Show that $v \in \ker T^{m+k+1}$ is also in $\ker T^{m+k}$ using $T^k v \in \ker T^{m+1}$. (8.3)

- We have $\ker T^k = \ker T^n$ for every $k \geqslant n$.

  Proof. Each time there is strict inclusion, the dimension increases by at least one.

- We have $V = \ker T^n \oplus \operatorname{range} T^n$.

  Proof. To show $\ker \cap \operatorname{range} = \{\mathbf{0}\}$, write $v = T^n u$, and show $u \in \ker T^{2n} = \ker T^n$. Finally, dimensions add up by Rank-Nullity Theorem.

  After-class exercise: Let $T(z_1, z_2, z_3) = (4z_2, 0, 5z_3)$. Check that $\ker T \cap \operatorname{range} T \neq \{\mathbf{0}\}$ and $\mathbb{C}^3 \neq \ker T + \operatorname{range} T$ but $\mathbb{C}^3 = \ker T^3 \oplus \operatorname{range} T^3$ by finding describing these subspaces explicitly. (8.5)

- A vector $v \in V$ is a *generalized eigenvector* associated to λ if $v \neq \mathbf{0}$ and if there exist a nonnegative integer j such that $(T - \lambda I)^j v = \mathbf{0}$.

  Remark: In this case λ must be an eigenvalue, because otherwise $(T - \lambda I)$ being injective would imply that $(T - \lambda I)^j$ is injective.

- Ideally, one should be able to decompose V as a direct sum of invariant subspaces where T behaves nicely. We would like to have $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$ but that is not always possible. Introducing generalized eigenvectors is an attempt to give a good description of linear operators, as simple and complete as possible.

- The subspace $G(\lambda, T) = \ker(T - \lambda I)^n$ is called the *generalized eigenspace* of T associated to the eigenvalue λ.

  Exercise: Let $T(z_1, z_2, z_3) = (4z_2, 0, 5z_3)$. Find the eigenvalues of T and the generalized eigenspaces. Check that $\mathbb{C}^3$ is a direct sum of generalized eigenspaces.

- Let $v_1, \ldots, v_m$ be generalized eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ of $T \in \mathcal{L}(V)$. Then the vectors $v_1 \ldots, v_m$ are linearly independent.

  Proof. Take $w = (T - \lambda_1 I)^k v_1$ as a λ_1-eigenvector. So $(T - \lambda I)^n w = (\lambda_1 - \lambda)^n w$. Apply $(T - \lambda_1 I)^k \prod_{j>1} (T - \lambda_j I)^n$ to both sides of $\sum_j a_j v_j = \mathbf{0}$ and conclude that $a_1 = 0$. To get $a_j = 0$ we repeat the same argument. (8.13)

13 Generalized eigenvectors and nilpotent operators

Main reference: Axler §8.A & §8.B

- An operator $N \in \mathcal{L}(V)$ is *nilpotent* if $N^n = \mathbf{0}$.

  Examples: The derivative on $\mathcal{P}_7(\mathbb{R})$, or $N(z_1, z_2, z_3, z_4) = (z_2 + z_3, z_3, z_4, 0)$ on $\mathbb{C}^4$.

- Let $N \in \mathcal{L}(V)$ be a nilpotent operator. There exists a basis $\mathcal{B}$ such that $[N]_{\mathcal{B}}$ is upper-triangular with only 0's on the diagonal.

  Proof. Choose basis for $\ker N$, extend it to a basis of $\ker N^2$, and so on. (8.19)

- Let $T \in \mathcal{L}(V)$ be an operator and let λ be an eigenvalue of T. The *algebraic multiplicity* of λ is defined as $\dim \ker(T - \lambda I)^n$. The *geometric multiplicity* of λ is defined as $\dim \ker(T - \lambda I)$ and can be smaller than the algebraic multiplicity.

  Exercise: For $T(z_1, z_2, z_3) = (6z_1 + 3z_2 + 4z_3, 6z_2 + 2z_3, 7z_3)$, find the eigenvalues, generalized eigenspaces, and a basis of $\mathbb{C}^3$ made of generalized eigenvectors.

Below we assume that $\mathbb{F} = \mathbb{C}$.

◦ Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of $T \in \mathcal{L}(V)$. Then

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T).$$

  In particular, there is a basis of generalized eigenvectors, and the sum of algebraic multiplicities of $\lambda_1, \ldots, \lambda_m$ equals n. Moreover, for all j, the subspace $G(\lambda_j, T)$ is invariant under T and the restriction of $T - \lambda_j I$ to $G(\lambda_j, T)$ is nilpotent.

  Proof. We use the observation that $\ker p(T)$ and $\operatorname{range} p(T)$ are invariant under T for any complex polynomial $p(z)$, in particular for $p(z) = (z - \lambda_j)^n$. Invariance of $G(\lambda_j, T)$ under T follow from this, nilpotence follows from definition of $G(\lambda, T)$. For the direct sum, recall that operators on complex spaces have at least one λ_1, write $V = G(\lambda_1, T) \oplus U$ with $U = \operatorname{range}(T - \lambda_1 I)^n$ invariant under T. Let $S = T_{|U} \in \mathcal{L}(U)$. The λ_1-eigenvectors are in $G(\lambda_1, T)$, hence not in U. By induction on n, $U = G(\lambda_2, S) \oplus \cdots \oplus G(\lambda_m, S)$. Since $G(\lambda, S) \subset G(\lambda, T)$, $G(\lambda_1, T) + \cdots + G(\lambda_m, T) = V$. Direct sum follows from linear independence of generalized eigenvectors. (8.21)

- Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of $T \in \mathcal{L}(V)$ and $d_1, \ldots, d_m$ their algebraic multiplicities. There exists a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}$ is block diagonal with blocks $A_1, \ldots, A_m$ such that A_j, $j \in \{1, \ldots, m\}$, is a $d_j \times d_j$ upper-triangular matrix with λ_j on the diagonal.

  Proof. Defining $N_j \in \mathcal{L}(G(\lambda_j, T))$ by $N_j v = (T - \lambda_j I)v$, we have N_j nilpotent, so there is a basis $\mathcal{B}_j$ of $G(\lambda_j, T)$ such that $[N_j]_{\mathcal{B}_j}$ is upper-triangular with zeros on the diagonal, adding back $\lambda_j I$ gives the upper-triangular matrix with λ_j on the diagonal. (8.29)

14 Characteristic and minimal polynomials

Main reference: Axler §8.C

- A *monic polynomial* is a polynomial whose leading coefficient is 1.
- There exists a unique monic polynomial $p \in \mathcal{P}(\mathbb{F})$ of smallest degree such that $p(T) = \mathbf{0}$. The polynomial p is called the *minimal polynomial* of T.

  Proof. Consider the largest m such that $I, T, T^2, \ldots, T^m$ is linearly independent (such m exists since $\mathcal{L}(V)$ is finite-dimensional). Then $I, T, T^2, \ldots, T^m$ is a basis for its span, which in turn contains T^{m+1} by definition of m. Therefore, there is a unique monic polynomial $p(z) = z^{m+1} + a_m z^m + \cdots + a_1 z + a_0$ of degree $m+1$ such that $p(T) = \mathbf{0}$. Moreover, if q is a polynomial of smaller degree such that $q(T) = \mathbf{0}$, by linear independence of $I, T, T^2, \ldots, T^m$ we have $q = \mathbf{0}$, so q cannot be monic. (8.40)

- Let p be the minimal polynomial of T and let $q \in \mathcal{P}(\mathbb{F})$ be a polynomial. We have that $q(T) = \mathbf{0}$ if and only if p divides q, i.e. $q = p \times s$ for some polynomial s.

  Proof. The more difficult direction uses polynomial division. (8.46)

- Let V be a finite-dimensional vector space and let $T \in \mathcal{L}(V)$. The zeros of the minimal polynomial of T are exactly the eigenvalues of T.

  Proof. If $p(\lambda) = 0$, we have $(T - \lambda I)q(T)v = \mathbf{0}$ for every v, and since $q(T) \neq \mathbf{0}$, λ is an eigenvalue. Conversely, if λ is an eigenvalue then for an eigenvector we get $\mathbf{0} = p(T)v = p(\lambda)v$, hence $p(\lambda) = 0$. (8.49)

Below we assume that $\mathbb{F} = \mathbb{C}$.

- Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of $T \in \mathcal{L}(V)$ and $d_1, \ldots, d_m$ their algebraic multiplicities. The *characteristic polynomial* of T is defined by

$$g(z) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}.$$

  Remark: $g(z)$ has degree n and its zeros are the eigenvalues.

- *Cayley-Hamilton Theorem*: $g(T) = \mathbf{0}$.

  Proof. Every v is a combination of vectors $v_j \in \ker(T - \lambda_j I)^{d_j}$, and since the operators $(T - \lambda_j I)^{d_j}$ commute, $g(T)v_j = \mathbf{0}$. (8.37)

- The characteristic polynomial is a multiple of the minimal polynomial.

  Proof. Combine Cayley-Hamilton and the property above. (8.48)

15 Jordan form

Main reference: Axler §8.D

For this lecture, $\mathbb{F} = \mathbb{C}$

- Suppose N is nilpotent. Then V has a basis which is of the form

  $$v_1, Nv_1, N^2v_1, \ldots, N^{d_1}v_1, v_2, Nv_2, \ldots, N^{d_2}v_2, \ldots, v_m, Nv_m, N^2v_m, \ldots, N^{d_m}v_m,$$

  and is such that $N^{d_j+1}v_j = \mathbf{0}$ for $j = 1, \ldots, m$.

  Remark: The only eigenvalue of N is 0, and m is just enumerating different families. Each family j starts with some vector v_j and ends when $NNN\cdots NNv_j = \mathbf{0}$.

  Proof. We will show how to find such a basis. Let $U_k = \operatorname{range} N^k$. A basis of a subspace $U \subset V$ with the above properties will be called a *good* basis for U.

  First, $U_n = \{\mathbf{0}\}$ and the empty set is a good basis for U_n.

  Let $\mathcal{B}_{k+1} = (v'_1, Nv'_1, \ldots, N^{d'_1}v'_1, \ldots, v'_{m'}, Nv'_{m'}, \ldots, N^{d'_m}v'_{m'})$ be a good basis for U_{k+1}. We will construct a good basis for U_k. Since $U_{k+1} = \operatorname{range}(N_{|U_k})$, we can find a solution v_k to $Nv_k = v'_k$, for $k = 1, \ldots, m'$. Let $d_j = d'_j+1$ for $j = 1, \ldots, m'$. Take the collection $\tilde{\mathcal{B}}_k = (v_1, Nv_1, \ldots, N^{d_1}v_1, \ldots, v_{m'}, Nv_{m'}, \ldots, N^{d_m}v_{m'})$ obtained by inserting $v_1, \ldots, v_{m'}$ before each subfamily of $\mathcal{B}_{k+1}$.

  To show that $\tilde{\mathcal{B}}_k$ is linearly independent, we consider a null linear combination, apply N and use linear independence of $\mathcal{B}_{k+1}$ to show that the all the coefficients except the last one of each family are zero. For the last coefficients of each family, use again linear independence of $\mathcal{B}_{k+1}$. Complete $\tilde{B}_k$ to a basis of U_k by adding $w_{m'+1}, \ldots, w_m$.

  Since $Nw_{m'+j} \in U_{k+1}$, it is a linear combination of $\mathcal{B}_{k+1}$, which means a linear combination of $N\tilde{\mathcal{B}}_k$ which means Nx_j for some $x \in \operatorname{span}\tilde{\mathcal{B}}_k$. Take $v_{m'+j} = w_{m'+j} - x_j$ so $Nv_{m'+j} = \mathbf{0}$. Take the collection $\mathcal{B}_k$ obtained by adding $v_{m'+1}, \ldots, v_m$ to $\tilde{\mathcal{B}}_k$. Then $\mathcal{B}_k$ spans U_k, so it is a basis, and moreover it is a good basis. (8.55)

- A basis $\mathcal{B}$ is a *Jordan basis* for T if $[T]_\mathcal{B}$ is block diagonal with each block A_i such that its diagonal consists entirely of λ_i's, where λ_i is an eigenvalue of T, and the line above the diagonal consists entirely of 1's.

  Remark: If N is nilpotent, then a good basis as in the previous proof, written backwards, is a Jordan basis for N.

- Let V be a complex finite-dimensional vector space. For any operator $T \in \mathcal{L}(V)$, there exists a Jordan basis for T.

  Proof. Recall that $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$, where $G(\lambda_j, T)$ is invariant under T and $N_j = (T - \lambda_j I)_{|G(\lambda_j,T)}$ is nilpotent. Find a Jordan basis for each N_j. These bases together provide a Jordan basis for T. (8.60)

16 Computing a Jordan basis

Main references: Jordan-compute.pdf and Jordan-thread.pdf For this lecture, $\mathbb{F} = \mathbb{C}$

- ○ To get a Jordan basis, we need the eigenvalues. Then work with $N = T - \lambda I$ for each λ separately. In fact, we can work with different λ in parallel in case we don't know the dimensions of each $G(\lambda, T)$, so we know when to stop. In order to start, we may need to find either some eigenvectors or generalized eigenvectors.
- A *thread* is a chain of non-zero vectors $v, Nv, N^2v, \ldots, N^kv$. Typically, $N^{k+1}v = \mathbf{0}$ in which case we say that it is a *closed thread.*

  Remark. This is not a standard terminology.
- We represent a thread by $v_0 \mapsto v_1 \mapsto v_2 \mapsto \cdots \mapsto v_k \mapsto \mathbf{0}$.
- A Jordan basis for a nilpotent operator can always be written as a collection of closed threads. Conversely, if a collection of closed threads has n vectors and they are linearly independent, these vectors form a Jordan basis. Moreover, it is enough to check linear independence of the last vector of the threads.
- ○ To find a Jordan basis we must work with rational numbers written as fractions. Numerical software cannot be used, because even the existence of "repeated eigenvalues" is a property that breaks down in the presence of rounding error. More precisely, we often need to apply row reduction to non-invertible matrices which become invertible (but horrible) in the presence of a small rounding error.
- The Quick and Dirty Method starts with some vector u and tries to find long threads $v_0 \mapsto v_1 \mapsto v_2 \mapsto \cdots \mapsto v_k \mapsto \mathbf{0}$ where $v_j = u$ for some j. To enlarge the thread forward just apply N, to enlarge it backward we need to solve an equation. The method is not guaranteed but we can always try to get n vectors and test whether $Q^{-1}AQ$ is in Jordan form.
- ○ The Thorough Method (a.k.a. Guaranteed Method) is based on the proof given in the previous lecture. It is probably the most economic if we wanted to find a Jordan basis that has a large number of threads with different lengths (which we don't want), but it is very rigid and really a pain in the neck.
- The Simpler Method is more playful and flexible. We can start with an *ad hoc* collection of threads, reduce the collection to get linear independence, and add more chains gradually if needed to span the whole space (or subspace).
- ○ These names for the methods are not standard terminology either.

Part IV: Operators on real vector spaces

For this part, $\mathbb{F} = \mathbb{R}$, $\dim V = n \geqslant 1$.

17 Complexification I

Main reference: Axler §9.A

- We have already given the best possible description to operators on complex spaces. The idea now is to import that knowledge and use it to help us study operators on real vector spaces.
- The *complexification* $V_{\mathbb{C}}$ of V is the complex vector space consisting of ordered pairs $(u, v) \in V_{\mathbb{C}} = V \times V$ with the addition defined naturally and the (complex) scalar multiplication defined by

  $$(a + ib)(u, v) = (au - bv, bu + av).$$

  If we use the notation $(u, v) = u + iv$, then the scalar multiplication follows the usual multiplication rules for complex numbers.

  Short description: $V_{\mathbb{C}} = V + iV$ with sum of two vectors and multiplication of a vector by a complex scalar multiplication defined in the obvious way.
- If $v_1, \ldots, v_n$ is a basis of V, then it is also a basis of $V_{\mathbb{C}}$. So $\mathsf{dim}_{\mathbb{R}} V = \mathsf{dim}_{\mathbb{C}} V_{\mathbb{C}}$.
- Definition. Let $T \in \mathcal{L}(V)$. The *complexification* $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ of T is defined by $T_{\mathbb{C}}(u + iv) = Tu + iTv$.

  Remark: Not every $S \in \mathcal{L}(V_{\mathbb{C}})$ corresponds to $Tu + iTv$ for some $T \in \mathcal{L}(V)$.
- Let $\mathcal{B}$ be a basis of V. Then $[T]_{\mathcal{B}} = [T_{\mathbb{C}}]_{\mathcal{B}}$.
- ∘ The minimal polynomial of T is equal to the minimal polynomial of $T_{\mathbb{C}}$.

  Proof. First note that $(T_{\mathbb{C}})^k(u + iv) = T^k u + iT^k v$ for every $u, v \in V$ and $k \in \mathbb{N}$, whence $g(T_{\mathbb{C}})u = g(T)u$ for real polynomial g and $u \in V$.

  Let p be the minimal polynomial of T, and let $v_1, \ldots, v_n$ be a basis for V. Then it is also a basis for $V_{\mathbb{C}}$, and since $p(T_{\mathbb{C}})v_j = p(T)v_j = \mathbf{0}$ for all j, we have $p(T_{\mathbb{C}}) = \mathbf{0}$.

  Now suppose $q = f + ig$ is a complex polynomial with smaller degree such that $q(T_{\mathbb{C}}) = \mathbf{0}$. Then $q(T_{\mathbb{C}})z = \mathbf{0}$ for all $z \in V_{\mathbb{C}}$, and in particular $q(T_{\mathbb{C}})u = \mathbf{0}$ for all $u \in V$. But $q(T_{\mathbb{C}})u = f(T)u + ig(T)u$, hence $f(T) = \mathbf{0}$ and $g(T) = \mathbf{0}$. By minimality of p, both f and g are the zero polynomial, and so is q. (here)
- ∘ If $\langle \cdot, \cdot \rangle$ is an inner product on V then $\langle u+iv, x+iy \rangle_{\mathbb{C}} = \langle u, x \rangle + \langle v, y \rangle + i\langle v, x \rangle - i\langle u, y \rangle$ defines a complex inner product on $V_{\mathbb{C}}$.

  Proof. Long computations show that it satisfies the 4-5 axioms. (Exercise 9.B.3)

18 Complexification II

Main reference: Axler §9.A

For convenience, fix a basis for V and work with matrix representation $A = [T] \in \mathbb{R}^{n\times n}$.

- Let $\lambda \in \mathbb{R}$. Then λ is an eigenvalue of T if and only if it is an eigenvalue of $T_{\mathbb{C}}$.

  Proof. Suppose λ is an eigenvalue for T. Then there is $u \in V \setminus \{\mathbf{0}\}$ such that $Tu = \lambda u$. On the other hand, $T_{\mathbb{C}}u = Tu = \lambda u$ so λ is also an eigenvalue of $T_{\mathbb{C}}$.

  Let λ be an eigenvalue for $T_{\mathbb{C}}$. Then $T_{\mathbb{C}}(u+iv) = \lambda(u+iv) = (\lambda u)+i(\lambda v)$ for some $z = u + iv \in V_{\mathbb{C}} \setminus \{\mathbf{0}\}$. On the other hand, $T_{\mathbb{C}}(u + iv) = Tu + iTv$, so $Tu = \lambda u$ and $Tv = \lambda v$. Since u and v cannot be both zero, λ is an eigenvalue. (9.11)

- ○ For any $\lambda \in \mathbb{C}$, $j \in \mathbb{N}$ and $u, v \in V$,

$$(T_{\mathbb{C}} - \lambda I)^j(u + iv) = \mathbf{0} \Leftrightarrow (T_{\mathbb{C}} - \overline{\lambda} I)^j(u - iv) = \mathbf{0}.$$

  Proof. $(A - \bar{\lambda}I)^j\bar{x} = (\bar{A} - \bar{\lambda}\bar{I})^j\bar{x} = \overline{(A - \lambda I)}^j\bar{x} = \overline{(A - \lambda I)^j x}$. (here)

- Let $\lambda \in \mathbb{C}$. Then λ is an eigenvalue of $T_{\mathbb{C}}$ if and only if $\overline{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$.

  Proof. Take $j = 1$.

- ○ For $T_{\mathbb{C}}$, the algebraic multiplicity of λ equals the algebraic multiplicity of $\overline{\lambda}$.

  Proof. If $x_1, \ldots, x_d$ is a basis for $G(\lambda, A)$, then $\bar{x}_1, \ldots, \bar{x}_d$ is a basis for $G(\bar{\lambda}, A)$.

- Every operator on an odd-dimensional real vector space has at least one eigenvalue.

  Proof. The non-real eigenvalues of $T_{\mathbb{C}}$ come in pairs with equal multiplicity. (9.19)

- ○ The (geometric and algebraic) multiplicities of $\lambda \in \mathbb{R}$ for T are the same for $T_{\mathbb{C}}$.

  Proof: The number of free variables in the echelon form of $A - \lambda I$ and $(A - \lambda I)^n$.

- The characteristic polynomial of $T_{\mathbb{C}}$ has real coefficients.

  Proof. $(z - \lambda)^m(z - \bar{\lambda})^m = (z^2 - 2\Re(\lambda)z + |\lambda|^2)^m$ has real coefficients. (9.20)

- ○ The *characteristic polynomial* of T is defined as the characteristic polynomial of $T_{\mathbb{C}}$. Its degree is $\dim V$ and its real roots are the eigenvalues of T.

- *Cayley-Hamilton Theorem.* $p(T) = \mathbf{0}$.

  Proof. By the complex Cayley-Hamilton Theorem, $[p(T)] = p([T]) = p([T_{\mathbb{C}}]) = \mathbf{0}$.

- The minimal polynomial divides the characteristic polynomial.

  Proof. The minimal polynomial divides any polynomial that annihilates T.

19 Normal operators on real inner product spaces

Main reference: Axler §9.B

- **Theorem.** If $T \in \mathcal{L}(V)$ is normal, then there exists an orthonormal basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is block diagonal with each block either 1×1 with a real eigenvalue or 2×2 of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ where $b > 0$ and $\lambda = a + ib$ an eigenvalue of $T_{\mathbb{C}}$.

  Lemma 1. Every operator has an invariant subspace of dimension 1 or 2.

  Proof. We know that $T_{\mathbb{C}}$ has at least one eigenvalue $a + ib$. So there exist $u, v \in V$ such that $u + iv \in V_{\mathbb{C}} \setminus \{\mathbf{0}\}$ is an eigenvector. Hence, $T_{\mathbb{C}}(u+iv) = (a+ib)(u+iv) = (au - bv) + i(av + bu)$. On the other hand, $T_{\mathbb{C}}(u + iv) = Tu + iTv$. Thus, $Tu = (au - bv)$ and $Tv = (av + bu)$. Therefore, $\mathsf{span}(u, v)$ is invariant and has dimension 1 or 2. (9.8)

  Lemma 2. If U is a 2-dimensional real inner product space, and $T \in \mathcal{L}(U)$ is normal but not self-adjoint, then there exists an orthonormal basis $\mathcal{B}$ of U such that $[T]_{\mathcal{B}} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ with $a, b \in \mathbb{R}$, $b > 0$.

  Proof. Pick any o.n. basis and check that $[T]$ must have this form with $\pm b$. (9.27)

  Lemma 3. If $T \in \mathcal{L}(V)$ is a normal operator and U is a subspace invariant under T, then $U^{\perp}$ is invariant under T, moreover $T_{|_U}$ and $T_{|_{U^{\perp}}}$ are normal operators.

  Proof. For invariance of $U^{\perp}$, extend an orthonormal basis for U and write $[T] = [A, B; \mathbf{0}, C]$. To show that $B = \mathbf{0}$, we use $\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m} \|T^* e_j\|^2$ and repeat the same trick as in the Complex Spectral Theorem. Invariance of U and $U^{\perp}$ under T^* follows from $[T^*] = [A^*, \mathbf{0}; \mathbf{0}, C^*]$. Normality of $T_{|_U}$ and $T_{|_{U^{\perp}}}$ follows from normality of A and C, which in turn follow from normality of $[T]$. (9.30)

  Proof of the theorem. Work by induction on $n = \mathsf{dim}\, V$.

  For $n = 1$ the description follows trivially, so suppose $n \geqslant 2$ and that the theorem holds for spaces of dimension smaller than n.

  Let U be an invariant subspace of dimension 1, if such subspace exists. If not, let U be an invariant subspace of dimension 2 (it exists by Lemma 1). In the former case, take $v \in U$ with $\|v\| = 1$ and let $\mathcal{B} = \{v\}$. In the latter case, we know that $T_{|_U}$ is normal (by Lemma 3) but not self-adjoint (otherwise T would have eigenvalues by the Real Spectral Theorem and we would not be in the latter case), hence by Lemma 2 there is a basis $\mathcal{B}$ of U for which $[T_{|_U}]$ is a 2×2 matrix of the above form.

  On the other hand, by Lemma 3, $U^{\perp}$ is invariant and $T_{|_{U^{\perp}}}$ is normal. By the induction hypothesis, there is an orthonormal basis $\mathcal{B}'$ for $U^{\perp}$ for which $[T_{|_{U^{\perp}}}]$ has the desired form. Then for the orthonormal basis $\mathcal{A} = \mathcal{B}' \cup \mathcal{B}$ of V we have $[T]_{\mathcal{A}}$ in the desired form, proving the theorem. (9.34)

20 Isometries and 2x2 block diagonalization

Main reference: these notes.

- If $S \in \mathcal{L}(V)$ is an isometry, then there exists an orthonormal basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is block diagonal with each block either $[\pm 1]_{1\times 1}$ or of the form $[\cos\theta, -\sin\theta; \sin\theta, \cos\theta]_{2\times 2}$ with $\theta \in (0, \pi)$.

  Proof. Use the previous theorem and the assumption that S is an isometry. (9.36)

- If the real vectors $u_1, \dots, u_n$ form a basis for $V_{\mathbb{C}}$, then they form a basis for V.

  Proof. If $\sum_j \alpha_j u_j = \mathbf{0}$ in V, then $\sum_j \alpha_j u_j = \mathbf{0}$ in $V_{\mathbb{C}}$, implying $\alpha_j = 0$.

- If $2m$ vectors $u_1 \pm iv_1, \dots, u_m \pm iv_m$ are L.I., then $u_1, v_1, \dots, u_m, v_m$ are L.I. in $V_{\mathbb{C}}$.

  Proof. These vectors span the same subspace of $V_{\mathbb{C}}$, hence they are L.I. in $V_{\mathbb{C}}$.

- If the m complex vectors $u_1 + iv_1, \dots, u_m + iv_m$ are L.I. in $V_{\mathbb{C}}$, then the real family $u_1, v_1, \dots, u_m, v_m$ has at least m vectors which are L.I. in $V_{\mathbb{C}}$.

  Proof. The span of $u_1, v_1, \dots, u_m, v_m$ contains the span of $u_1 + iv_1, \dots, u_m + iv_m$, which has complex dimension m. Hence, at least m of them are L.I. in $V_{\mathbb{C}}$.

- If $u + iv$ is an eigenvector of $T_{\mathbb{C}}$ with eigenvalue $a + ib$ with $b > 0$, then $\mathsf{span}\{u, v\}$ is two-dimensional and invariant under T, and $[T|_{\mathsf{span}\{u,v\}}]_{\{v,u\}} = \left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]$.

  Proof. Since $u - iv$ is an eigenvector for $a - ib$, they are linearly independent and $\mathsf{span}\{u \pm iv\} = \mathsf{span}\{u, v\}$ is two-dimensional. Now observe that $Tu + iTv = T_{\mathbb{C}}(u + iv) = (a + ib)(u + iv) = (au - bv) + i(bu + av)$, whence $Tu = au - bv$ and $Tv = bu + av$. Hence $\mathsf{span}\{u, v\}$ is invariant and $[T|_{\mathsf{span}\{u,v\}}]_{\{v,u\}} = \left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]$.

- Suppose $T_{\mathbb{C}}$ is diagonalizable. Denote the eigenvalues of $T_{\mathbb{C}}$ listed with multiplicity by $a_1 + ib_1, \dots, a_m + ib_m, a_1 - ib_1, \dots, a_m - ib_m, a_{m+1}, \dots, a_n$ where $a_{m+1} \leqslant \dots \leqslant a_n$ and $b_j > 0$. Then there is a basis $\mathcal{B}$ for V such that $[T]_{\mathcal{B}}$ is block diagonal, with m blocks of the form $\left[\begin{smallmatrix} a_j & -b_j \\ b_j & a_j \end{smallmatrix}\right]$ and $n - 2m$ blocks of the form $[a_j]_{1\times 1}$.

  How to get $\mathcal{B}$? Let $z_j = u_j + iv_j$ denote the elements of the original eigenbasis. Replace z_{m+j} by $u_j - iv_j$ for $j = 1, \dots, m$. For each string of identical eigenvalues $a_k, \dots, a_{k+d}$, choose $d + 1$ linearly independent vectors $x_k, \dots, x_{k+d}$ from $u_k, v_k, \dots, u_{k+d}, v_{k+d}$, and replace them in the list. The new collection has n eigenvectors (why?) and spans all the eigenspaces (why?), hence it spans $V_{\mathbb{C}}$, so it is an eigenbasis. And $[T]_{\mathcal{B}}$ has the claimed form if $\mathcal{B} = v_1, u_1, \dots, v_m, u_m, x_{2m+1}, x_n$.

  Remark. When T is normal, the orthonormal basis of eigenvalues provided by the Complex Spectral Theorem will provide an orthogonal basis whose existence we proved in the previous lecture. Gram-Schmidt might be needed for the real part.

  Remark. If $\mathcal{A}$ is a Jordan basis for $T_{\mathbb{C}}$, a similar procedure will give a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ has blocks in the diagonal as above, some 1's above the real eigenvalues, and some $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ blocks above the $\left[\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right]$ blocks.

Part V: Trace and determinant

For this part, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $\mathsf{dim}\, V = n \geqslant 1$. For convenience, $T_{\mathbb{C}}$ and $V_{\mathbb{C}}$ mean T and V if $\mathbb{F} = \mathbb{C}$, or their complexification if $\mathbb{F} = \mathbb{R}$.

21 Trace

Main reference: Axler §10.A

- We define the *trace of an operator* $T \in \mathcal{L}(V)$, denoted $\mathsf{trace}\, T$, as the sum of all n eigenvalues of $T_{\mathbb{C}}$. We define the *trace of a matrix* $A = (a_{jk})_{jk} \in \mathbb{C}^{n\times n}$ by $\mathsf{trace}\, A = \sum_{k=1}^{n} a_{kk}$, which is the sum of its n diagonal entries.
- If $p_T(z) = \sum_j a_j z^j$ is the characteristic polynomial of T, then $\mathsf{trace}\, T = -a_{n-1}$.
- *Theorem.* For every basis $\mathcal{B}$, $\mathsf{trace}[T]_{\mathcal{B}} = \mathsf{trace}\, T$. In particular, $\mathsf{trace}[T]_{\mathcal{B}}$ does not depend on the choice of basis.
- To prove the previous theorem, we recall the following facts:

  * There is a basis $\mathcal{B}$ of $V_{\mathbb{C}}$ such that $[T_{\mathbb{C}}]_{\mathcal{B}}$ is triangular superior.

  * For every basis $\mathcal{B}$ of $V_{\mathbb{C}}$, if $[T_{\mathbb{C}}]_{\mathcal{B}}$ is triangular superior then its diagonal elements coincide with the n eigenvalues of $T_{\mathbb{C}}$.

  And we use the following property of matrix trace:

  * If A is invertible, then $\mathsf{trace}(ABA^{-1}) = \mathsf{trace}(B)$.

  The last property implies that $\mathsf{trace}[T]_{\mathcal{B}}$ does not depend on the basis $\mathcal{B}$. Hence, it is enough to show that the definitions of trace of operator and trace of matrix coincide for *some* basis $\mathcal{B}$. The first property ensures the existence of a convenient basis, which together with the second property shows that this basis gives the desired equality, proving the theorem.
- Proof that $\mathsf{trace}(ABA^{-1}) = \mathsf{trace}(B)$.

  This follows from another property: $\mathsf{trace}(AB) = \mathsf{trace}(BA)$, given by

  $$\mathsf{trace}\, AB = \sum_j (AB)_{jj} = \sum_j \sum_k a_{jk} b_{kj} = \sum_k \sum_j b_{kj} a_{jk} = \sum_k (BA)_{kk} = \mathsf{trace}\, BA.$$
- Interesting property: Although there is no relationship between the eigenvalues of $S + T$ with sums of eigenvalues of S and of T, when added together we have that the sum of all eigenvalues of $S + T$ equals the sum of all eigenvalues of S and T.
- Interesting application: There are no operators S and T such that $ST - TS = I$.

22 Determinant

Main reference: Axler §10.B

- We define the *determinant of an operator* $T \in \mathcal{L}(V)$, denoted $\mathsf{det}\, T$, as the product of all n eigenvalues of $T_{\mathbb{C}}$. It equals $(-1)^n a_0$ in the characteristic polynomial.
- The operator T is invertible if and only if $\mathsf{det}\, T \neq 0$. Also, $p_T(z) = \mathsf{det}(zI - T)$.
- We define the *determinant of a matrix* $A = (a_{jk})_{jk} \in \mathbb{C}^{n \times n}$ by

$$\mathsf{det}\, A = \sum_{\sigma} \mathsf{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j},$$

where the sum is over all the $n!$ permutations σ of $\{1, \ldots, n\}$, and

$$\mathsf{sgn}(\sigma) = (-1)^{\#\{(j,k)\,:\, j<k \text{ and } \sigma(j)>\sigma(k)\}}.$$

- $\mathsf{det}(AB) = (\mathsf{det}\, A)(\mathsf{det}\, B)$
- For every basis $\mathcal{B}$ of V, we have $\mathsf{det}[T]_{\mathcal{B}} = \mathsf{det}\, T$.

  Proof. Since $\mathsf{det}\, I = 1$, by the previous property $\mathsf{det}[T]_{\mathcal{B}}$ does not depend on $\mathcal{B}$, so it suffices to show the identity for *some* basis of $V_{\mathbb{C}}$. Take a basis $\mathcal{B}$ of $T_{\mathbb{C}}$ for which $[T_{\mathbb{C}}]_{\mathcal{B}}$ is upper-triangular, so the n elements in the diagonal are the n eigenvalues of $T_{\mathbb{C}}$. Finally, notice that the determinant of an upper-triangular matrix is the product of the elements in the diagonal, because other permutations always pick an entry below it.

◦ For $\Omega \subset \mathbb{R}^n$ open and bounded, $\mathsf{vol}(T(\Omega)) = |\,\mathsf{det}\, T| \times \mathsf{vol}(\Omega)$.

The volume is defined as

$$\mathsf{vol}(\Omega) = \inf_{\mathcal{A}} \sum_{k} \mathsf{vol}(A_k).$$

The infimum is taken over all sequences of boxes $\mathcal{A} = (A_1, A_2, \ldots)$ whose faces are orthogonal to the axes and whose union contains Ω. The volume of a box (A_k) is defined as the product of its n dimensions (width, depth, height, etc.).

The proof uses the following facts:

- A diagonal operator D stretches volume by a factor $|\,\mathsf{det}\, D|$. (exercise)
- Isometries preserve volume. (accepted without proof)
- An isometry S has $|\,\mathsf{det}\, S| = 1$. (exercise)
- $\mathsf{det}(RT) = (\mathsf{det}\, R)(\mathsf{det}\, T)$. (exercise)
- Singular Value Decomposition: $T = S_1 S_2 D S_3$. (write $P = S_2 D S_3$)

Proof. $\mathsf{vol}(T(\Omega)) = \mathsf{vol}(S_1 S_2 D S_3(\Omega)) = \mathsf{vol}(D S_3(\Omega)) = |\,\mathsf{det}\, D| \times \mathsf{vol}(S_3(\Omega)) = |\,\mathsf{det}\, D| \times \mathsf{vol}(\Omega)$. On the other hand, $|\,\mathsf{det}\, T| = |\,\mathsf{det}\, D|$, concluding the proof.