April 16, 2026*

These notes refer to the numbering in the fifth edition.

The numbers of statements/pages/etc. in the textbook are for now marked with a †; e.g., Theorem 5.2 (5.20†) below.

NOTE. Unless mentioned otherwise, $T : V \to V$ is linear and V is finite dimensional.

Notation 0.1 The identity matrix is denoted by I and – if not specified – has the size that fits. The identity transformation is denoted by Id, sometimes without the space on which it acts.

We denote the standard bases by *st*.

To make it more precise, we denote the change of basis matrix (Q in the book, see Section 2.5†) as follows:

- let α and β be two (ordered) bases of V
- for $x \in V$, denote its coordinates w.r.t. α by $[x]^\alpha$ (the book uses a lower index, but that does not work so nicely with the rest)
- $Q^\beta_\alpha := [\mathrm{Id}_V]^\beta_\alpha$ is the representation of the identity transformation on V with respect to the initial basis α and final basis β (so column j is the representation in β of the j-th vector of α)
- the same notation, $[T]^\beta_\alpha$, will be used for the matrix of a linear transformation $T : V \to W$ w.r.t. the ordered bases $\alpha \subset V$, $\beta \subset W$. That is: column j of $[T]^\beta_\alpha$ is the representation in β of the image through T of the j-th vector in α.
- then $[x]^\beta = Q^\beta_\alpha [x]^\alpha$, more generally $[Tx]^\beta = [T]^\beta_\alpha [x]^\alpha$, and $Q^\gamma_\alpha = Q^\gamma_\beta Q^\beta_\alpha$ since $[TS]^\gamma_\alpha = [T]^\gamma_\beta [S]^\beta_\alpha$ for composition of linear maps (see details in the book).

---

[*] If you find typos/errors/omissions/etc. please let me know, will correct them for a subsequent edition.

Contents

Tentative syllabus:

The topics indicated with a * may be covered if time permits.

5.1 Eigenvalues and Eigenvectors (Review)
5.2 Diagonalizability (Review)
5.3* Matrix Limits and Markov Chains
5.4 Invariant Subspaces and the Cayley–Hamilton Theorem

6.1 Inner Products and Norms
6.2 The Gram–Schmidt Orthogonalization Process and Orthogonal Complements
6.3 The Adjoint of a Linear Operator
6.4 Normal and Self-Adjoint Operators
6.5 Unitary and Orthogonal Operators and Their Matrices
6.6 Orthogonal Projections and the Spectral Theorem
6.7* The Singular Value Decomposition and the Pseudoinverse
6.8* Bilinear and Quadratic Forms
6.9* Einstein's Special Theory of Relativity
6.10* Conditioning and the Rayleigh Quotient
6.11* The Geometry of Orthogonal Operators

7.1 The Jordan Canonical Form I
7.2 The Jordan Canonical Form II
7.3 The Minimal Polynomial
7.4* The Rational Canonical Form

5 Chapter $5^\dagger$

5.1 Section $5.1^\dagger$: Eigenvalues and Eigenvectors

Characteristic polynomial

For a square matrix A, its ***characteristic polynomial*** is $\mathrm{char}_A(\lambda) := \det(A - \lambda\,\mathrm{I})$ (denoted by f_A in the textbook).

For $T : V \to V$ a linear transformation (on a finite dimensional vector space), compute its characteristic polynomial by $\mathrm{char}_T(\lambda) := \det([T]_\beta^\beta - \lambda\,\mathrm{I})$ where β is any ordered basis of V (the results does not depend on β).

Theorem 5.1 (Definitions, Theorems $5.1^\dagger$, $5.2^\dagger$, etc.)

- *The value* $\lambda \in \mathbb{F}$ *is an* **eigenvalue** *of the linear transformation* $T : V \to V$ *if there is* $\boldsymbol{v} \in V$ *such that*
  - $\boldsymbol{v} \neq \boldsymbol{0}$
  - $T\boldsymbol{v} = \lambda \boldsymbol{v} \iff (T - \lambda\,\mathrm{Id}_V)\boldsymbol{v} = \boldsymbol{0}$

  Such a vector $\boldsymbol{v}$ *is an* **eigenvector** *corresponding to the eigenvalue* λ.

- *The set* $E_\lambda := \{\boldsymbol{v} \in V \mid (T - \lambda\,\mathrm{Id}_V)\boldsymbol{v} = \boldsymbol{0}\} = \mathrm{N}(T - \lambda\,\mathrm{Id}_V)$ *is a vector subspace of* V*, called the* **eigenvspace** *of* T *corresponding to the eigenvalue* λ*; we need a* **basis** *of it, that is, a* **basis of eigenvectors** *for the eigenvalue* λ.

- *Assuming* V *is finite dimensional:*
  - λ *is an eigenvalue of* $T \iff \mathrm{char}_T(\lambda) = 0$
  - char_T *is a polynomial of degree* $\dim_{\mathbb{F}} V$*, with leading coefficient* $(-1)^{\dim_{\mathbb{F}} V}$
    - *therefore,* char_T *has at most* $\dim_{\mathbb{F}} V$ *roots, counting multiplicities*
    - *by the Fundamental Theorem of Algebra: a polynomial* $p(x)$ *with complex coefficients has exactly* $\mathrm{degree}(p)$ *complex*[1] *roots, counting multiplicities*
    - *for a polynomial with real coefficients: complex roots come in conjugate pairs,* $a \pm ib$*, and the same multiplicity*
  - *the matrix of* T *in the basis* β *of* V *is diagonal* $\iff$ β *consists of eigenvectors for* T
  - *such a basis exists* $\iff$ T *is called* **diagonalizable**

---

[1] The field of complex numbers is denoted by $\mathbb{C}$.

Similarity, change of basis

Two $n \times n$ matrices A and B are ***similar*** if $A = QBQ^{-1}$ for an invertible matrix Q (which has to have the same size); see page $116^{\dagger}$. Similarity is an equivalence relation (symmetric, reflexive, transitive).

Assume $\beta, \gamma \subset V$ are (ordered) bases. If $T : V \to V$ is linear, then we saw that

$$[T]_{\beta}^{\beta} = [\mathrm{Id}_V]_{\beta}^{\gamma}[T]_{\gamma}^{\gamma}[\mathrm{Id}_V]_{\beta}^{\gamma}$$

and the change of basis matrices $Q_{\beta}^{\gamma} := [\mathrm{Id}_V]_{\beta}^{\gamma}$ and $Q_{\gamma}^{\beta} := [\mathrm{Id}_V]_{\gamma}^{\beta}$ are inverses to each other. Therefore $[T]_{\gamma}^{\gamma}$ and $[T]_{\beta}^{\beta}$ are similar matrices.

In the case of an $n \times n$ matrix A, A is the matrix of the linear transformation determined by A on $\mathbb{F}^n$ (denoted by L_A in the book) w.r.t. the standard basis (check this). If β is another basis, then Q_{β}^{st} is easy to compute (why?) and

$$[A]_{\beta}^{\beta} = \left(Q_{\beta}^{st}\right)^{-1} A \, Q_{\beta}^{st} \quad \text{or, equivalently} \quad A = Q_{\beta}^{st} \, [A]_{\beta}^{\beta} \, \left(Q_{\beta}^{st}\right)^{-1}$$

5.2 Section 5.2†: Diagonalizability

Idea: *The central result (from which most of the rest follows) is that*

$$\sum_{\lambda} E_{\lambda} = \oplus_{\lambda} E_{\lambda}$$

where the sum is over the **distinct** *eigenvalues λ of $T : V \to V$ (take each eigenvalue once, even if it is has a higher multiplicity). Here $E_{\lambda} := \mathrm{N}(T - \lambda\,\mathrm{Id})$, the eigenspace corresponding to λ. This follows from Theorem 5.5†. See Theorem 5.9† for details about direct sums.*[2]

Here $E_{\lambda} := N(T - \lambda\,\mathrm{Id})$, the eigenspace corresponding to λ; the dimension of E_{λ} is called the geometric multiplicity of λ.

The LHS above is the vector subspace spanned by **all the eigenvectors***; for the transformation to be diagonalizable, this sum must be equal to V; because the sum is actually a direct sum, its dimension is equal to the sum of the dimensions of E_{λ}.*

Definition 5.1 Assume $T : V \to V$ with $\dim V = n$, or A an $n \times n$ matrix.

- the ***algebraic multiplicity*** is the multiplicity of λ as a root of the characteristic polynomial;
- the ***geometric multiplicity*** is the dimension of the eigenspace corresponding to λ, $\dim E_{\lambda} := N(A - \lambda\,\mathrm{I})$; the latter is equal to $\dim V - \mathrm{rank}\,(T - \lambda\,\mathrm{Id})$ or $n - \mathrm{rank}\,(A - \lambda\,\mathrm{I})$.

Theorem 5.2 (Theorem 5.7†) *For any λ:*

$$\textit{geometric multiplicity of } \lambda \quad \leq \quad \textit{algebraic multiplicity of } \lambda.$$

Note that if λ is an eigenvalue, then its geometric multiplicity is at least 1 (because E_{λ} is not the zero subspace).

- A matrix can be diagonalizable even if it has repeated eigenvalues (e.g., the identity matrix).

---

[2] This was Theorem 5.10 in the 4th edition.

- An example where geometric multiplicity is less than the algebraic multiplicity:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

  - $\det(A - \lambda \mathrm{I}) = (1-\lambda)^3$, so 1 has algebraic multiplicity 3.
  - Because $\operatorname{rank}(A - \mathrm{I}) = 1$ (check that), the geometric multiplicity of 1 is 3-1=2; thus there are only two linearly independent solutions to $(A-\mathrm{I})\boldsymbol{v} = \mathbf{0}$. **Therefore A is not diagonalizable.**
  - Two such vectors are $\boldsymbol{e}_1$ and $\boldsymbol{e}_3$, where $\{\boldsymbol{e}_k\}$ is the canonical basis of $\mathbb{R}^3$ (check this too).

- Another example:

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & -2 & -2 \\ 0 & 0 & 2 \end{pmatrix}$$

  - $\det(A - \lambda \mathrm{I}) = -(\lambda-2)^2(\lambda+3)$, so $\lambda_{1,2} = 2$ has algebraic multiplicity 2, $\lambda_3 = -3$ has algebraic multiplicity 1.
  - Compute eigenvectors (or, at least $\operatorname{rank}(A - 2\,\mathrm{I})$ to decide what the geometric multiplicity of $\lambda = 2$ is (recall: geometric multiplicity = nullity = $n-$ rank); after row reduction, we see that this rank is 1, so geometric multiplicity is 3-1=2.

    Conclusion: the matrix is diagonalizable.
  - To find a basis of eigenvectors for $\lambda = 2$, need further computations. We get

$$(A - 2\,\mathrm{I})\boldsymbol{v} = \mathbf{0} \implies \boldsymbol{v} = y \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

    so can take the above two vectors as linearly independent eigenvectors for $\lambda = 2$ (they form a basis of E_2).
  - Do the same for $\lambda = -3$ (we know that its geometric multiplicity is one), get after row reduction

$$\boldsymbol{v}_3 \in \left\{ y \begin{pmatrix} -1/2 \\ 1 \\ 0 \end{pmatrix} \mid y \in \mathbb{R} \right\}$$

    so can take

$$\boldsymbol{v}_3 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}$$

- One more example:
  - Say A is 5×5 with coefficients in $\mathbb{R}$ and has three (real of complex) distinct eigenvalues, λ_1, λ_2, λ_3, with algebraic multiplicities 1, 3, and 1.
  - to decide diagonalizability, have only to compute the geometric multiplicity of λ_2 (for λ_1 and λ_3 the geometric multiplicity is 1).
  - if the geometric multiplicity of λ_2 is 3, then A is diagonalizable, otherwise (that is, if it less than 3), then A is not diagonalizable.

5.4 Section 5.4†: Invariant subspaces; the Cayley-Hamilton Theorem

Invariant subspaces

Definition 5.1 A vector subspace $W \subset V$ is ***T-invariant*** if $T(W) \subset W$.

> **Idea:** *Invariant subspaces allow to "break-up" V into "smaller" spaces.*
>
> *This is good b/c computations are now done on "smaller" matrices. E.g., say a linear transformation is on a 20-dimensional space, but there is a vector which generates a 5-dimensional cyclic subspace; then the behavior of the transformation on this subspace requires only a 5×5 matrix.*

Example 1. Some T-invariant subspaces:

- $\{\mathbf{0}\}$, V are the trivial invariant subspaces
- $\mathrm{N}(T)$, $\mathrm{R}(T)$, E_λ for any eigenvalue λ

Example 2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(a, b, c) = (a, a + b, b + c)$
Then the yz-plane is T-invariant. The z-axis is also T-invariant.

NOTE. If W is a T-invariant subspace then:

- can consider the restriction $T_W : W \to W$ of T to W; T_W is also linear (another common notation for T_W is $T|_W$, with "$|$" denoting "restriction")
- the matrix of T with respect to an ordered basis β of V that extends a basis β_W of W has lower left corner equal to zero (this is equivalent to W being invariant):

$$[T]_\beta = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix} \text{ with } B_1 = [T_W]_{\beta_W}$$

Theorem 5.2 (5.20†) *If $W \subset V$ is T-invariant then the characteristic polynomial of T_W divides the characteristic polynomial of T.*

Proof: Use a basis as above to get a "block upper-triangular" matrix for $[T]_\beta - t\,\mathrm{I}$. Then compute the determinant using the Claim below. ∎

Claim 5.3 *(Exercise 21, §4.3†)*
For square matrices A_1 and A_3,

$$\det \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix} = \det A_1 \det A_3$$

To prove the Claim, write

$$\begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix} = \begin{pmatrix} \mathrm{I} & 0 \\ 0 & A_3 \end{pmatrix} \begin{pmatrix} \mathrm{I} & A_2 \\ 0 & \mathrm{I} \end{pmatrix} \begin{pmatrix} A_1 & 0 \\ 0 & I \end{pmatrix}$$

and compute the determinant of each of the three factors: for the first expand successively (or use induction) along the rows at the top, for the third do the same along the bottom rows; the middle one is upper-triangular, so its determinant is the product of the diagonal entries, therefore 1. Now use the fact that $\det(AB) = \det(A)\det(B)$ for square matrices A and B (see Chapter 4†).

Cyclic subspaces

Definition 5.4 The T***-cyclic subspace generated by*** $v \in V$ is

$$W = \operatorname{span}\{v, Tv, T^2v, T^3v, \dots\}$$

NOTE.

- any cyclic subspace is T-invariant
- the T-cyclic subspace generated by v is the smallest T-invariant subspace that contains v

Theorem 5.5 (5.21†) *Assume* $T : V \to V$ *is linear,* V *finite dimensional,* $v \neq 0$ *a vector in* V*,* W *the cyclic subspace generated by* v*.*

If $\dim W = k$ *then*

(a) $\beta_v := \{v, Tv, \dots, T^{k-1}v\}$ *is a basis of* W

(b) Thus there are (unique) scalars such that $a_0v + a_1Tv + \dots a_{k-1}T^{k-1}v + T^kv = \mathbf{0}$*. Then the characteristic polynomial of* T_W *is*

$$\operatorname{char}_{T_W}(t) = (-1)^k(a_0 + a_1t + \dots a_{k-1}t^{k-1} + t^k)$$

Proof:

(a) Take the largest j such that $\beta := \{v, Tv, \dots, T^{j-1}v\}$ is linearly independent (since $v \neq \mathbf{0}$, $j \geq 1$). Thus $T^jv \in \operatorname{span}(\beta)$ – otherwise could have increased j, so $\operatorname{span}(\beta)$ is T-invariant (because $T\beta \subset \operatorname{span}(\beta)$), so $\operatorname{span}(\beta) = \operatorname{span}(\{v, Tv, T^2v, \dots\}) = W$, the cyclic subspace generated by v. In particular $j = k$ and $\beta = \beta_v$ is a basis of W.

(b) Write the matrix of T_W in the basis β_v and use this to compute the characteristic polynomial of T_W. To do that, e.g. expand along the first row and use induction. Alternatively, by row operations, eliminate the $-t$'s on the diagonal starting with the one on the next-to-last row and moving upward; at the end expand along the first row.

∎

The Cayley-Hamilton Theorem

Theorem 5.6 (5.22†) *If* $\dim V < \infty$ *then* $\operatorname{char}_T(T) = 0 : V \to V$ *(that is, the characteristic polynomial of* T *computed at* T *gives the zero transformation).*

Preliminaries. See Appendix E, from page 565. This can be called "functional calculus with polynomials".

- For $p(x) = a_0 + a_1x + \dots a_nx^n$ a polynomial with coefficients in $\mathbb{F}$, define $p(T)$ to be

  $$p(T) := a_0\,\mathrm{I} + a_1T + \dots a_nT^n,$$

  which is a linear map from V to V.
- If $p(x) = q(x)r(x)$ then $p(T) = q(T)r(T)$, where in the first equality we have product of polynomials, in the second composition of linear maps.

- $q(T)r(T) = r(T)q(T)$
- This actually gives a ring homomorphism $p \in \mathcal{P}(\mathbb{F}) \mapsto p(T) \in \mathcal{L}(V, V)$, from where the "functional calculus" name.

Proof of the Cayley-Hamilton Theorem:

- Enough to check that $\mathrm{char}_T(T)|_v = \mathbf{0}$ for any $v \in V$. [3]
- For $v \neq \mathbf{0}$ let W be the T-cyclic subspace it generates. If $\dim W = k$, there are scalars such that
$$a_0 v + a_1 T v + \ldots a_{k-1} T^{k-1} v + T^k v = \mathbf{0}$$
By Theorem 5.5(5.21†), the LHS is exactly $\mathrm{char}_{T_W}(T)$ computed at v.
- Since $\mathrm{char}_T(t) = g(t)\,\mathrm{char}_{T_W}(t)$, then (using "|" to denote "evaluated at")
$$\mathrm{char}_T(T)|_v = g(T)\,\mathrm{char}_{T_W}(T)|_v = g(T)|_{\mathbf{0}} = \mathbf{0}$$

∎

Remark 5.7 Why Cayley-Hamilton is reasonable: assume A is a diagonal matrix, say

$$A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

Then $\mathrm{char}_A(t) = (a - t)(b - t)$ and it is easy to see that $(a - A)(b - A)$ is the zero matrix.

Note however that ***not all matrices are diagonalizable***; the "best" that happens is the Jordan Canonical Form (see Chapter 7). The proof we gave does not rely on the matrix of T having a particular form.

---

[3] $f|_x$ denotes f evaluated at x; use this when there would be too many parenthesis.

6 Chapter 6†: Inner product spaces

NOTE. ***In this chapter the field $\mathbb{F}$ of scalars is either $\mathbb{R}$ or $\mathbb{C}$.*** It is more convenient to work over $\mathbb{C}$ because, as we saw earlier, over the complex numbers all polynomials split, so we always have enough eigenvalues.

> **Idea:** *By endowing a vector space with the additional structure of* **inner product**, *one can talk about length of vectors, orthogonality, angles (if the scalars are $\mathbb{R}$), and orthogonal projections.*
>
> *A particular case of inner product is the dot product on $\mathbb{R}^n$, see the examples in §6.1†.*
>
> *We will prove that certain linear transformations have a nice description (the Spectral Theorem), which is a restatement of Thm. 6.16† in §6.4†:*
>
> ***Theorem A.*** *If $T : V \to V$ is a* **normal** *linear transformation on a finite dimensional inner product space over $\mathbb{C}$, then T has an* **orthonormal basis of eigenvectors**.
>
> **Theorem A is the main result of this chapter.**
>
> *Using that*
>
> ***Theorem B.*** *The eigenvalues of a self-adjoint matrix are all real (consider the matrix as a complex matrix, so it has eigenvalues in $\mathbb{C}$, to start with).*
>
> *we easily obtain that*
>
> ***Theorem C.*** *If $T : V \to V$ is a* **self-adjoint** *linear transformation on a finite dimensional inner product space over $\mathbb{R}$ or $\mathbb{C}$, then T has an* **orthonormal basis of eigenvectors**, *and all eigenvalues are real.*
>
> *[The emphasis in Theorem C is the case of transformations over real vector spaces, which are not covered by Theorem A.]*

6.1 Section 6.1†: Inner products and norms

Definition 6.1 (inner product) See the definition on page 327†.

Properties (a) and (b) say that the inner product is linear in the first variable.

Examples 6.2 See Examples 1, 3, 5 on page 328-329†.

Definition 6.3 (page 329†) The ***adjoint*** A^* of a matrix A is the matrix obtained by transposing A and complex conjugating each entry.

[The complex conjugation only matters if the scalars are $\mathbb{C}$; over $\mathbb{R}$, $A^* = A^t$.]

Recall that the trace of a square matrix is the sum of its diagonal entries.

Properties: $A \mapsto \operatorname{tr}(A)$ is linear, $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for A and B rectangular matrices of the proper size.

NOTE. In this chapter the ***standard inner product*** of $\mathbb{F}^n$ and $\mathrm{Mat}_{n\times n}(\mathbb{F})$ are those described above.

Theorem 6.4 ($6.1^\dagger$) *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space.*

- *(a), (b) the inner product is* **conjugate** *linear in the second variable (only makes a difference if $\mathbb{F} = \mathbb{C}$)*
- *(c)* $\langle x, \mathbf{0}\rangle = \langle \mathbf{0}, x\rangle = 0$
- *(d)* $\langle x, x\rangle \geq 0$ *for all x, and is zero iff $x = 0$.*
- *(e)* $\langle x, y\rangle = \langle x, z\rangle$ *for all $x \in V \implies y = z$*

Proof:

- (a), (b), (c), (d) follow from the definition of the inner product
- for (e), write $\langle x, y\rangle = \langle x, z\rangle \iff \langle x, y - z\rangle = 0$, take $x = y - z$ and use (d) above.

∎

Definition 6.5 (page $331^\dagger$) The ***norm*** or ***length*** of a vector in $(V, \langle\cdot,\cdot\rangle)$ is

$$\|x\| = \sqrt{\langle x, x\rangle}$$

Example 6.6 (#6, page $331^\dagger$) For $\mathbb{R}^n$ with the standard inner product, this is just the Euclidean length of a vector.

Theorem 6.7 ($6.2^\dagger$) *Let $(V, \langle\cdot,\cdot\rangle)$, $x, y \in V$, $c \in \mathbb{F}$. Then*

- *(a)* $\|cx\| = |c|\|x\|$
- *(b)* $\|x\| \geq 0$*, with equality iff $x = \mathbf{0}$*
- *(c)* **Cauchy-Schwarz***:* $|\langle x, y\rangle| \leq \|x\|\|y\|$
- *(d)* **the triangle inequality***:* $\|x + y\| \leq \|x\| + \|y\|$

Moreover: equality holds in Cauchy-Schwarz iff the vectors are proportional, in the triangle inequality iff they are positively proportional. To see these, follow the proofs.

Proof:

- (a) This is a direct computation.
- (b) Follows from Theorem 6.4, (d) as written in these notes.
- (c) **Idea:** *we know that $\langle x - cy, x - cy\rangle \geq 0$ for all $c \in \mathbb{F}$; if we expand, we get a "quadratic function" in c that is always ≥ 0; pick c that minimizes this function.*

  If we do the computations, we conclude that, unless $y = \mathbf{0}$ (when the inequality is clear) $c = \langle x, y\rangle / \langle y, y\rangle$ is the right value.

(d) Expand $\|x+y\|^2$, and use Cauchy-Schwarz for $\mathrm{Re}\langle x,y\rangle \leq |\langle x,y\rangle| \leq \|x\|\|y\|$ [4]

∎

Remark 6.8 For many applications, a norm on V (that is, a function $\|\cdot\| : V \to \mathbb{R}$ that satisfies (a), (b) and (d) in the previous Theorem) suffices. We see that one can obtain such a function from an inner product.

6.2 Section 6.2†: The Gram–Schmidt Orthogonalization Process and Orthogonal Complements

See Exercise 13† for more properties of the orthogonal complement. Part (c) of Exercise 13 (that $(W^\perp)^\perp = W$) follows from Thm. 6.7 if $\dim V$ is finite.

Exercise 23*† shows that $(W^\perp)^\perp = W$ need not hold if $\dim W$ is infinite.

Theorem 6.9 (follows from Thm. 6.7†) *Let V be a finite dimensional inner product vector space. If $W \subset V$ is a subspace, then $W \oplus W^\perp = V$.*

Remark 6.10 See also Exercise 6.2.13†. The above Theorem remains true for V infinite dimensional if W has finite dimension. Exercise 6.2.23† shows that can have $(W^\perp)^\perp \neq W$ if W is infinite dimensional, so the Theorem need not hold in that case.

6.3 Section 6.3†: The Adjoint of a Linear Operator

We also proved the following structure result:

Theorem 6.11 (Theorem not in the book) *Let $T : (V, \langle\cdot,\cdot\rangle_V) \to (W, \langle\cdot,\cdot\rangle_W)$ be a linear map between finite-dimensional inner product spaces. Then*

(a) $R(T)^\perp = N(T^*)$ $\quad$ ($\iff$ $R(T) = N(T^*)^\perp$ *because W has finite dimension*) *and therefore*

$$W = R(T) \oplus N(T^*) \quad \textit{(orthogonal direct sum)}$$

Do the same for $T^* : W \to V$ to get

$$V = R(T^*) \oplus N(T) \quad \textit{(orthogonal direct sum)}$$

(b) $R(T^*T) = R(T^*)$ *and* $N(T^*T) = N(T)$.

Therefore, $\mathrm{rank}(T^*T) = \mathrm{rank}(T^*) = \mathrm{rank}(T)$, *and thus (applying this to T^*) T restricted to $R(T^*)$ is a linear bijection from $R(T^*) \subset V$ to $R(T) \subset W$.*

(c) If $T : V \to V$ and $W \subset V$ is a subspace:

$$T(W) \subset W \implies T^*(W^\perp) \subset W^\perp$$

That is, if W is T-invariant, then $W^\perp$ is T^*-invariant.

---

[4] $\mathrm{Re}\, z$ stands for the real part of the complex number z.

Proof: (a) Will prove that $R(T)^\perp = N(T^*)$:

$$\boldsymbol{v} \in R(T)^\perp \iff \langle \boldsymbol{v}, T(\boldsymbol{x})\rangle = 0 \quad \forall\, \boldsymbol{x} \in V \iff \langle T^*(\boldsymbol{v}), \boldsymbol{x}\rangle = 0 \quad \forall\, \boldsymbol{x} \in V \iff T^*(\boldsymbol{v}) = \boldsymbol{0}$$

(b) It is immediate that $R(T^*T) \subset R(T^*)$ and $N(T) \subset N(T^*T)$.

We check first that $N(T^*T) \subset N(T)$. Let $\boldsymbol{v} \in N(T^*T) \iff T^*T\boldsymbol{v} = \boldsymbol{0}$. Then $\langle T^*T\boldsymbol{v}, \boldsymbol{v}\rangle = 0$, and $\langle T^*T\boldsymbol{v}, \boldsymbol{v}\rangle = \langle T\boldsymbol{v}, T\boldsymbol{v}\rangle = \|T\boldsymbol{v}\|^2$, so conclude that $T\boldsymbol{v} = \boldsymbol{0}$, as desired.

This implies the claim about the ranges. Since $(T^*T)^* = T^*T$, using (a) twice and the relation for null-spaces: $R(T^*T) = N((T^*T)^*)^\perp = N(T^*T)^\perp = N(T)^\perp = R(T^*)$.

(c) Take $\boldsymbol{v} \in W^\perp$, want to show that $T^*(\boldsymbol{v}) \perp W$ (so is also in $W^\perp$). Indeed, for any $\boldsymbol{w} \in W$:

$$\langle T^*\boldsymbol{v}, \boldsymbol{w}\rangle = \langle \boldsymbol{v}, T\boldsymbol{w}\rangle = 0$$

because $T\boldsymbol{w} \in W$ and $\boldsymbol{v} \perp W$. ∎

The first part above gives a geometric explanation to Thm's $6.12^\dagger$ and $6.13^\dagger$ about "best approximation" and "minimal solution".

See Exercises 12 and 13 for more about these.

Theorem 6.12 (Theorem $6.12^\dagger$, Least Squares Approximation) *Let $A \in \mathrm{Mat}_{m\times n}(\mathbb{F})$ and $\boldsymbol{y} \in \mathbb{F}^m$.*

There exists $\boldsymbol{x}_0 \in \mathbb{F}^n$ such that $A^*A\boldsymbol{x}_0 = A^*\boldsymbol{y}$; the point $A\boldsymbol{x}_0$ is the (unique) **point in** $\mathrm{R}(A)$ **closest to** $\boldsymbol{y}$ *(that is, $\|A\boldsymbol{x}_0 - \boldsymbol{y}\| \le \|A\boldsymbol{x} - \boldsymbol{y}\|$ for all $\boldsymbol{x} \in \mathbb{F}^n$, so $A\boldsymbol{x}_0$ is the orthogonal projection of $\boldsymbol{y}$ onto the range of A).*

Furthermore, if $\mathrm{rank}(A) = n$ *then $\boldsymbol{x}_0 = (A^*A)^{-1}A^*\boldsymbol{y}$.*

Note that $\mathrm{rank}(A) = n \iff A$ *is 1-1* $\iff A^*A$ *is 1-1 (see Theorem 6.11 (b)).*

Theorem 6.13 (Theorem $6.13^\dagger$, Minimal Solution) *Let $A \in \mathrm{Mat}_{m\times n}(\mathbb{F})$ and $\boldsymbol{b} \in \mathbb{F}^m$ such that $A\boldsymbol{x} = \boldsymbol{b}$ has a solution (that is, $\boldsymbol{b} \in \mathrm{R}(A)$).*

Then:

- *(a) There exists exactly one* **minimal solution** *(that is, of shortest length) $\boldsymbol{s}$ to $A\boldsymbol{x} = \boldsymbol{b}$; this $\boldsymbol{s}$ is in $\mathrm{R}(A^*)$.*
- *(b) To find $\boldsymbol{s}$: solve $AA^*\boldsymbol{u} = \boldsymbol{b}$, then $\boldsymbol{s} = A^*\boldsymbol{u}$. [Note that $\boldsymbol{u}$ need not be unique, but any such $\boldsymbol{u}$ gives the same $\boldsymbol{s}$.]*

Remark 6.14 The "Least Squares Approximation" Theorem $6.12^\dagger$ gives the following method to compute the orthogonal projection onto a subspace $W \subset \mathbb{F}^m$ having basis $\beta = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ (an alternative to computing from β an orthogonal basis of W by Gram–Schmidt, etc.): consider the matrix A with columns the vectors of β; then $W = \mathrm{R}(A)$, $\mathrm{rank}(A) = n$, and the orthogonal projection of $\boldsymbol{y} \in \mathbb{F}^m$ to W is given by $A(A^*A)^{-1}A^*\boldsymbol{y}$.

This is because the "Least Squares Approximation" $A\boldsymbol{x}_0$ of $\boldsymbol{y}$ given by Theorem $6.12^\dagger$ is the orthogonal projection of $\boldsymbol{y}$ onto the range of A.

6.4 Section $6.4^\dagger$: Normal and self-adjoint operators

Theorem 6.15 (Thm $6.14^\dagger$, Schur) *Let $T : V \to V$ be a linear operator on a finite dimensional $\underline{\text{inner product}}$ vector space over the field $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Assume that* char_T *splits in $\mathbb{F}$ (satisfied if $\mathbb{F} = \mathbb{C}$).*

Then there is an $\underline{\text{orthonormal}}$ basis of V in which T has an upper-triangular matrix.

Remark 6.16 If T has an upper-triangular matrix in any basis, then its eigenvalues are the diagonal entries, so its characteristic polynomial splits; thus this is a <u>necessary</u> condition.

Remark 6.17 Schur's theorem holds for any a finite dimensional vector space V over $\mathbb{R}$ or $\mathbb{C}$, because can define an inner product on V (see Exercise 6.1.22†). Given any (finite or infinite dimensional) vector space V over $\mathbb{R}$ or $\mathbb{C}$, one can introduce an inner product as follows: take any basis $\beta = \{v_k\}$, and define $\langle \sum a_k v_k, \sum b_k v_k \rangle := \sum a_k \overline{b_k}$. [Check that this is an inner product, and β is now an orthonormal basis.]

Actually, Schur's theorem holds in an even more general setting, over any field (so $\mathbb{F}$ does not have to be $\mathbb{R}$ or $\mathbb{C}$), if one drops the requirement that the basis be orthonormal.

Theorem 6.18 (Schur restated) *If the characteristic polynomial of a linear map $T : V \to V$ on a finite dimensional vector space V over* **any field** $\mathbb{F}$ *splits in $\mathbb{F}$, then there is a basis of V in which T has an upper-triangular matrix.*

Proof: *The proof given below works in that case too, instead of taking W the orthonormal complement, take it to be any complement to $N(T - \lambda I)$.*

See Exercise 5.2.12† for another proof. ∎

Proof of Schur's Theorem 6.14†: See the book for one proof, using Exercise 5.2.12† and Gram–Schmidt.

<u>Another proof (more direct, based on a similar idea):</u>

Proceed by induction on $n = \dim V$.

$\dim V = 1 \implies$ OK

Assume the claim holds for $\dim V < n$, and let V be of dimension n.

Pick an eigenvalue λ of T (exists because char_T splits).

Decompose $V = N(T - \lambda I) \oplus W$ where $W = N(T - \lambda I)^\perp$, and write the "matrix" of T w.r.t. this decomposition (<u>check that this is indeed the form one obtains</u>):

$$T = \begin{pmatrix} \lambda \, \mathrm{Id}_m & * \\ 0 & S \end{pmatrix} : \begin{array}{c} N(T-\lambda I) \\ \oplus \\ W \end{array} \to \begin{array}{c} N(T-\lambda I) \\ \oplus \\ W \end{array}$$

(here $m = \dim N(T - \lambda I) \geq 1$ and thus $\dim W < \dim V = n$).

The lower right corner is $S : W \to W$. Since the matrix is "block upper-triangular", char_S divides char_T, so char_S also splits. Thus, by the induction hypothesis, can find an orthonormal basis of W in which S is upper-triangular. But then T is upper-triangular as well.

For an explicit o.n. basis, take any o.n. basis of $N(T - \lambda I)$, and extend it with the o.n. basis of W found for S by Schur & the induction hypothesis. ∎

An example of finding an ONB that diagonalizes a normal transformation:

$$A = \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix}$$

Answer: the eigenvalues are $1, 1, -2$, and an orthogonal family of eigenvectors is $(-1, 1, 0), (1, 1, -2)$ (have to do Gram–Schmidt on a basis of the eigenspace for $\lambda = 1$) and $(1, 1, 1)$.

6.5 Section 6.5†: Unitary and orthogonal operators

Began with the definition/meaning of orthogonal and unitary (linear) transformations: these are the ***isomorphisms*** of the "inner product vector space" structure.

Remark 6.19

- See the definition of ***similarity*** on page 116†: $n \times n$ matrices A and B are similar if there is an invertible matrix Q such that $A = Q^{-1}BQ$. This is an equivalence relation.

  The ***unitary/orthogonal equivalence*** requires in addition that Q be a unitary/orthogonal matrix.

  Moreover:

  - A and B similar $\Longrightarrow$ they have the same eigenvalues (because they have the same characteristic polynomial)
  - $A = Q^{-1}BQ \Longrightarrow Q(N(A - \lambda I)) = N(B - \lambda I)$

  These are also true for linear transformations, not only matrices.

- The question in sections 5.1-5.2† was: When is a (square) matrix similar to a diagonal matrix?

  In section 5.1† (e.g., second part of page 251†) we discussed how to diagonalize a matrix: $A = QDQ^{-1} \iff D = Q^{-1}AQ$ with D diagonal exactly when Q has columns the eigenvectors of A (so need a basis of eigenvectors for A) and then D has on the diagonal the corresponding eigenvalues. This is because the standard basis is a basis of eigenvectors for D, and Q has to take eigenvectors of D to eigenvectors of A by the above.

Reflections in $\mathbb{R}^2$

We compute the matrix of a reflection T_L in the line $L \subset \mathbb{R}^2$ with slope α, so having direction $(\cos\alpha, \sin\alpha)$: in the standard basis the matrix is

$$[T_L]_{st} = \begin{pmatrix} \cos(2\alpha) & \sin(2\alpha) \\ \sin(2\alpha) & -\cos(2\alpha) \end{pmatrix}$$

One can see that T_L is a symmetric orthogonal transformation.

- All such reflections are orthogonally equivalent: T_L is orthogonally equivalent to the reflection T_x in the x-axis via a rotation by α:

  $$T_L = R_\alpha T_x R_{-\alpha}, \qquad [R_\alpha]_{st} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \qquad [T_x]_{st} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

  We have $R_{-\alpha}$ on the right because it has to map the eigenvectors of T_L (which are $\beta = \{f_1 = (\cos\alpha, \sin\alpha), f_2 = (\cos(\pi/2 + \alpha), \sin(\pi/2 + \alpha) = (-\sin(\alpha), +\cos(\alpha))\}$) to the eigenvectors of T_x corresponding to the same eigenvalues, (so $\{e_1, e_2\}$, the standard ordered basis of $\mathbb{R}^2$).

- Alternatively, can easily compute $[T_L]_\beta$ because β are eigenvectors of T_L (we get exactly the matrix $[T_x]_{st}$ from above), and then do the basis change from β to standard.

- Direct computation: compute the image of the standard basis from the geometry: the vector e_1 is mapped to the (ordered) direction 2α, so to $(\cos 2\alpha, \sin 2\alpha)$; the vector e_2 is mapped to the direction $2\alpha - \pi/2$, [5] so to $(\cos(2\alpha - \pi/2), \sin(2\alpha - \pi/2)) = (\sin 2\alpha, -\cos 2\alpha)$.

---

[5] E.g., if $\alpha = 0$ then e_2 is mapped to $-e_2$.

6.6 Section $6.6^\dagger$ Orthogonal Projections and the Spectral Theorem

The Spectral Theorem $6.25^\dagger$ is only a restatement of the previous results about normal operators. One way to state it is:

Theorem 6.20 *Let V be a finite dimensional inner product space, and $T : V \to V$ linear.*
If T is normal over $\mathbb{C}$ (or normal and diagonalizable over $\mathbb{R}$ $\iff$ T self-adjoint), then

$$T = \sum_i \lambda_i P_{\lambda_i}$$

where λ_i are the distinct eigenvalues of T and P_λ is the orthogonal projection onto E_λ, the eigenspace corresponding to λ.

Because $\oplus_i E_{\lambda_i} = V$ is in this case an **orthogonal** *direct sum, the projections $\{P_{\lambda_i}\}_i$ form a* **resolution of the identity**.

More about orthogonal projections

Let $W \subset V$ be a subspace in the finite dimensional inner product space V and $P : V \to V$ the orthogonal [6] projection onto W. Then:

- $I - P$ is also an orthogonal projection, onto $W^\perp$.
- The projection P is orthogonal $\iff$ P is a self-adjoint projection (that is, $P = P^2 = P^*$)
- $\mathrm{R}(P) \oplus \mathrm{N}(P) = V$, with the direct sum being *orthogonal* myiff the projection P is orthogonal. In particular, its matrix in an orthonormal basis is self-adjoint.
- The eigenvalues of P are 1 (with multiplicity $\dim W$) and 0 (with multiplicity $\dim V - \dim W$).
- Therefore, the trace [7] of P is equal to $\dim W$.

7 Chapter $7^\dagger$: Jordan normal forms

7.1 The Jordan Canonical Form I
7.2 The Jordan Canonical Form II
7.3 The Minimal Polynomial
7.4* The Rational Canonical Form

7.1 Sections $7.1^\dagger$ and $7.2^\dagger$: The Jordan normal form

Let $T : V \to V$ a linear transformation on a finite dimensional space. Will discuss only the case when char_T splits (this holds when the scalars are $\mathbb{C}$).

Theorem 7.1 *Assume $T : V \to V$ is linear and* char_T *splits. Then:*

---

[6] Many of the next properties also hold without an inner product, in that case P is an idempotent – called a projection too – with image W.

[7] The trace of a linear map $T : V \to V$ is computed from any matrix representation of T, same idea as for $\det(T)$. For a matrix the trace is equal to the sum of its diagonal entris, and also to the sum of its eigenvalues (computed in $\mathbb{C}$), e.g. because of Schur's Theorem 6.18.

- *There is a (Jordan) basis of V in which the matrix of T is block-diagonal, with each block a Jordan block (that is, the diagonal equals an eigenvalue λ, there are 1's in the 1st "line" above the diagonal, and zeros everywhere else).*
- *The Jordan normal form is unique, up to permuting the blocks.*
- *Two matrices are similar $\iff$ they have the same Jordan normal form, up to permuting the blocks.*

Definition 7.2 (Generalized eigenvectors) Given an eigenvalue λ of T, the space of generalized eigenvectors corresponding to λ is

$$K_\lambda := \{\boldsymbol{v} \in V \mid (T - \lambda I)^p \boldsymbol{v} = 0 \text{ for some } p \geq 1\}$$

For an eigenvalue λ:

- have an increasing sequence of null-spaces

  $$E_\lambda = N(T - \lambda I) \underset{\neq}{\subsetneq} N((T-\lambda I)^2) \underset{\neq}{\subsetneq} \cdots \underset{\neq}{\subsetneq} N((T-\lambda I)^p) = N((T-\lambda I)^{p+1})$$

  and $K_\lambda = N((T - \lambda I)^p)$
- the dimensions of the above null-spaces determines the Jordan blocks for the eigenvalue λ; can use the "dot-diagram" introduced in section $7.2^\dagger$
- $\dim E_\lambda$ = number of Jordan blocks for λ
- the biggest Jordan block for λ has size p
- $\dim K_\lambda$ = algebraic multiplicity of λ

Moreover

- each K_λ is T-invariant
- $\oplus_{\text{distinct } \lambda' s} K_\lambda = V$

7.3 Section $7.3^\dagger$: The minimal polynomial

Definition 7.3 (Minimal polynomial) The minimal polynomial of a linear transformation $T : V \to V$, $\dim V < \infty$, is $p(t)$ if:

- $p(t)$ is the polynomial of lowest degree for which $p(T) = \mathbf{0}$, where $\mathbf{0} : V \to V$ is the zero linear transformation on V
- $p(t)$ is monic (that is, its leading coefficient is 1)

Theorem 7.4 *If $p(t)$ is a minimal polynomial for T and $f(T) = \mathbf{0}$ for some polynomial $f(t)$, then $p(t)$ divides $f(t)$.*

This implies that the minimal polynomial is unique.

Theorem 7.5 *The minimal polynomial and the characteristic polynomial have the same roots.*

Theorem 7.6 *$T : V \to V$ is diagonalizable $\iff$ its minimal polynomial has only degree-one factors, that is $p(t) = (t - \lambda_1)(t - \lambda_2)\dots(t - \lambda_k)$ with distinct λ_i's.*

Proof: This is a simpler proof then the one in the book, relying on the Jordan normal form.

If T is diagonalizable then the minimal polynomial is clearly as claimed.

Conversely, assume the minimal polynomial is as above. Then the characteristic polynomial splits (since it has the same roots as the minimal polynomial), therefore T has a Jordan normal form. In that case the minimal polynomial has each $(t - \lambda_j)$ at the power given by the largest Jordan block with that eigenvalue, so all blocks are 1×1. ∎