

9.7 Triangularization

Section 4 found two ways for diagonalization to fail. The field can be too small, as for a rotation of $\mathbb{R}^2$. Or there can be too few eigenvectors, as for $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, and no change of field helps. In the second case we should ask for less. The next simplest shape after a diagonal matrix is an upper triangular one, and triangular matrices are almost as good for computation: their eigenvalues, trace, and determinant are visible on the diagonal. This section proves that an operator has a triangular matrix exactly when its characteristic polynomial splits, so every complex matrix is similar to a triangular one. The proof is the induction on dimension that Section 1 promised: take one eigenvector, pass to the quotient, and repeat.

9.7.1 Triangular matrices and chains of invariant subspaces

As in Section 2, start with the wish and unwind it. Let $\mathcal{B} = (v_1, \dots, v_n)$ be a basis of V and $T \in \mathcal{L}(V)$. Column j of $[T]_{\mathcal{B}}^{\mathcal{B}}$ holds the coordinates of Tv_j (Definition 4.1.1). The matrix is upper triangular when every entry below the diagonal is 0 (Definition 0.9.5), that is, when Tv_j is a combination of $v_1, \dots, v_j$ only. So the first basis vector must be sent to a multiple of itself, the second into the plane of the first two, and so on. In the language of Section 1, each of the growing spans must be invariant. Chapter 2 proved this as an exercise (Exercise (C1: Upper triangular matrices and invariant subspaces)), where the word *invariant* was introduced just for that problem; here it is again, with one more piece of information about the diagonal.

Theorem 9.7.1. Triangular Matrices and Chains of Invariant Subspaces

Let V be finite-dimensional with basis $\mathcal{B} = (v_1, \dots, v_n)$, let $T \in \mathcal{L}(V)$, and put $U_0 = \{0\}$ and $U_k = \text{span}(v_1, \dots, v_k)$ for $1 \leq k \leq n$. The following are equivalent:

- (a) $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular;
- (b) $Tv_k \in U_k$ for every $k = 1, \dots, n$;
- (c) each U_k is T -invariant.

In that case, if $d_1, \dots, d_n$ are the diagonal entries of $[T]_{\mathcal{B}}^{\mathcal{B}}$, then $(T - d_k \text{id}_V)(U_k) \subseteq U_{k-1}$ for every k .

Proof. Write $[T]_{\mathcal{B}}^{\mathcal{B}} = (a_{ij})$, so that $Tv_j = \sum_i a_{ij}v_i$ with unique coefficients (Definition 4.1.1, Theorem 1.5.1).

- (a) $\Leftrightarrow$ (b). If $a_{ij} = 0$ for $i > j$, then $Tv_j = \sum_{i \leq j} a_{ij}v_i \in U_j$. Conversely, if $Tv_j \in U_j$, it has a representation using only $v_1, \dots, v_j$; by uniqueness, $a_{ij} = 0$ for $i > j$.
- (b) $\Rightarrow$ (c). For $j \leq k$, $Tv_j \in U_j \subseteq U_k$, so U_k is invariant by Lemma 9.1.1.
- (c) $\Rightarrow$ (b). $v_k \in U_k$, so $Tv_k \in U_k$ by invariance.

Finally, suppose (a) holds and let $u = \sum_{j \leq k} b_j v_j \in U_k$. For $j < k$, $(T - d_k \text{id}_V)v_j = Tv_j - d_k v_j \in U_j \subseteq U_{k-1}$. For $j = k$, $(T - d_k \text{id}_V)v_k = \sum_{i \leq k} a_{ik}v_i - a_{kk}v_k = \sum_{i < k} a_{ik}v_i \in U_{k-1}$, since $d_k = a_{kk}$. By linearity, $(T - d_k \text{id}_V)u \in U_{k-1}$.

■

A chain $U_1 \subset U_2 \subset \cdots \subset U_n = V$ of subspaces with $\dim U_k = k$ is often called a **flag**, after the picture of a point on a line in a plane in space. The theorem says: an upper triangular matrix is the same thing as a flag of invariant subspaces, together with a basis adapted to it. The last statement says what the diagonal entry d_k measures. Modulo the smaller space U_{k-1} , the operator T acts on the new basis vector v_k as multiplication by d_k .

The operators for which such a basis exists deserve a name.

Definition 9.7.1. Triangularizable

Let V be a finite-dimensional vector space over F . An operator $T \in \mathcal{L}(V)$ is **triangularizable over F** if there **exists** a basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular. A matrix $\mathbf{A} \in M_n(F)$ is **triangularizable over F** if there **exist** an **invertible** $\mathbf{P} \in M_n(F)$ and an upper triangular $\mathbf{U} \in M_n(F)$ with $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{U}$.

As for diagonalizability, $\mathbf{A}$ is triangularizable over F exactly when $T_{\mathbf{A}}$ is, by Theorem 4.2.4.

Examples.

- **Diagonalizable operators** are triangularizable, since a diagonal matrix is upper triangular. So are all operators on a space of dimension 1, and the zero and identity operators.
- **Differentiation.** For D on $\mathbb{R}[x]_{\leq n}$ in the basis $(1, x, \dots, x^n)$, the matrix is upper triangular with zero diagonal: $D(x^k) = kx^{k-1} \in \text{span}(1, \dots, x^k)$. The flag is $\mathbb{R}[x]_{\leq 0} \subset \mathbb{R}[x]_{\leq 1} \subset \cdots \subset \mathbb{R}[x]_{\leq n}$, a chain of invariant subspaces we met in Section 1. D is not diagonalizable (its only eigenvectors are the constants), but it is triangularizable.
- **Too few eigenvectors.** $\mathbf{J} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is already triangular, with flag $\text{span}(\mathbf{e}_1) \subset F^2$.

Nonexample by minimal change. Reverse the basis for D on $\mathbb{R}[x]_{\leq 2}$: take $(x^2, x, 1)$. The spans are $\text{span}(x^2) \subset \text{span}(x^2, x) \subset \mathbb{R}[x]_{\leq 2}$, still a chain of subspaces of dimensions 1, 2, 3. What fails is invariance of the first one: $D(x^2) = 2x \notin \text{span}(x^2)$. Accordingly, the matrix in this basis is

$$\begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

which is lower triangular, not upper. Being triangularizable is a property of T ; being triangular is a property of T **and** a basis.

A triangularizable operator on $V \neq \{0\}$ must have an invariant line $U_1 = \text{span}(v_1)$, hence an eigenvector (Proposition 9.1.2). So the rotation R of $\mathbb{R}^2$ by a right angle, which has no invariant line (Example (A Rotation Has No Invariant Lines)), is **not** triangularizable over $\mathbb{R}$. An eigenvector to start the flag is exactly what is needed, and the next theorem shows that it is also enough, provided eigenvectors keep existing at every stage.

Quick check. Is $\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \in M_2(\mathbb{R})$ triangularizable over $\mathbb{R}$? If so, find a basis in which its matrix is upper triangular.

Solution. Yes. We need a first basis vector that is an eigenvector. The matrix sends $\mathbf{e}_2$ to $(0, 2) = 2\mathbf{e}_2$, so start with $\mathbf{e}_2$, and complete with $\mathbf{e}_1$. In the basis $(\mathbf{e}_2, \mathbf{e}_1)$, the image of $\mathbf{e}_2$

is $2\mathbf{e}_2$ and the image of $\mathbf{e}_1$ is $(2, 1) = \mathbf{e}_2 + 2\mathbf{e}_1$, so the matrix is $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, which is upper triangular. (The standard basis gives a lower triangular matrix: $\text{span}(\mathbf{e}_1)$ is not invariant.)

9.7.2 The Triangularization Theorem

If T has a triangular matrix, its characteristic polynomial is a product of linear factors, one for each diagonal entry. The converse is the real content: splitting of p_T is enough to build the flag.

Theorem 9.7.2. Triangularization Theorem

Let V be a finite-dimensional vector space over F with $\dim V = n \geq 1$, and let $T \in \mathcal{L}(V)$. Then T is triangularizable over F if and only if p_T splits over F . The same holds for $\mathbf{A} \in M_n(F)$ and $p_{\mathbf{A}}$.

Idea. Take one eigenvector, extend, recurse. Since p_T splits, it has a root in F , so T has an eigenvector v_1 . Put $U = \text{span}(v_1)$ and extend v_1 to a basis $\mathcal{B}$, with $\bar{\mathcal{B}}$ the basis it induces on V/U . By Theorem 9.1.3 the matrix looks like

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \lambda & * \\ \mathbf{0} & [\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}} \end{pmatrix},$$

where $\bar{T}$ is the induced operator on the smaller space V/U . The first column is already right. If $[\bar{T}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}}$ could be made triangular, the whole matrix would be. Now $\bar{T}$ lives on a space of dimension $n - 1$, and $p_T = (x - \lambda)p_{\bar{T}}$, so $p_{\bar{T}}$ splits too: this is the induction on dimension. The only technical point is that a basis of V/U lifts to vectors that, together with v_1 , form a basis of V .

Proof. ($\Rightarrow$) Let $[T]_{\mathcal{B}}^{\mathcal{B}}$ be upper triangular with diagonal entries $d_1, \dots, d_n$. The matrix $x\mathbf{I}_n - [T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular with diagonal entries $x - d_i$, and Theorem 7.4.6 holds over $F[x]$ (the remark at the end of Chapter 7, §4). So $p_T = (x - d_1) \cdots (x - d_n)$ by Definition 7.10.2, which splits (Definition 6.4.2).

($\Leftarrow$) We prove, by induction on $n \geq 1$, that every operator on an n -dimensional space whose characteristic polynomial splits is triangularizable. For $n = 1$, every 1×1 matrix is upper triangular.

Let $n \geq 2$, and assume the statement for spaces of dimension $n - 1$. Since p_T is monic of degree n (Theorem 7.10.2) and splits, $p_T = (x - c_1)(x - c_2) \cdots (x - c_n)$ with $c_i \in F$. As $p_T(c_1) = 0$, Theorem 9.2.1 gives an eigenvector v_1 with $Tv_1 = c_1v_1$. Let $U = \text{span}(v_1)$. It is T -invariant (Proposition 9.1.2), $\dim U = 1$, and $U \neq V$ since $n \geq 2$. Let $\bar{T}$ be the induced operator on V/U (Definition 9.1.2), a space of dimension $n - 1 \geq 1$ (Theorem 4.4.5).

$p_{\bar{T}}$ splits. The restriction $T|_U$ is $c_1 \text{id}_U$, whose matrix in the basis (v_1) is (c_1) , so $p_{T|_U} = x - c_1$. By Theorem 9.1.3 (b),

$$(x - c_1)(x - c_2) \cdots (x - c_n) = p_T = (x - c_1)p_{\bar{T}}.$$

Canceling the nonzero factor $x - c_1$ (Corollary 0.8.4 (2)) gives $p_{\bar{T}} = (x - c_2) \cdots (x - c_n)$,

which splits.

Apply the induction hypothesis. There is a basis $\bar{\mathcal{C}} = (w_2 + U, \dots, w_n + U)$ of V/U with $[\bar{T}]_{\bar{\mathcal{C}}}^{\bar{\mathcal{C}}}$ upper triangular, for some $w_2, \dots, w_n \in V$.

Lift the basis. Let $\mathcal{B} = (v_1, w_2, \dots, w_n)$. Suppose $av_1 + b_2w_2 + \dots + b_nw_n = 0$. Passing to cosets, and using $v_1 \in U$, gives $b_2(w_2 + U) + \dots + b_n(w_n + U) = 0 + U$ (Theorem 4.4.4). Since $\bar{\mathcal{C}}$ is a basis, all $b_i = 0$. Then $av_1 = 0$ with $v_1 \neq 0$, so $a = 0$ (Theorem 1.1.9). Thus $\mathcal{B}$ is a linearly independent list of n vectors in V , hence a basis (Theorem 1.6.4).

The matrix. $\mathcal{B}$ extends the basis (v_1) of U , and the cosets of the added vectors form $\bar{\mathcal{C}}$. By Theorem 9.1.3 (b),

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} c_1 & * \\ \mathbf{0} & [\bar{T}]_{\bar{\mathcal{C}}}^{\bar{\mathcal{C}}} \end{pmatrix},$$

where the $\mathbf{0}$ is a column of $n - 1$ zeros. The block $[\bar{T}]_{\bar{\mathcal{C}}}^{\bar{\mathcal{C}}}$ is upper triangular, so $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular. This completes the induction.

For $\mathbf{A} \in M_n(F)$, apply the result to $T_{\mathbf{A}}$, using Theorem 4.2.4. ■

The proof is an algorithm, and its output is a flag. The first invariant subspace is the eigenline U ; the k -th is spanned by v_1 and the lifts $w_2, \dots, w_k$ of the first $k - 1$ vectors of the flag for $\bar{T}$. The eigenvalue found at each stage is the next diagonal entry.

Over the complex numbers the hypothesis is automatic.

Corollary 9.7.3. Complex Operators Are Triangularizable

Every operator on a finite-dimensional complex vector space $V \neq \{0\}$ is triangularizable over $\mathbb{C}$. Every $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$, is similar to an upper triangular matrix in $M_n(\mathbb{C})$.

Proof. By Theorem 7.10.2, p_T is nonzero, and by Corollary 6.5.5 (a) it splits over $\mathbb{C}$. Apply Theorem 9.7.2. ■

So over $\mathbb{C}$ the question of Chapter 4 always has at least this answer: a basis exists in which the matrix has zeros below the diagonal. A real matrix can always be triangularized as a complex matrix, but perhaps only with complex entries.

Warning. Triangularizing may require leaving the field, and the triangular form is far from unique. The rotation $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has $p_{\mathbf{R}} = x^2 + 1$, which does not split over $\mathbb{R}$, so no **real** invertible $\mathbf{P}$ makes $\mathbf{P}^{-1}\mathbf{R}\mathbf{P}$ triangular; over $\mathbb{C}$ it becomes $\text{diag}(i, -i)$. And a triangularizable matrix has many triangular forms. $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ and $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ are both triangular forms of the same operator (the first has distinct eigenvalues 1, 2, so it is diagonalizable, and similar to the second). The off-diagonal entries change with the flag and the basis, and even the order of the diagonal entries is not fixed. Only the diagonal entries, as a list with

repetitions, are determined, as the next theorem shows.

9.7.3 What the diagonal records

A diagonalization displays the eigenvalues. A triangularization does too, with their algebraic multiplicities.

Theorem 9.7.4. The Diagonal of a Triangular Form

Let V be finite-dimensional with $\dim V = n \geq 1$, $T \in \mathcal{L}(V)$, and let $\mathcal{B}$ be a basis with $[T]_{\mathcal{B}}^{\mathcal{B}}$ upper triangular, with diagonal entries $d_1, \dots, d_n$. Then:

- (a) $p_T = (x - d_1)(x - d_2) \cdots (x - d_n)$;
- (b) $\text{spec}(T) = \{d_1, \dots, d_n\}$, and each eigenvalue λ occurs among $d_1, \dots, d_n$ exactly $a_T(\lambda)$ times;
- (c) $\text{tr } T = d_1 + \cdots + d_n$ and $\det T = d_1 \cdots d_n$.

In particular, any two triangular forms of T have the same diagonal entries, up to order.

Proof.

- (a) was shown in the proof of [Theorem 9.7.2](#), ($\Rightarrow$).
- (b) By Theorem 9.2.1, the eigenvalues are the roots of p_T in F , and λ is a root of $\prod_i (x - d_i)$ exactly when $\prod_i (\lambda - d_i) = 0$, that is, when $\lambda = d_i$ for some i (Theorem 0.7.1). By Theorem 6.4.2, $a_T(\lambda) = \text{mult}_{\lambda}(p_T) = \sum_i \text{mult}_{\lambda}(x - d_i)$. Each term is 1 if $d_i = \lambda$ and 0 otherwise, by Theorem 6.1.3 (b), since $x - d_i$ is not divisible by $(x - \lambda)^2$ for degree reasons. So the sum counts the indices with $d_i = \lambda$.
- (c) By Definition 4.5.2, $\text{tr } T = \text{tr}[T]_{\mathcal{B}}^{\mathcal{B}} = \sum_i d_i$. By Definition 7.5.1 and Theorem 7.4.6, $\det T = \det[T]_{\mathcal{B}}^{\mathcal{B}} = \prod_i d_i$.

The final statement follows from (b): the number of times λ occurs on the diagonal is $a_T(\lambda)$, which does not depend on the basis. ■

Part (c) recovers Theorem 9.3.3. In Section 3 it came from comparing coefficients of p_T ; here it is read off a matrix. For instance, a complex 3×3 matrix with $p_{\mathbf{A}} = (x - 1)^2(x + 4)$ is similar to a triangular matrix with diagonal 1, 1, -4 in some order, so $\text{tr } \mathbf{A} = -2$ and $\det \mathbf{A} = -4$, however complicated $\mathbf{A}$ looks.

The theorem does **not** say that a triangular form with **distinct** diagonal entries is itself diagonal: $\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ is a counterexample. What does follow is that T is then diagonalizable (Corollary 9.4.2), so some **other** basis makes its matrix diagonal. And an operator with repeated diagonal entries may or may not be diagonalizable: compare $\mathbf{I}_2$ with $\mathbf{J}$.

9.7.4 Triangularizing a matrix in practice

For a matrix $\mathbf{A} \in M_n(F)$ whose characteristic polynomial splits, the proof of [Theorem 9.7.2](#) becomes a procedure.

1. Find an eigenvalue λ_1 and an eigenvector $\mathbf{v}_1$.
2. Extend $\mathbf{v}_1$ to a basis $(\mathbf{v}_1, \mathbf{w}_2, \dots, \mathbf{w}_n)$ of F^n , choosing the $\mathbf{w}_i$ as simple as possible, for example standard basis vectors.
3. Compute the matrix of the induced operator on $F^n / \text{span}(\mathbf{v}_1)$ in the basis $(\mathbf{w}_2 + U, \dots, \mathbf{w}_n + U)$. Its entries are the coordinates of $\mathbf{A}\mathbf{w}_j$ along $\mathbf{w}_2, \dots, \mathbf{w}_n$, ignoring the $\mathbf{v}_1$ -coordinate.
4. Triangularize this $(n-1) \times (n-1)$ matrix in the same way, and lift each coset back to its representative.

Example. Triangularizing a Matrix with a Single Eigenvector Line

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 1 \\ 1 & -1 & 2 \end{pmatrix} \in M_3(\mathbb{R}).$$

Show that $\mathbf{A}$ is not diagonalizable but is triangularizable over $\mathbb{R}$, and find an invertible $\mathbf{P}$ with $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ upper triangular.

Solution. The characteristic polynomial. Expanding along the first row (Theorem 7.6.1 over $\mathbb{R}[x]$),

$$\begin{aligned} p_{\mathbf{A}} &= \det \begin{pmatrix} x-2 & 0 & -1 \\ 0 & x-2 & -1 \\ -1 & 1 & x-2 \end{pmatrix} \\ &= (x-2)[(x-2)^2 + 1] + (-1) \det \begin{pmatrix} 0 & x-2 \\ -1 & 1 \end{pmatrix}. \end{aligned}$$

The last determinant is $0 \cdot 1 - (x-2)(-1) = x-2$, so $p_{\mathbf{A}} = (x-2)[(x-2)^2 + 1 - 1] = (x-2)^3$. It splits, so $\mathbf{A}$ is triangularizable by [Theorem 9.7.2](#).

Eigenvectors. $\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \end{pmatrix}$ has rank 2, so $g(2) = 1 < 3 = a(2)$, and $\mathbf{A}$ is not

diagonalizable (Theorem 9.4.1 (e)). The system gives $z = 0$ and $x = y$, so $E_2(\mathbf{A}) = \text{span}(\mathbf{v}_1)$ with $\mathbf{v}_1 = (1, 1, 0)$: all eigenvectors lie on one line.

Step 1. Let $U = \text{span}(\mathbf{v}_1)$, and extend to the basis $(\mathbf{v}_1, \mathbf{e}_2, \mathbf{e}_3)$ of $\mathbb{R}^3$ (the matrix with these columns has determinant 1).

Step 2: the induced operator. Express the images of $\mathbf{e}_2$ and $\mathbf{e}_3$ in the basis:

$$\mathbf{A}\mathbf{e}_2 = (0, 2, -1) = 0\mathbf{v}_1 + 2\mathbf{e}_2 - \mathbf{e}_3, \quad \mathbf{A}\mathbf{e}_3 = (1, 1, 2) = 1\mathbf{v}_1 + 0\mathbf{e}_2 + 2\mathbf{e}_3.$$

Dropping the $\mathbf{v}_1$ -coordinates, the induced operator $\bar{\mathbf{A}}$ on $\mathbb{R}^3/U$ has, in the basis $\bar{\mathcal{B}} = (\mathbf{e}_2 + U, \mathbf{e}_3 + U)$, the matrix

$$[\bar{\mathbf{A}}]_{\bar{\mathcal{B}}}^{\bar{\mathcal{B}}} = \begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix},$$

with $p_{\bar{\mathbf{A}}} = (x - 2)^2$, as the proof predicts.

Step 3: triangularize $\bar{\mathbf{A}}$. Its second column says $\bar{\mathbf{A}}(\mathbf{e}_3 + U) = 2(\mathbf{e}_3 + U)$, so $\mathbf{e}_3 + U$ is an eigenvector. In the basis $(\mathbf{e}_3 + U, \mathbf{e}_2 + U)$, $\bar{\mathbf{A}}(\mathbf{e}_2 + U) = -(\mathbf{e}_3 + U) + 2(\mathbf{e}_2 + U)$, and the matrix becomes $\begin{pmatrix} 2 & -1 \\ 0 & 2 \end{pmatrix}$.

Step 4: lift. The basis $\mathcal{B} = (\mathbf{v}_1, \mathbf{e}_3, \mathbf{e}_2)$ of $\mathbb{R}^3$ gives

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Check. $\mathbf{A}\mathbf{v}_1 = (2, 2, 0) = 2\mathbf{v}_1$; $\mathbf{A}\mathbf{e}_3 = (1, 1, 2) = \mathbf{v}_1 + 2\mathbf{e}_3$; $\mathbf{A}\mathbf{e}_2 = (0, 2, -1) = -\mathbf{e}_3 + 2\mathbf{e}_2$. These are the columns $(2, 0, 0)$, $(1, 2, 0)$, $(0, -1, 2)$ of the triangular matrix, read in the basis $\mathcal{B}$. Also $\det \mathbf{P} = -1 \neq 0$.

The flag is $\text{span}((1, 1, 0)) \subset \{(a, a, b) : a, b \in \mathbb{R}\} \subset \mathbb{R}^3$. One can check the plane directly: $\mathbf{A}(a, a, b) = (2a + b, 2a + b, 2b)$ lies in it again. The diagonal is $2, 2, 2$, as [Theorem 9.7.4](#) requires.

Different choices in Step 2 give different triangular matrices, with the same diagonal. Chapter 10 shows how to choose the basis so that the entries above the diagonal are as simple as possible (the Jordan form). Chapter 12 shows that over $\mathbb{C}$ with an inner product, the basis can even be chosen orthonormal (Schur's theorem).

9.7.5 Exercises

9.7.5.1 A. Check your understanding

Exercise. A1

- Define what it means for $\mathbf{A} \in M_n(F)$ to be triangularizable over F .
- State the Triangularization Theorem.
- For a basis $(v_1, \dots, v_n)$, which subspaces must be invariant for $[T]_{\mathcal{B}}^{\mathcal{B}}$ to be upper triangular?
- True or false: every $\mathbf{A} \in M_n(\mathbb{R})$ is triangularizable over $\mathbb{R}$. Justify your answer.
- True or false: every triangularizable matrix is diagonalizable. Justify your answer.
- Let $\mathbf{A} \in M_4(\mathbb{C})$ have $p_{\mathbf{A}} = (x - 3)^3(x + 1)$. What are the diagonal entries of any upper triangular matrix similar to $\mathbf{A}$, and what are $\text{tr } \mathbf{A}$ and $\det \mathbf{A}$?

Solution.

- There are an invertible $\mathbf{P} \in M_n(F)$ and an upper triangular $\mathbf{U} \in M_n(F)$ with $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{U}$ ([Definition 9.7.1](#)).
- An operator on a space of dimension $n \geq 1$ over F is triangularizable over F if and only if its characteristic polynomial splits over F ([Theorem 9.7.2](#)).
- All the spans $\text{span}(v_1, \dots, v_k)$, $1 \leq k \leq n$ ([Theorem 9.7.1](#)).

- (d) False. The rotation $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has $p = x^2 + 1$, which does not split over $\mathbb{R}$.
- (e) False. $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is triangular but not diagonalizable.
- (f) By [Theorem 9.7.4](#), the diagonal is $3, 3, 3, -1$ in some order, so $\operatorname{tr} \mathbf{A} = 8$ and $\det \mathbf{A} = -27$.

9.7.5.2 B. Practice

Exercise. B1: Triangularizing matrices

For each real matrix, find an invertible $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is upper triangular, and write down $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$.

- (a) $\mathbf{A} = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$.
- (b) $\mathbf{A} = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 3 & 1 \\ 1 & 0 & 2 \end{pmatrix}$, with $p_{\mathbf{A}} = (x-1)(x-3)^2$ (Example (A 3×3 Characteristic Polynomial)).

Solution.

- (a) $\operatorname{tr} \mathbf{A} = 4$, $\det \mathbf{A} = 4$, so $p_{\mathbf{A}} = (x-2)^2$, which splits. $\mathbf{A} - 2\mathbf{I} = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$ gives the eigenvector $\mathbf{v}_1 = (1, -1)$. Extend with $\mathbf{e}_2$: $\mathbf{A}\mathbf{e}_2 = (1, 1) = \mathbf{v}_1 + 2\mathbf{e}_2$. So with

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.$$

(In size 2, one eigenvector already fixes the flag.)

- (b) $\mathbf{A}\mathbf{e}_2 = (0, 3, 0) = 3\mathbf{e}_2$, so $\mathbf{v}_1 = \mathbf{e}_2$ is an eigenvector for 3. Let $U = \operatorname{span}(\mathbf{e}_2)$ and extend to $(\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3)$. Then $\mathbf{A}\mathbf{e}_1 = (2, 1, 1) = \mathbf{e}_2 + 2\mathbf{e}_1 + \mathbf{e}_3$ and $\mathbf{A}\mathbf{e}_3 = (1, 1, 2) = \mathbf{e}_2 + \mathbf{e}_1 + 2\mathbf{e}_3$, so the induced operator has matrix $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ in the basis $(\mathbf{e}_1 + U, \mathbf{e}_3 + U)$. Its characteristic polynomial is $x^2 - 4x + 3 = (x-3)(x-1)$, with eigenvectors $(1, 1)$ for 3 and $(1, -1)$ for 1, that is, $\mathbf{e}_1 + \mathbf{e}_3 + U$ and $\mathbf{e}_1 - \mathbf{e}_3 + U$. Lifting gives the basis $(\mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_3, \mathbf{e}_1 - \mathbf{e}_3)$. Now

$$\begin{aligned} \mathbf{A}(\mathbf{e}_1 + \mathbf{e}_3) &= (3, 2, 3) = 2\mathbf{e}_2 + 3(\mathbf{e}_1 + \mathbf{e}_3), \\ \mathbf{A}(\mathbf{e}_1 - \mathbf{e}_3) &= (1, 0, -1) = \mathbf{e}_1 - \mathbf{e}_3, \end{aligned}$$

so

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \quad \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{pmatrix} 3 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with $\det \mathbf{P} = 2 \neq 0$ (expand along the second row). The diagonal $3, 3, 1$ agrees with $p_{\mathbf{A}}$.

Exercise. B2: Nilpotent matrices are strictly triangular

Let $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$, with $\mathbf{A}^k = \mathbf{O}$ for some $k \geq 1$. Prove that $\mathbf{A}$ is similar to a **strictly** upper triangular matrix, that is, an upper triangular matrix with zero diagonal.

Solution. By [Corollary 9.7.3](#), $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{U}$ is upper triangular for some invertible $\mathbf{P}$. By [Theorem 9.7.4](#) (b), each diagonal entry of $\mathbf{U}$ is an eigenvalue of $\mathbf{A}$. By Exercise (C2: Nilpotent operators) (a), the only eigenvalue of a nilpotent operator on a nonzero space is 0. So every diagonal entry of $\mathbf{U}$ is 0, and $\mathbf{U}$ is strictly upper triangular. (The statement holds over every field, with a different proof: use the chain $\ker \mathbf{A} \subseteq \ker \mathbf{A}^2 \subseteq \cdots \subseteq \ker \mathbf{A}^k = F^n$ to build the basis.)

Exercise. B3: Eigenvalues of powers from a triangular form

Let $\mathbf{A} \in M_n(\mathbb{C})$, $n \geq 1$, and let $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{U}$ be upper triangular with diagonal entries $d_1, \dots, d_n$.

- Show that the product of two upper triangular matrices is upper triangular, with diagonal entries the products of the corresponding diagonal entries.
- Deduce that $p_{\mathbf{A}^k} = (x - d_1^k) \cdots (x - d_n^k)$ and $\text{tr}(\mathbf{A}^k) = d_1^k + \cdots + d_n^k$ for every $k \geq 1$.
- Suppose $p_{\mathbf{A}} = (x - 1)(x + 1)(x - 2)$. Find $p_{\mathbf{A}^2}$, $\text{tr} \mathbf{A}^2$, and $\det \mathbf{A}^2$.

Solution.

- Let $\mathbf{U}, \mathbf{U}'$ be upper triangular. The (i, j) -entry of $\mathbf{U}\mathbf{U}'$ is $\sum_l u_{il}u'_{lj}$. A term can be nonzero only if $u_{il} \neq 0$ and $u'_{lj} \neq 0$, which forces $i \leq l \leq j$. So the entry is 0 when $i > j$, and for $i = j$ only $l = i$ survives, giving $u_{ii}u'_{ii}$.
- $\mathbf{P}^{-1}\mathbf{A}^k\mathbf{P} = (\mathbf{P}^{-1}\mathbf{A}\mathbf{P})^k = \mathbf{U}^k$, by induction on k . By (a) and induction, $\mathbf{U}^k$ is upper triangular with diagonal entries d_i^k . By [Theorem 9.7.4](#) (a) and (c), applied to $T_{\mathbf{A}^k}$ in the basis given by the columns of $\mathbf{P}$, $p_{\mathbf{A}^k} = \prod_i (x - d_i^k)$ and $\text{tr} \mathbf{A}^k = \sum_i d_i^k$.
- By [Theorem 9.7.4](#), the diagonal of $\mathbf{U}$ is $1, -1, 2$ in some order. By (b), $p_{\mathbf{A}^2} = (x - 1)(x - 1)(x - 4) = (x - 1)^2(x - 4)$, $\text{tr} \mathbf{A}^2 = 1 + 1 + 4 = 6$, and $\det \mathbf{A}^2 = 1 \cdot 1 \cdot 4 = 4$. The last agrees with $\det \mathbf{A}^2 = (\det \mathbf{A})^2 = (-2)^2$.

9.7.5.3 C. Going deeper**Exercise. C1: A second proof of Cayley–Hamilton**

Let V be finite-dimensional with $\dim V = n \geq 1$, and let $T \in \mathcal{L}(V)$ have a basis $\mathcal{B} = (v_1, \dots, v_n)$ with $[T]_{\mathcal{B}}^{\mathcal{B}}$ upper triangular, with diagonal entries $d_1, \dots, d_n$. Let $U_0 = \{0\}$ and $U_k = \text{span}(v_1, \dots, v_k)$.

- Let $S_k = (T - d_1 \text{id}_V)(T - d_2 \text{id}_V) \cdots (T - d_k \text{id}_V)$. Prove by induction on k that $S_k(U_k) = \{0\}$ for $1 \leq k \leq n$.
- Deduce that $p_T(T) = 0$, without using the adjugate.
- Deduce the Cayley–Hamilton Theorem for every $\mathbf{A} \in M_n(\mathbb{C})$, and then for every $\mathbf{A} \in$

$M_n(\mathbb{R})$.

- (d) Why does this argument, on its own, not prove the theorem for $\mathbf{A} \in M_2(\mathbb{F}_2)$ with $p_{\mathbf{A}} = x^2 + x + 1$?

Hint: for (a), use the last statement of Theorem 9.7.1.

Solution.

- (a) By Theorem 9.7.1, each U_k is invariant and $(T - d_k \text{id}_V)(U_k) \subseteq U_{k-1}$. For $k = 1$, $S_1(U_1) = (T - d_1 \text{id}_V)(U_1) \subseteq U_0 = \{0\}$. Suppose $S_{k-1}(U_{k-1}) = \{0\}$ for some $k \geq 2$. Since $S_k = S_{k-1}(T - d_k \text{id}_V)$, for $u \in U_k$ we get $S_k u = S_{k-1}((T - d_k \text{id}_V)u) \in S_{k-1}(U_{k-1}) = \{0\}$.
- (b) By Theorem 9.7.4 (a), $p_T = \prod_{i=1}^n (x - d_i)$, so $p_T(T) = S_n$ by Theorem 6.7.1 (b). By (a), S_n kills $U_n = V$, so $p_T(T) = 0$.
- (c) Let $\mathbf{A} \in M_n(\mathbb{C})$. By Corollary 9.7.3, $T_{\mathbf{A}}$ has a basis in which its matrix is upper triangular, and (b) gives $p_{\mathbf{A}}(T_{\mathbf{A}}) = 0$, so $p_{\mathbf{A}}(\mathbf{A}) = \mathbf{O}$ by Corollary 4.1.3. For $\mathbf{A} \in M_n(\mathbb{R})$, regard $\mathbf{A}$ as an element of $M_n(\mathbb{C})$. The Leibniz formula gives the same polynomial $p_{\mathbf{A}}$ with real coefficients, and $p_{\mathbf{A}}(\mathbf{A})$ is computed from the same entries by the same sums and products; it is $\mathbf{O}$ by the complex case.
- (d) Over $\mathbb{F}_2$, $x^2 + x + 1$ has no root (its values at 0 and 1 are both 1), so it does not split, and the matrix is not triangularizable over $\mathbb{F}_2$ by Theorem 9.7.2. To run the argument we would need a larger field containing $\mathbb{F}_2$ in which the polynomial splits, and we have not constructed one. The adjugate proof of Theorem 9.6.1 needs no such field. (For instance $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ has this characteristic polynomial, and indeed $\mathbf{A}^2 + \mathbf{A} + \mathbf{I} = \mathbf{O}$ over $\mathbb{F}_2$.)

Exercise. C2: Real operators have small invariant subspaces

Let $\mathbf{A} \in M_n(\mathbb{R})$, $n \geq 1$, regarded also as a complex matrix.

- (a) Let $\lambda = a + bi$ ($a, b \in \mathbb{R}$) be a complex eigenvalue of $\mathbf{A}$ with eigenvector $\mathbf{z} = \mathbf{x} + i\mathbf{y}$, where $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Show that $\mathbf{A}\mathbf{x} = a\mathbf{x} - b\mathbf{y}$ and $\mathbf{A}\mathbf{y} = b\mathbf{x} + a\mathbf{y}$.
- (b) Deduce that $\text{span}(\mathbf{x}, \mathbf{y}) \subseteq \mathbb{R}^n$ is $\mathbf{A}$ -invariant, and has dimension 1 or 2.
- (c) Deduce that every operator on a real vector space V with $1 \leq \dim V < \infty$ has an invariant subspace of dimension 1 or 2.
- (d) For $\mathbf{C} = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}$, check that $\mathbf{z} = (0, -2, 1) + i(-2, 1, 0)$ is an eigenvector for i , and find the invariant plane of (b).

Solution.

- (a) The eigenvalue exists by Theorem 9.2.4. Since $\mathbf{A}$ has real entries, $\mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x} + i\mathbf{A}\mathbf{y}$ with $\mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \in \mathbb{R}^n$. On the other hand, $\lambda\mathbf{z} = (a + bi)(\mathbf{x} + i\mathbf{y}) = (a\mathbf{x} - b\mathbf{y}) + i(b\mathbf{x} + a\mathbf{y})$. A vector of $\mathbb{C}^n$ determines its real and imaginary parts entry by entry, so $\mathbf{A}\mathbf{x} = a\mathbf{x} - b\mathbf{y}$ and $\mathbf{A}\mathbf{y} = b\mathbf{x} + a\mathbf{y}$.

- (b) By (a), $\mathbf{Ax}$ and $\mathbf{Ay}$ lie in $W = \text{span}(\mathbf{x}, \mathbf{y})$, so W is invariant by Lemma 9.1.1. Since $\mathbf{z} \neq \mathbf{0}$, at least one of $\mathbf{x}, \mathbf{y}$ is nonzero, so $\dim W \geq 1$, and $\dim W \leq 2$ since W is spanned by two vectors.
- (c) Choose a basis $\mathcal{B}$ of V , let $\mathbf{A} = [T]_{\mathcal{B}}^{\mathcal{B}} \in M_n(\mathbb{R})$, and let W be as in (b). Let $\Phi: V \rightarrow \mathbb{R}^n$ be the coordinate isomorphism $v \mapsto [v]_{\mathcal{B}}$ (Corollary 2.5.5). Then $\Phi^{-1}(W)$ is a subspace of the same dimension as W , and for $v \in \Phi^{-1}(W)$, $\Phi(Tv) = \mathbf{A}\Phi(v) \in W$ (Theorem 4.1.1), so $Tv \in \Phi^{-1}(W)$.
- (d) With $\mathbf{x} = (0, -2, 1)$ and $\mathbf{y} = (-2, 1, 0)$: $\mathbf{Cx} = (2, -1, 0) = -\mathbf{y}$ and $\mathbf{Cy} = (0, -2, 1) = \mathbf{x}$. So $\mathbf{Cz} = \mathbf{Cx} + i\mathbf{Cy} = -\mathbf{y} + i\mathbf{x} = i(\mathbf{x} + i\mathbf{y}) = i\mathbf{z}$, and $\mathbf{z}$ is an eigenvector for i (here $a = 0, b = 1$, matching (a)). The invariant plane is $\text{span}((0, -2, 1), (-2, 1, 0))$, which is the plane $x + 2y + 4z = 0$. Over $\mathbb{R}$, $\mathbf{C}$ also has the eigenvalue 2 with eigenvector $(1, 0, 1)$, an invariant line.

Exercise. C3: Traces of powers detect nilpotency

Let $\mathbf{C} \in M_n(\mathbb{C})$, $n \geq 1$, with $\text{tr}(\mathbf{C}^k) = 0$ for $k = 1, \dots, n$. The goal is to prove that $\mathbf{C}^n = \mathbf{O}$.

- (a) Let $d_1, \dots, d_n$ be the diagonal entries of a triangular form of $\mathbf{C}$. Show that $d_1^k + \dots + d_n^k = 0$ for $k = 1, \dots, n$.
- (b) Suppose some $d_i \neq 0$. Let $\mu_1, \dots, \mu_r$ ($r \geq 1$) be the distinct nonzero values among the d_i , and let $m_j \geq 1$ be the number of indices i with $d_i = \mu_j$. Show that $\sum_{j=1}^r m_j \mu_j^k = 0$ for $k = 1, \dots, r$.
- (c) Show that the $r \times r$ matrix with (k, j) -entry μ_j^k is invertible, and derive a contradiction from (b).
- (d) Conclude that every $d_i = 0$, that $p_{\mathbf{C}} = x^n$, and that $\mathbf{C}^n = \mathbf{O}$.
- (e) Show that the statement fails over $\mathbb{F}_2$.

Hint: for (c), factor the matrix as a transposed Vandermonde matrix times a diagonal matrix.

Solution.

- (a) By Corollary 9.7.3 a triangular form exists, and by Exercise (B3: Eigenvalues of powers from a triangular form) (b), $\text{tr } \mathbf{C}^k = \sum_i d_i^k$, which is 0 by hypothesis.
- (b) Indices with $d_i = 0$ contribute nothing to $\sum_i d_i^k$ for $k \geq 1$. Grouping the other indices by their value, $\sum_i d_i^k = \sum_j m_j \mu_j^k$. Since $r \leq n$, (a) applies for $k = 1, \dots, r$.
- (c) Let $\mathbf{M}$ be the matrix with (k, j) -entry μ_j^k . The Vandermonde matrix $\mathbf{V} = \mathbf{V}(\mu_1, \dots, \mu_r)$ has (j, k) -entry μ_j^{k-1} , so $\mathbf{V}^{\top}$ has (k, j) -entry μ_j^{k-1} , and multiplying column j by μ_j gives $\mathbf{M} = \mathbf{V}^{\top} \text{diag}(\mu_1, \dots, \mu_r)$. Since the μ_j are distinct, $\mathbf{V}$ is invertible (Corollary 7.8.2), hence so is $\mathbf{V}^{\top}$ (Theorem 7.4.5); since the μ_j are nonzero, the diagonal matrix is invertible. So $\mathbf{M}$ is invertible. By (b), $\mathbf{M}(m_1, \dots, m_r) = \mathbf{0}$, so $(m_1, \dots, m_r) = \mathbf{0}$ in $\mathbb{C}^r$. But each m_j is a positive integer, which is nonzero in $\mathbb{C}$. This contradiction shows that no d_i is nonzero.
- (d) By (c), all $d_i = 0$, so $p_{\mathbf{C}} = x^n$ by Theorem 9.7.4 (a). By Theorem 9.6.1, $\mathbf{C}^n = p_{\mathbf{C}}(\mathbf{C}) = \mathbf{O}$.
- (e) Over $\mathbb{F}_2$, $\mathbf{C} = \mathbf{I}_2$ has $\text{tr } \mathbf{C}^k = \text{tr } \mathbf{I}_2 = 1 + 1 = 0$ for every k , but $\mathbf{C}^k = \mathbf{I}_2 \neq \mathbf{O}$. The step that fails is (c): the positive integer $m_1 = 2$ is 0 in $\mathbb{F}_2$. Compare Exercise (C2: When $\mathbf{A}$ commutes with $[\mathbf{A}, \mathbf{B}]$), where in characteristic 0 the same conclusion was reached

for a commutator by a route that avoids eigenvalues.