

12.1 Schur Triangularization

Chapter 9 proved that every operator on a nonzero complex vector space has an upper triangular matrix in **some** basis (Corollary 9.7.3). That basis was built by extending an eigenvector and passing to a quotient, and it came out as skewed as the operator happened to be. Now we have lengths and angles, so we can be fussier and ask for an **orthonormal** basis. The remarkable thing is that the answer is still yes, at no extra cost in hypotheses, and the proof is the same induction run inside an orthogonal complement rather than in a quotient. This section proves it, reads off the first consequences, and gives the real version, where a conjugate pair of eigenvalues costs one 2×2 block.

Throughout, V is a finite-dimensional inner product space over $F = \mathbb{R}$ or $F = \mathbb{C}$, and every matrix statement uses the standard inner product on F^n .

12.1.1 Why the obvious route fails

Here is the natural first attempt, and it is worth watching it fail, because the repair is the whole idea of the proof.

We want an orthonormal basis $\mathcal{B} = (v_1, \dots, v_n)$ with $[T]_{\mathcal{B}}^{\mathcal{B}}$ upper triangular, that is, with

$$Tv_j \in \text{span}(v_1, \dots, v_j) \quad \text{for every } j.$$

For $j = 1$ this says v_1 is an eigenvector of T , and over $\mathbb{C}$ one exists (Theorem 9.2.4). Normalize it, set $U = (\text{span}(v_1))^{\perp}$, and try to finish by induction on U . The attempt stops at once: to speak of T restricted to U we need U to be T -invariant, and it need not be.

Example. An Eigenline Whose Complement Is Not Invariant

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ act on $\mathbb{C}^2$. Check that $\mathbf{e}_1$ is an eigenvector and decide whether $(\text{span}(\mathbf{e}_1))^{\perp}$ is invariant.

Solution. $\mathbf{A}\mathbf{e}_1 = (1, 0) = \mathbf{e}_1$, so $\mathbf{e}_1$ is an eigenvector for the eigenvalue 1. The orthogonal complement of $\text{span}(\mathbf{e}_1)$ is $\text{span}(\mathbf{e}_2)$, and $\mathbf{A}\mathbf{e}_2 = (1, 2)$, which is not a multiple of $\mathbf{e}_2$. So $(\text{span}(\mathbf{e}_1))^{\perp}$ is **not** invariant, and the induction has nothing to run on.

So the first basis vector is the wrong end to start from. Look at the target instead. Suppose $\mathcal{B}$ is orthonormal and $[T]_{\mathcal{B}}^{\mathcal{B}} = \mathbf{T}$ is upper triangular. By Theorem 11.6.3 the matrix of T^* in the same basis is $\mathbf{T}^*$, which is **lower** triangular; and the last column of a lower triangular matrix has only its bottom entry. Reading that column off,

$$T^*v_n = \overline{t_{nn}}v_n.$$

In the basis we are trying to build, the last vector is an eigenvector of T^* — not of T .

That is the vector to produce first, and it is exactly the one whose orthogonal complement behaves.

Lemma 12.1.1. Eigenvectors of the Adjoint Cut Out Invariant Subspaces

Let V be a finite-dimensional inner product space over F , let $T \in \mathcal{L}(V)$, and let $W \subseteq V$ be a T^* -invariant subspace. Then $W^\perp$ is T -invariant.

Proof. Let $u \in W^\perp$ and $w \in W$. Since W is T^* -invariant, $T^*w \in W$, so by Definition 11.6.1

$$\langle Tu, w \rangle = \langle u, T^*w \rangle = 0.$$

As $w \in W$ was arbitrary, $Tu \in W^\perp$. Hence $T(W^\perp) \subseteq W^\perp$. ■

The special case we need is $W = \text{span}(v)$ with $T^*v = \mu v$: one vector, one line, and an invariant complement of dimension $n - 1$ to recurse into. And over $\mathbb{C}$ such a v always exists, because T^* is an operator on a nonzero complex space and the Fundamental Theorem of Algebra gives it an eigenvalue. That is where the field enters, and it is the only place it enters.

12.1.2 The theorem**Theorem 12.1.2. Schur's Unitary Triangularization**

Let V be a finite-dimensional inner product space over $\mathbb{C}$ with $\dim V = n \geq 1$, and let $T \in \mathcal{L}(V)$.

- (a) There is an **orthonormal** basis $\mathcal{B}$ of V such that $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular.
- (b) More precisely, let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of T , each listed as often as its algebraic multiplicity, in **any** order we please. Then $\mathcal{B}$ can be chosen so that the i -th diagonal entry of $[T]_{\mathcal{B}}^{\mathcal{B}}$ is λ_i .

Idea. Induction on $\dim V$, building the basis from its **last** vector backwards. ① Pick the eigenvalue λ_n we want in the bottom-right corner; the conjugate $\overline{\lambda_n}$ is then an eigenvalue of T^* , because starring turns a noninvertible operator into a noninvertible operator. ② Take a unit eigenvector e of T^* for $\overline{\lambda_n}$ and put $U = (\text{span}(e))^\perp$; by Lemma 12.1.1 it is T -invariant, so $T|_U$ exists and lives on a space of dimension $n - 1$. ③ Check that the characteristic polynomial of $T|_U$ is p_T with the factor $x - \lambda_n$ removed, so the inductive hypothesis may be applied with the shortened list. ④ Append e to the basis it returns.

Proof. We induct on $n = \dim V$. Since $\mathbb{C}$ is algebraically closed, p_T splits (Corollary 6.5.5), so a list $\lambda_1, \dots, \lambda_n$ as in (b) exists and $p_T = (x - \lambda_1) \cdots (x - \lambda_n)$; proving (b) proves (a).

For $n = 1$, any unit vector v_1 is a basis, every 1×1 matrix is upper triangular, and its single entry is the single eigenvalue λ_1 .

Let $n \geq 2$ and assume the statement for all inner product spaces of dimension $n - 1$. Fix a list $\lambda_1, \dots, \lambda_n$ as in (b).

$\overline{\lambda_n}$ is an eigenvalue of T^* . Since λ_n is an eigenvalue of T , the operator $T - \lambda_n \text{id}_V$ is not

invertible (Theorem 9.2.1). By Theorem 11.6.2 (a), (b), and (d), its adjoint is $T^* - \overline{\lambda_n} \text{id}_V$, and the adjoint of that is $T - \lambda_n \text{id}_V$ again. If $T^* - \overline{\lambda_n} \text{id}_V$ were invertible, then Theorem 11.6.2 (e) would make its adjoint $T - \lambda_n \text{id}_V$ invertible too. So $T^* - \overline{\lambda_n} \text{id}_V$ is not invertible, and $\overline{\lambda_n} \in \text{spec}(T^*)$.

Choose an eigenvector of T^* for $\overline{\lambda_n}$; it is nonzero, so dividing by its norm gives a unit vector e with $T^*e = \overline{\lambda_n}e$. Put $U = (\text{span}(e))^\perp$. By Theorem 11.3.2 (c), $\dim U = n - 1 \geq 1$, and U is T -invariant by Lemma 12.1.1. Write $S = T|_U \in \mathcal{L}(U)$, where U carries the inner product of V .

The bottom-right entry is λ_n . Using Definition 11.6.1 and conjugate-linearity in the second slot,

$$\langle Te, e \rangle = \langle e, T^*e \rangle = \langle e, \overline{\lambda_n}e \rangle = \lambda_n \|e\|^2 = \lambda_n.$$

The shortened list belongs to S . Take any orthonormal basis $(f_1, \dots, f_{n-1})$ of U (Theorem 11.2.2) and set $\mathcal{C} = (f_1, \dots, f_{n-1}, e)$, an orthonormal list of n vectors, hence an orthonormal basis of V by Theorem 11.1.4 and Theorem 1.6.4. Since U is T -invariant, Theorem 9.1.3 (b) applies with the basis $\mathcal{C}$ and gives $p_T = p_S p_{\bar{T}}$, where $\bar{T}$ is the induced operator on the one-dimensional quotient V/U . Its matrix is the single bottom-right entry of $[T]_{\mathcal{C}}^{\mathcal{C}}$, which is $\langle Te, e \rangle = \lambda_n$ by Theorem 11.2.1; so $p_{\bar{T}} = x - \lambda_n$. Canceling that factor in $\mathbb{C}[x]$,

$$p_S = (x - \lambda_1) \cdots (x - \lambda_{n-1}).$$

So $\lambda_1, \dots, \lambda_{n-1}$ is a list of the eigenvalues of S with their algebraic multiplicities.

Induct. By the inductive hypothesis there is an orthonormal basis $(v_1, \dots, v_{n-1})$ of U in which the matrix of S is upper triangular with diagonal $\lambda_1, \dots, \lambda_{n-1}$. Put $\mathcal{B} = (v_1, \dots, v_{n-1}, e)$, again an orthonormal basis of V for the reason given above. For $j \leq n-1$ we have $Tv_j = Sv_j \in \text{span}(v_1, \dots, v_j)$, so the j -th column of $[T]_{\mathcal{B}}^{\mathcal{B}}$ vanishes below row j and has λ_j in row j . The n -th column has nothing below row n , and its entry there is $\langle Te, e \rangle = \lambda_n$. Hence $[T]_{\mathcal{B}}^{\mathcal{B}}$ is upper triangular with diagonal $\lambda_1, \dots, \lambda_n$. This proves the theorem. ■

Two hypotheses did all the work, and it is worth naming which. **Complex** produced the eigenvalue of T^* , through the Fundamental Theorem of Algebra; over $\mathbb{R}$ the theorem is false as stated, and the last part of this section says what survives. **Inner product** produced the invariant complement, through Lemma 12.1.1; without one we would be back in Chapter 9's quotient. Notice also what the theorem does **not** claim: nothing was assumed about T , and nothing beyond triangularity is promised. Diagonal is a different and much stronger request, and Sections 4 and 5 say exactly which operators grant it.

12.1.3 The matrix form

Translating through an orthonormal basis turns the theorem into a factorization. Recall that a square complex matrix is **unitary** when $\mathbf{U}^* \mathbf{U} = \mathbf{I}$ (Definition 11.7.2), and call two matrices **unitarily similar** when $\mathbf{B} = \mathbf{U}^* \mathbf{A} \mathbf{U}$ for some unitary $\mathbf{U}$; this is similarity by a matrix that is also an isometry, so it preserves the inner product as well as the operator.

Corollary 12.1.3. Schur Factorization

Let $\mathbf{A} \in M_n(\mathbb{C})$ with $n \geq 1$, and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of $\mathbf{A}$, each listed as often as its algebraic multiplicity, in any prescribed order. Then there are a unitary $\mathbf{U} \in M_n(\mathbb{C})$ and an upper triangular $\mathbf{T} \in M_n(\mathbb{C})$ with

$$\mathbf{A} = \mathbf{UTU}^*, \quad t_{ii} = \lambda_i \quad (1 \leq i \leq n).$$

Proof. Apply [Theorem 12.1.2](#) to $T_{\mathbf{A}}$ on $\mathbb{C}^n$ with the standard inner product and the given list: there is an orthonormal basis $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ of $\mathbb{C}^n$ with $\mathbf{T} := [T_{\mathbf{A}}]_{\mathcal{B}}^{\mathcal{B}}$ upper triangular and $t_{ii} = \lambda_i$. Let $\mathbf{U}$ be the matrix whose j -th column is $\mathbf{u}_j$. Its columns are orthonormal, so $\mathbf{U}$ is unitary (Theorem 11.7.3 (f), Definition 11.7.2), and $\mathbf{U} = [\text{id}]_{\mathcal{B}}^{\mathcal{E}}$ for the standard basis $\mathcal{E}$ (Definition 4.2.1). Since $[T_{\mathbf{A}}]_{\mathcal{E}}^{\mathcal{E}} = \mathbf{A}$, Theorem 4.2.3 gives $\mathbf{T} = \mathbf{U}^{-1}\mathbf{AU} = \mathbf{U}^*\mathbf{AU}$, the last step because $\mathbf{U}^{-1} = \mathbf{U}^*$ for a unitary matrix. Rearranging gives $\mathbf{A} = \mathbf{UTU}^*$. ■

The order of the diagonal is free because at each step of the induction we chose which eigenvalue to send to the bottom-right corner, and any of them was available. That freedom is used constantly: it lets us put a distinguished eigenvalue first, or collect equal eigenvalues into a block.

Remark. How to compute one. The induction above runs from the bottom-right corner, because that is the corner where an operator on a subspace is available. For a **matrix** there is a second route, running from the top-left, and it is the one to use by hand. Let $\mathbf{q}_1$ be a unit eigenvector of $\mathbf{A}$ for an eigenvalue λ , extend it to an orthonormal basis of $\mathbb{C}^n$ (Corollary 11.2.3) and let $\mathbf{Q}$ be the unitary matrix with those columns. Then $\mathbf{Q}^*\mathbf{A}\mathbf{q}_1 = \lambda\mathbf{Q}^*\mathbf{q}_1 = \lambda\mathbf{e}_1$, so

$$\mathbf{Q}^*\mathbf{A}\mathbf{Q} = \begin{pmatrix} \lambda & \mathbf{b}^* \\ \mathbf{0} & \mathbf{A}_1 \end{pmatrix}$$

for some $\mathbf{b} \in \mathbb{C}^{n-1}$ and $\mathbf{A}_1 \in M_{n-1}(\mathbb{C})$. Triangularize $\mathbf{A}_1 = \mathbf{VT}_1\mathbf{V}^*$ and put $\mathbf{U} = \mathbf{Q}(1 \oplus \mathbf{V})$; the product is unitary and $\mathbf{U}^*\mathbf{A}\mathbf{U}$ is upper triangular. This is legitimate for matrices precisely because we may replace $\mathbf{A}$ by the compressed block $\mathbf{A}_1$, which is not the restriction of any operator: $(\text{span}(\mathbf{q}_1))^{\perp}$ is not invariant, as [Example \(An Eigenline Whose Complement Is Not Invariant\)](#) showed.

Example. A Two-by-Two Schur Factorization

Find a unitary $\mathbf{U}$ and an upper triangular $\mathbf{T}$ with $\mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{T}$, as in [Corollary 12.1.3](#), for

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ 1 & 4 \end{pmatrix}.$$

Solution. Here $\text{tr } \mathbf{A} = 6$ and $\det \mathbf{A} = 8 + 1 = 9$, so $p_{\mathbf{A}} = x^2 - 6x + 9 = (x - 3)^2$ and the only eigenvalue is 3, with algebraic multiplicity 2.

Solving $(\mathbf{A} - 3\mathbf{I})\mathbf{x} = \mathbf{0}$ with $\mathbf{A} - 3\mathbf{I} = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}$ gives the single eigenline $\text{span}((1, -1))$.

Take $\mathbf{q}_1 = \frac{1}{\sqrt{2}}(1, -1)$ and complete with $\mathbf{q}_2 = \frac{1}{\sqrt{2}}(1, 1)$, a unit vector orthogonal to $\mathbf{q}_1$. With $\mathbf{U}$ the matrix of those columns,

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{U}^* \mathbf{A} \mathbf{U} = \begin{pmatrix} 3 & -2 \\ 0 & 3 \end{pmatrix}.$$

Both diagonal entries are 3, as they must be. The off-diagonal -2 cannot be removed: a matrix unitarily similar to $\text{diag}(3, 3) = 3\mathbf{I}$ would equal $\mathbf{U}(3\mathbf{I})\mathbf{U}^* = 3\mathbf{I}$, and $\mathbf{A} \neq 3\mathbf{I}$.

Example. A Three-by-Three Schur Factorization

Find a unitary $\mathbf{U}$ and an upper triangular $\mathbf{T}$ with $\mathbf{U}^* \mathbf{A} \mathbf{U} = \mathbf{T}$, as in [Corollary 12.1.3](#), for

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 4 & -3 & -2 \end{pmatrix}.$$

Solution. Step 1: peel off one eigenvector. Expanding $\det(x\mathbf{I} - \mathbf{A})$ gives $p_{\mathbf{A}} = (x - 1)(x - 3)(x + 4)$, so 1 is an eigenvalue. Solving $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$ by row reduction gives the eigenline spanned by $(3, 4, 0)$, of norm 5. Put

$$\mathbf{q}_1 = \frac{1}{5}(3, 4, 0), \quad \mathbf{q}_2 = \frac{1}{5}(4, -3, 0), \quad \mathbf{q}_3 = (0, 0, 1),$$

an orthonormal basis of $\mathbb{C}^3$ (each has norm 1 and the three pairwise inner products are 0; the entries are real, so no conjugates intrude). With $\mathbf{Q} = (\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3)$,

$$\mathbf{Q}^* \mathbf{A} \mathbf{Q} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 5 & -2 \end{pmatrix}, \quad \mathbf{A}_1 = \begin{pmatrix} 1 & 2 \\ 5 & -2 \end{pmatrix}.$$

Step 2: triangularize the block. Here $\text{tr } \mathbf{A}_1 = -1$ and $\det \mathbf{A}_1 = -12$, so $p_{\mathbf{A}_1} = x^2 + x - 12 = (x - 3)(x + 4)$. For the eigenvalue 3, $\mathbf{A}_1 - 3\mathbf{I} = \begin{pmatrix} -2 & 2 \\ 5 & -5 \end{pmatrix}$ gives the eigenvector $(1, 1)$. Take $\mathbf{p}_1 = \frac{1}{\sqrt{2}}(1, 1)$ and $\mathbf{p}_2 = \frac{1}{\sqrt{2}}(1, -1)$, and let $\mathbf{V}$ be the matrix of those columns. Then

$$\mathbf{V}^* \mathbf{A}_1 \mathbf{V} = \begin{pmatrix} 3 & 3 \\ 0 & -4 \end{pmatrix}.$$

Step 3: assemble. Put $\mathbf{U} = \mathbf{Q}(1 \oplus \mathbf{V})$, whose columns are $\mathbf{q}_1$, $(\mathbf{q}_2 + \mathbf{q}_3)/\sqrt{2}$, and $(\mathbf{q}_2 - \mathbf{q}_3)/\sqrt{2}$:

$$\mathbf{U} = \frac{1}{5\sqrt{2}} \begin{pmatrix} 3\sqrt{2} & 4 & 4 \\ 4\sqrt{2} & -3 & -3 \\ 0 & 5 & -5 \end{pmatrix}, \quad \mathbf{U}^* \mathbf{A} \mathbf{U} = \begin{pmatrix} 1 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 3 & 3 \\ 0 & 0 & -4 \end{pmatrix}.$$

The diagonal reads 1, 3, -4 , and indeed $p_{\mathbf{A}} = (x - 1)(x - 3)(x + 4)$, with trace $0 = 1 + 3 - 4$ and determinant $-12 = 1 \cdot 3 \cdot (-4)$.

Quick check. Why does [Theorem 12.1.2](#) not prove that every complex matrix is unitarily similar to a **diagonal** matrix?

Solution. Because the proof controls only the entries below the diagonal. The entries above it are whatever they happen to be, and [Example \(A Two-by-Two Schur Factorization\)](#) shows they need not vanish: the matrix there has a -2 in the corner, and no unitary can remove it. A matrix unitarily similar to a diagonal one satisfies strong conditions — Section 4 identifies them exactly — and most matrices do not.

12.1.4 Reading the diagonal

A triangular form with an explicit factorization makes two Chapter 9 facts fall out in a line each.

Corollary 12.1.4. Trace Is the Sum of the Eigenvalues

Let $\mathbf{A} \in M_n(\mathbb{C})$ have eigenvalues $\lambda_1, \dots, \lambda_n$, listed with algebraic multiplicity. Then $\operatorname{tr} \mathbf{A} = \lambda_1 + \dots + \lambda_n$.

Proof. Write $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ as in [Corollary 12.1.3](#). By Theorem 0.9.6 (3) applied to the pair $\mathbf{U}$ and $\mathbf{T}\mathbf{U}^*$,

$$\operatorname{tr} \mathbf{A} = \operatorname{tr}(\mathbf{U}(\mathbf{T}\mathbf{U}^*)) = \operatorname{tr}((\mathbf{T}\mathbf{U}^*)\mathbf{U}) = \operatorname{tr} \mathbf{T},$$

since $\mathbf{U}^*\mathbf{U} = \mathbf{I}$. The diagonal of $\mathbf{T}$ is the list $\lambda_1, \dots, \lambda_n$. ■

Corollary 12.1.5. Determinant Is the Product of the Eigenvalues

Let $\mathbf{A} \in M_n(\mathbb{C})$ have eigenvalues $\lambda_1, \dots, \lambda_n$, listed with algebraic multiplicity. Then $\det \mathbf{A} = \lambda_1 \cdots \lambda_n$.

Proof. With $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ as above, Theorem 7.5.4 gives $\det \mathbf{A} = \det \mathbf{U} \det \mathbf{T} \det \mathbf{U}^*$, and $\det \mathbf{U} \det \mathbf{U}^* = \det(\mathbf{U}\mathbf{U}^*) = \det \mathbf{I} = 1$. So $\det \mathbf{A} = \det \mathbf{T}$, which is the product of the diagonal entries by Theorem 7.4.6. ■

Both statements are Theorem 9.3.3 over $\mathbb{C}$, recovered from the diagonal of $\mathbf{T}$ instead of from the coefficients of $p_{\mathbf{A}}$. Nothing is new; what is new is how short the argument became once a triangular form was available in hand.

The next consequence is not a restatement of anything. It says that the matrices we know how to handle sit arbitrarily close to the ones we do not.

Corollary 12.1.6. Diagonalizable Matrices Are Everywhere

Let $\mathbf{A} \in M_n(\mathbb{C})$ and let $\varepsilon > 0$. Then there is a **diagonalizable** $\mathbf{A}' \in M_n(\mathbb{C})$ with

$$|a_{ij} - a'_{ij}| < \varepsilon \quad \text{for all } i, j.$$

In this sense every complex square matrix is a limit of diagonalizable ones.

Idea. Diagonalizability is hard to arrange directly, but n distinct eigenvalues force it (Corollary 9.4.2), and in a triangular form the eigenvalues are sitting on the diagonal where we can nudge them apart. The only thing to check is that a small nudge inside the triangular form is still a small nudge after conjugating back by $\mathbf{U}$, and that is Cauchy–Schwarz on the rows of a unitary matrix.

Proof. Write $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ as in Corollary 12.1.3. Choose $\delta_1, \dots, \delta_n \in \mathbb{C}$ with $|\delta_k| < \varepsilon$ and with $t_{11} + \delta_1, \dots, t_{nn} + \delta_n$ pairwise distinct. This is possible one index at a time: having chosen $\delta_1, \dots, \delta_{k-1}$, the forbidden values of δ_k are the $k-1$ numbers $t_{jj} + \delta_j - t_{kk}$, and the disc $\{|z| < \varepsilon\}$ has infinitely many points.

Put $\mathbf{D} = \text{diag}(\delta_1, \dots, \delta_n)$ and $\mathbf{A}' = \mathbf{U}(\mathbf{T} + \mathbf{D})\mathbf{U}^*$. The matrix $\mathbf{T} + \mathbf{D}$ is upper triangular with n distinct diagonal entries, so by Theorem 9.7.4 (b) it has n distinct eigenvalues; $\mathbf{A}'$ is similar to it, hence has the same eigenvalues, and is diagonalizable by Corollary 9.4.2.

For the entries, $\mathbf{A} - \mathbf{A}' = -\mathbf{U}\mathbf{D}\mathbf{U}^*$, whose (i, j) entry is $-\sum_k u_{ik}\delta_k\overline{u_{jk}}$. Hence

$$|a_{ij} - a'_{ij}| \leq \max_k |\delta_k| \sum_{k=1}^n |u_{ik}| |u_{jk}| \leq \max_k |\delta_k| < \varepsilon,$$

where the middle inequality is Theorem 11.1.5 applied to the vectors of moduli, together with the fact that the rows of $\mathbf{U}$ are unit vectors, which is $\mathbf{U}\mathbf{U}^* = \mathbf{I}$. This proves the corollary. ■

This is the tool behind a standard strategy: *prove it for diagonalizable matrices, then take a limit*. Chapter 16, where matrices acquire norms and limits can be spoken of properly, runs arguments of exactly this shape; and Exercise (C1: Cayley–Hamilton from the triangular factors) uses the triangular form directly, with no limits at all, to reprove the Cayley–Hamilton theorem over $\mathbb{C}$.

12.1.5 The real version

Over $\mathbb{R}$ the theorem fails, and it fails for the reason it always does: a rotation has no real eigenvector, so no real basis makes its matrix triangular (Theorem 9.7.2). The most one can hope for is to let a conjugate pair occupy a 2×2 block, and that much is true.

Theorem 12.1.7. Real Schur Form

Let $\mathbf{A} \in M_n(\mathbb{R})$ with $n \geq 1$. Then there are an orthogonal $\mathbf{Q} \in O(n)$ and a matrix $\mathbf{T} \in M_n(\mathbb{R})$ with

$$\mathbf{A} = \mathbf{Q}\mathbf{T}\mathbf{Q}^\top,$$

where $\mathbf{T}$ is **block** upper triangular with diagonal blocks of size 1×1 and 2×2 , and every 2×2 diagonal block has a conjugate pair of nonreal eigenvalues.

Idea. The same induction, with one change. The eigenvector of T^* may fail to exist over $\mathbb{R}$, but an invariant subspace of dimension 1 or 2 always does — that is Exercise (C2: Real operators have small invariant subspaces), proved in Chapter 9 from a complex eigenvector $\mathbf{x} + i\mathbf{y}$ by taking $\text{span}(\mathbf{x}, \mathbf{y})$. Apply it to T^* , take the orthogonal complement, and induct. Dimension 1 occurs exactly when T^* has a real eigenvalue, and that is what keeps the 2×2 blocks honest.

Proof. Let $T = T_{\mathbf{A}}$ on $\mathbb{R}^n$ with the standard inner product, so that $T^* = T_{\mathbf{A}^\top}$ by Theorem 11.6.3. We induct on n .

Choose a T^* -invariant subspace W as follows. If T^* has a real eigenvalue, let W be a corresponding eigenline, so $d := \dim W = 1$. If it does not, take a complex eigenvalue $\mu = a + bi$ of $\mathbf{A}^\top$ with $b \neq 0$, which exists by Theorem 9.2.4 applied to $\mathbf{A}^\top$ over $\mathbb{C}$, and a complex eigenvector $\mathbf{x} + i\mathbf{y}$ with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Comparing real and imaginary parts of $\mathbf{A}^\top(\mathbf{x} + i\mathbf{y}) = \mu(\mathbf{x} + i\mathbf{y})$ gives $\mathbf{A}^\top \mathbf{x} = a\mathbf{x} - b\mathbf{y}$ and $\mathbf{A}^\top \mathbf{y} = b\mathbf{x} + a\mathbf{y}$, so $W := \text{span}(\mathbf{x}, \mathbf{y})$ is T^* -invariant. It has dimension 2: were $\mathbf{x}, \mathbf{y}$ dependent, W would be a T^* -invariant line and T^* would have a real eigenvalue, contrary to the case we are in. So $d = 2$. (Exercise (C2: Real operators have small invariant subspaces) runs the same argument in Chapter 9.)

Put $U = W^\perp$, which is T -invariant by Lemma 12.1.1 and has dimension $n - d$ (Theorem 11.3.2 (c)). Pick an orthonormal basis $(\mathbf{v}_1, \dots, \mathbf{v}_{n-d})$ of U and one $(\mathbf{v}_{n-d+1}, \dots, \mathbf{v}_n)$ of W (Theorem 11.2.2); together they form an orthonormal basis $\mathcal{B}$ of $\mathbb{R}^n$, since the two blocks are orthogonal to each other. Because $T(U) \subseteq U$,

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} \mathbf{C} & * \\ \mathbf{O} & \mathbf{B} \end{pmatrix}$$

with $\mathbf{C} \in M_{n-d}(\mathbb{R})$ the matrix of $T|_U$ and $\mathbf{B} \in M_d(\mathbb{R})$. Writing $\mathbf{Q}_{\mathcal{B}}$ for the matrix whose columns are the vectors of $\mathcal{B}$, which is orthogonal because those columns are orthonormal, Theorem 4.2.3 gives $\mathbf{Q}_{\mathcal{B}}^\top \mathbf{A} \mathbf{Q}_{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{B}}$.

The block $\mathbf{B}$ is of the allowed kind. If $d = 1$ it is 1×1 . If $d = 2$, then by Theorem 11.6.3 the matrix of T^* in $\mathcal{B}$ is the transpose of the display, whose bottom-right block is $\mathbf{B}^\top$; as W is T^* -invariant, $\mathbf{B}^\top$ is the matrix of $T^*|_W$. A real eigenvalue of $\mathbf{B}^\top$ would be a real eigenvalue of T^* , which this case excludes. So $p_{\mathbf{B}} = p_{\mathbf{B}^\top}$ has no real root, and its two complex roots form a conjugate pair by Theorem 9.2.5.

Close the induction. If $U = \{\mathbf{0}\}$ then $n = d \leq 2$, the list $\mathcal{B}$ is a basis of W , and $\mathbf{Q}_{\mathcal{B}}^\top \mathbf{A} \mathbf{Q}_{\mathcal{B}} = \mathbf{B}$ is already a single allowed block; take $\mathbf{Q} = \mathbf{Q}_{\mathcal{B}}$. Otherwise $1 \leq n - d < n$, and by the inductive hypothesis $\mathbf{C} = \mathbf{Q}_1 \mathbf{T}_1 \mathbf{Q}_1^\top$ with $\mathbf{Q}_1 \in O(n - d)$ and $\mathbf{T}_1$ of the required shape. Put $\mathbf{Q} = \mathbf{Q}_{\mathcal{B}}(\mathbf{Q}_1 \oplus \mathbf{I}_d)$, a product of orthogonal matrices and so orthogonal (Proposition 11.7.4 (a)). Then

$$\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = (\mathbf{Q}_1 \oplus \mathbf{I}_d)^\top \begin{pmatrix} \mathbf{C} & * \\ \mathbf{O} & \mathbf{B} \end{pmatrix} (\mathbf{Q}_1 \oplus \mathbf{I}_d) = \begin{pmatrix} \mathbf{T}_1 & * \\ \mathbf{O} & \mathbf{B} \end{pmatrix},$$

which is block upper triangular with diagonal blocks those of $\mathbf{T}_1$ followed by $\mathbf{B}$. This proves the theorem.

Two remarks on the shape. The 1×1 blocks carry the real eigenvalues of $\mathbf{A}$ and the 2×2 blocks carry the nonreal ones in conjugate pairs, so the block sizes are forced by the spectrum, even though the blocks themselves are not. And when all the eigenvalues of $\mathbf{A}$ are real, every block is 1×1 and the real Schur form is an honest orthogonal triangularization.

Warning. Neither $\mathbf{T}$ nor $\mathbf{U}$ is unique; only the diagonal multiset is. For $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, taking $\mathbf{U} = \mathbf{I}$ gives $\mathbf{T} = \mathbf{A}$, with diagonal $(1, 2)$. Taking instead $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ gives

$$\mathbf{U}^* \mathbf{A} \mathbf{U} = \begin{pmatrix} 2 & -1 \\ 0 & 1 \end{pmatrix},$$

with diagonal $(2, 1)$ and a different corner entry. So a Schur form is **not** a canonical form: it does not decide similarity, and two matrices can have the same triangular form without being equal. The canonical form for similarity is Jordan's (Theorem 10.3.1), and Jordan's cannot in general be reached unitarily — [Example \(A Two-by-Two Schur Factorization\)](#) is already a Jordan block in disguise whose off-diagonal entry is -2 , not 1 . What the Schur form gives up in uniqueness it buys back in stability: the change of basis is an isometry.

12.1.6 Exercises

12.1.6.1 A. Check your understanding

Exercise. A1

- State [Theorem 12.1.2](#), including the field and every hypothesis.
- In the proof, an eigenvector of T^* is taken rather than one of T . Say why, in one sentence, and name the property of $(\text{span}(v))^\perp$ that is wanted.
- True or false: if $\mathbf{A} \in M_n(\mathbb{C})$ is upper triangular in some orthonormal basis, then its eigenvalues are the diagonal entries of that triangular matrix. Justify your answer.
- Which step of the proof of [Theorem 12.1.2](#) fails over $\mathbb{R}$?
- Give a matrix in $M_2(\mathbb{C})$ that is upper triangular in an orthonormal basis but is **not** diagonal in any.

Solution.

- Let V be a finite-dimensional inner product space over $\mathbb{C}$ with $\dim V \geq 1$ and let $T \in \mathcal{L}(V)$. Then V has an orthonormal basis $\mathcal{B}$ with $[T]_{\mathcal{B}}^{\mathcal{B}}$ upper triangular, and the diagonal may be made to read the eigenvalues of T , with algebraic multiplicity, in any prescribed order.
- Because the induction needs $(\text{span}(v))^\perp$ to be T -invariant, and [Lemma 12.1.1](#) gives that exactly when $\text{span}(v)$ is T^* -invariant, that is, when v is an eigenvector of T^* .
- True. The matrix of T in **any** basis has the same characteristic polynomial (The-

orem 7.10.3), and for a triangular matrix that polynomial is the product of $x - t_{ii}$ (Theorem 9.7.4 (a)). Orthonormality is irrelevant to this particular point.

- (d) The Fundamental Theorem of Algebra, twice: once so that p_T splits into the list of n roots the statement speaks of (Corollary 6.5.5), and once for the existence of an eigenvalue of T^* at each stage of the induction. It came from Theorem 9.2.4, whose proof rests on the Fundamental Theorem of Algebra (Theorem 6.5.1); over $\mathbb{R}$ the rotation $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has no eigenvalue at all. Every other step is valid over $\mathbb{R}$.
- (e) $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. It is already upper triangular in the standard orthonormal basis. It is not diagonal in any basis, orthonormal or not: its only eigenvalue is 0, so a diagonal form would be the zero matrix, and $\mathbf{A} \neq \mathbf{O}$.

12.1.6.2 B. Practice

Exercise. B1: A Schur form with complex entries

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$, regarded as a complex matrix.

- (a) Find the eigenvalues of $\mathbf{A}$.
- (b) Find a unitary $\mathbf{U} \in M_2(\mathbb{C})$ with $\mathbf{U}^* \mathbf{A} \mathbf{U}$ upper triangular, and compute that triangular matrix.
- (c) Explain why no **real** orthogonal $\mathbf{Q}$ makes $\mathbf{Q}^\top \mathbf{A} \mathbf{Q}$ upper triangular.

Solution.

- (a) $\text{tr } \mathbf{A} = 4$ and $\det \mathbf{A} = 3 + 2 = 5$, so $p_{\mathbf{A}} = x^2 - 4x + 5 = (x - 2)^2 + 1$, with roots $2 + i$ and $2 - i$.
- (b) For $\lambda = 2 + i$,

$$\mathbf{A} - \lambda \mathbf{I} = \begin{pmatrix} -1 - i & 2 \\ -1 & 1 - i \end{pmatrix}.$$

The second row gives $-x + (1 - i)y = 0$, so $(1 - i, 1)$ is an eigenvector; its norm is $\sqrt{2 + 1} = \sqrt{3}$. Put $\mathbf{q}_1 = \frac{1}{\sqrt{3}}(1 - i, 1)$. A unit vector orthogonal to it is $\mathbf{q}_2 = \frac{1}{\sqrt{3}}(1, -1 - i)$, since

$$\langle \mathbf{q}_2, \mathbf{q}_1 \rangle = \frac{1}{3}(1 \cdot \overline{1 - i} + (-1 - i) \cdot \overline{1}) = \frac{1}{3}((1 + i) - (1 + i)) = 0$$

and $\|\mathbf{q}_2\|^2 = \frac{1}{3}(1 + 2) = 1$. With $\mathbf{U} = (\mathbf{q}_1 \quad \mathbf{q}_2)$,

$$\mathbf{U}^* \mathbf{A} \mathbf{U} = \begin{pmatrix} 2 + i & -1 - 2i \\ 0 & 2 - i \end{pmatrix}.$$

- (c) A real upper triangular matrix has its diagonal entries as eigenvalues (Theorem 9.7.4), and those entries are real. But $\mathbf{Q}^\top \mathbf{A} \mathbf{Q}$ is similar to $\mathbf{A}$, whose eigenvalues $2 \pm i$ are not real. So no such $\mathbf{Q}$ exists. (This is [Theorem 12.1.7](#) doing the only thing it can: here $\mathbf{T}$ is a single 2×2 block, and $\mathbf{Q} = \mathbf{I}$ already works.)

Exercise. B2: A three-by-three Schur form

Let $\mathbf{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 3 \\ 0 & 0 & 4 \end{pmatrix}$.

- (a) Verify that $(1, -1, 0)$ is an eigenvector and find the corresponding eigenvalue.
- (b) Extend it to an orthonormal basis of $\mathbb{R}^3$ and compute $\mathbf{Q}^\top \mathbf{A} \mathbf{Q}$ for the resulting $\mathbf{Q}$.
- (c) Hence write down the eigenvalues of $\mathbf{A}$ with their algebraic multiplicities.

Solution.

- (a) $\mathbf{A}(1, -1, 0) = (2 - 1, 1 - 2, 0) = (1, -1, 0)$, so the eigenvalue is 1.
- (b) Take $\mathbf{q}_1 = \frac{1}{\sqrt{2}}(1, -1, 0)$, $\mathbf{q}_2 = \frac{1}{\sqrt{2}}(1, 1, 0)$, and $\mathbf{q}_3 = (0, 0, 1)$. These are pairwise orthogonal unit vectors, hence an orthonormal basis of $\mathbb{R}^3$. With $\mathbf{Q} = (\mathbf{q}_1 \ \mathbf{q}_2 \ \mathbf{q}_3)$,

$$\mathbf{Q}^\top \mathbf{A} \mathbf{Q} = \begin{pmatrix} 1 & 0 & -\sqrt{2} \\ 0 & 3 & 2\sqrt{2} \\ 0 & 0 & 4 \end{pmatrix}.$$

This is already upper triangular, so a single step sufficed and $\mathbf{U} = \mathbf{Q}$. (In general the lower-right 2×2 block would need a second step, as in [Example \(A Three-by-Three Schur Factorization\)](#).)

- (c) The diagonal is 1, 3, 4, so by Theorem 9.7.4 these are the eigenvalues, each of algebraic multiplicity 1. As a check, $\text{tr } \mathbf{A} = 8 = 1 + 3 + 4$ and $\det \mathbf{A} = 4(4 - 1) = 12 = 1 \cdot 3 \cdot 4$.

Exercise. B3: The field matters

Let $\mathbf{A} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \in M_2(\mathbb{R})$ with $b \neq 0$.

- (a) Find the eigenvalues of $\mathbf{A}$ over $\mathbb{C}$ and show that $\mathbf{A}$ is unitarily similar to a diagonal matrix in $M_2(\mathbb{C})$.
- (b) Prove that there is no invertible $\mathbf{P} \in M_2(\mathbb{R})$ at all, orthogonal or otherwise, with $\mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ upper triangular.
- (c) Say what [Theorem 12.1.7](#) produces for this $\mathbf{A}$.

Solution.

- (a) $\text{tr } \mathbf{A} = 2a$ and $\det \mathbf{A} = a^2 + b^2$, so $p_{\mathbf{A}} = x^2 - 2ax + a^2 + b^2$ with roots $a \pm bi$, distinct because $b \neq 0$. For $\lambda = a + bi$, $\mathbf{A}(1, -i) = (a + bi, b - ai) = (a + bi)(1, -i)$, so $\mathbf{q}_1 = \frac{1}{\sqrt{2}}(1, -i)$ is a unit eigenvector; likewise $\mathbf{q}_2 = \frac{1}{\sqrt{2}}(1, i)$ is a unit eigenvector for $a - bi$, and

$$\langle \mathbf{q}_1, \mathbf{q}_2 \rangle = \frac{1}{2}(1 \cdot 1 + (-i)\bar{i}) = \frac{1}{2}(1 - 1) = 0.$$

So $\mathbf{U} = (\mathbf{q}_1 \ \mathbf{q}_2)$ is unitary and $\mathbf{U}^* \mathbf{A} \mathbf{U} = \text{diag}(a + bi, a - bi)$.

- (b) Suppose $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{T}$ were upper triangular with $\mathbf{P} \in M_2(\mathbb{R})$ invertible. Then $\mathbf{T} \in M_2(\mathbb{R})$ is similar to $\mathbf{A}$, so $p_{\mathbf{T}} = p_{\mathbf{A}}$ (Theorem 7.10.3); but $p_{\mathbf{T}} = (x - t_{11})(x - t_{22})$ has the real roots t_{11}, t_{22} , while $p_{\mathbf{A}}$ has none. Contradiction.
- (c) Since $\mathbf{A}$ has no real eigenvalue, the real Schur form of $\mathbf{A}$ is a single 2×2 diagonal block; $\mathbf{Q} = \mathbf{I}_2$ and $\mathbf{T} = \mathbf{A}$ already work. There is nothing to simplify: over $\mathbb{R}$ this matrix is as reduced as it gets.

12.1.6.3 C. Going deeper

Exercise. C1: Cayley–Hamilton from the triangular factors

Let $\mathbf{T} \in M_n(\mathbb{C})$ be upper triangular with diagonal entries $\lambda_1, \dots, \lambda_n$, and set $\mathbf{N}_k = (\mathbf{T} - \lambda_1 \mathbf{I})(\mathbf{T} - \lambda_2 \mathbf{I}) \cdots (\mathbf{T} - \lambda_k \mathbf{I})$ for $1 \leq k \leq n$.

- (a) Prove by induction on k that the first k columns of $\mathbf{N}_k$ are zero.
- (b) Deduce that $p_{\mathbf{T}}(\mathbf{T}) = \mathbf{O}$.
- (c) Deduce the Cayley–Hamilton Theorem (Theorem 9.6.1) for every $\mathbf{A} \in M_n(\mathbb{C})$, using [Corollary 12.1.3](#).

Hint for (a): the j -th column of $\mathbf{T} - \lambda_j \mathbf{I}$ has its only possibly nonzero entries in rows $1, \dots, j-1$.

Solution.

- (a) For $k = 1$: the first column of $\mathbf{T} - \lambda_1 \mathbf{I}$ is $(\lambda_1 - \lambda_1, 0, \dots, 0) = \mathbf{0}$, since $\mathbf{T}$ is upper triangular. Assume the first $k-1$ columns of $\mathbf{N}_{k-1}$ are zero, where $2 \leq k \leq n$. Then $\mathbf{N}_k = \mathbf{N}_{k-1}(\mathbf{T} - \lambda_k \mathbf{I})$, so the j -th column of $\mathbf{N}_k$ is $\mathbf{N}_{k-1} \mathbf{c}_j$ with $\mathbf{c}_j$ the j -th column of $\mathbf{T} - \lambda_k \mathbf{I}$. For $j \leq k-1$ we get $\mathbf{N}_k \mathbf{e}_j = \mathbf{N}_{k-1}(\mathbf{T} - \lambda_k \mathbf{I}) \mathbf{e}_j$, and since $(\mathbf{T} - \lambda_k \mathbf{I}) \mathbf{e}_j \in \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_j)$ by upper triangularity, this is a combination of the first $j \leq k-1$ columns of $\mathbf{N}_{k-1}$, all zero. For $j = k$ the column $\mathbf{c}_k$ of $\mathbf{T} - \lambda_k \mathbf{I}$ has its (k, k) entry equal to $\lambda_k - \lambda_k = 0$ and zeros below, so $\mathbf{c}_k \in \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_{k-1})$, and $\mathbf{N}_{k-1} \mathbf{c}_k$ is a combination of the first $k-1$ columns of $\mathbf{N}_{k-1}$, again zero. So all of the first k columns of $\mathbf{N}_k$ vanish.
- (b) By Theorem 9.7.4 (a), $p_{\mathbf{T}} = (x - \lambda_1) \cdots (x - \lambda_n)$, so $p_{\mathbf{T}}(\mathbf{T}) = \mathbf{N}_n$. By (a) with $k = n$, all n columns of $\mathbf{N}_n$ are zero, that is $p_{\mathbf{T}}(\mathbf{T}) = \mathbf{O}$.
- (c) Write $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ by [Corollary 12.1.3](#). Similar matrices have the same characteristic polynomial (Theorem 7.10.3), so $p_{\mathbf{A}} = p_{\mathbf{T}}$, and $\mathbf{A}^k = \mathbf{U}\mathbf{T}^k\mathbf{U}^*$ for every $k \geq 0$ because the inner factors $\mathbf{U}^*\mathbf{U}$ cancel. Hence

$$p_{\mathbf{A}}(\mathbf{A}) = \mathbf{U} p_{\mathbf{T}}(\mathbf{T}) \mathbf{U}^* = \mathbf{U} \mathbf{O} \mathbf{U}^* = \mathbf{O}.$$

Chapter 9 reached the same conclusion from the flag of invariant subspaces (Exercise (C1: A second proof of Cayley–Hamilton)); here the whole argument happens inside one matrix product.

Exercise. C2: The Frobenius norm is unitarily invariant

Chapter 11 §07 introduced the Frobenius norm $\|\mathbf{A}\|_F = (\sum_{i,j} |a_{ij}|^2)^{1/2}$ and proved (a) and

(b) below as Lemma 11.7.5. Prove them again here without looking, since the argument is two lines and (c) is what this section is for.

- (a) Prove that $\|\mathbf{A}\|_F^2 = \text{tr}(\mathbf{A}^* \mathbf{A})$.
- (b) Prove that $\|\mathbf{U}\mathbf{A}\|_F = \|\mathbf{A}\|_F = \|\mathbf{A}\mathbf{V}\|_F$ whenever $\mathbf{U} \in M_m(\mathbb{C})$ and $\mathbf{V} \in M_n(\mathbb{C})$ are unitary.
- (c) Deduce that if $\mathbf{A} = \mathbf{U}\mathbf{T}\mathbf{U}^*$ is a Schur factorization (Corollary 12.1.3), then $\|\mathbf{A}\|_F = \|\mathbf{T}\|_F$.

Solution.

- (a) The (j, j) entry of $\mathbf{A}^* \mathbf{A}$ is $\sum_i \overline{a_{ij}} a_{ij} = \sum_i |a_{ij}|^2$. Summing over j gives $\text{tr}(\mathbf{A}^* \mathbf{A}) = \sum_{i,j} |a_{ij}|^2 = \|\mathbf{A}\|_F^2$.
- (b) By (a) and $\mathbf{U}^* \mathbf{U} = \mathbf{I}_m$,

$$\|\mathbf{U}\mathbf{A}\|_F^2 = \text{tr}((\mathbf{U}\mathbf{A})^*(\mathbf{U}\mathbf{A})) = \text{tr}(\mathbf{A}^* \mathbf{U}^* \mathbf{U} \mathbf{A}) = \text{tr}(\mathbf{A}^* \mathbf{A}) = \|\mathbf{A}\|_F^2.$$

For the other side, use $\text{tr}(\mathbf{B}\mathbf{C}) = \text{tr}(\mathbf{C}\mathbf{B})$ (Theorem 0.9.6 (3)) and $\mathbf{V}\mathbf{V}^* = \mathbf{I}_n$:

$$\|\mathbf{A}\mathbf{V}\|_F^2 = \text{tr}(\mathbf{V}^* \mathbf{A}^* \mathbf{A} \mathbf{V}) = \text{tr}(\mathbf{A}^* \mathbf{A} \mathbf{V} \mathbf{V}^*) = \text{tr}(\mathbf{A}^* \mathbf{A}) = \|\mathbf{A}\|_F^2.$$

Norms are nonnegative, so the squares being equal gives the norms equal.

- (c) Apply (b) twice: $\|\mathbf{A}\|_F = \|\mathbf{U}\mathbf{T}\mathbf{U}^*\|_F = \|\mathbf{T}\mathbf{U}^*\|_F = \|\mathbf{T}\|_F$, using that $\mathbf{U}^*$ is unitary as well. So a Schur factorization redistributes the size of $\mathbf{A}$ between the diagonal and the strictly upper triangle without changing the total. Section 11 turns that observation into a measurement.

Exercise. C3: Distinct eigenvalues are dense, but not over the reals

- (a) Sharpen Corollary 12.1.6: prove that for every $\mathbf{A} \in M_n(\mathbb{C})$ and every $\varepsilon > 0$ there is $\mathbf{A}'$ with n **distinct** eigenvalues and $|a_{ij} - a'_{ij}| < \varepsilon$ for all i, j .
- (b) Show that the analogous statement fails over $\mathbb{R}$: for $\mathbf{R} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ there is an $\varepsilon > 0$ such that **no** real matrix within ε of $\mathbf{R}$, entry by entry, has a real eigenvalue at all.

Hint for (b): a real 2×2 matrix with entries a, b, c, d has real eigenvalues exactly when $(a - d)^2 + 4bc \geq 0$.

Solution.

- (a) This is exactly what the proof of Corollary 12.1.6 produced, before the conclusion was weakened. There the δ_k were chosen so that the diagonal entries of $\mathbf{T} + \mathbf{D}$ are pairwise distinct, and by Theorem 9.7.4 (b) those n distinct numbers are the eigenvalues of $\mathbf{T} + \mathbf{D}$, hence of the similar matrix $\mathbf{A}' = \mathbf{U}(\mathbf{T} + \mathbf{D})\mathbf{U}^*$. The entrywise estimate is unchanged.

- (b) The characteristic polynomial of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $x^2 - (a+d)x + (ad-bc)$, whose discriminant

is

$$(a + d)^2 - 4(ad - bc) = (a - d)^2 + 4bc,$$

so the eigenvalues are real exactly when this is ≥ 0 . Take $\varepsilon = \frac{1}{4}$ and let $\mathbf{A}'$ have entries within ε of those of $\mathbf{R}$, so $|a|, |d| \leq \frac{1}{4}$, $b \in [-\frac{5}{4}, -\frac{3}{4}]$, and $c \in [\frac{3}{4}, \frac{5}{4}]$. Then $(a - d)^2 \leq \frac{1}{4}$ and $bc \leq -\frac{9}{16}$, so

$$(a - d)^2 + 4bc \leq \frac{1}{4} - \frac{9}{4} = -2 < 0.$$

Hence $\mathbf{A}'$ has no real eigenvalue, let alone two distinct ones. The reason is the one this section keeps meeting: over $\mathbb{C}$ the eigenvalues move continuously and can always be separated, while over $\mathbb{R}$ a conjugate pair cannot be pushed onto the real line by a small change.