Finite Dimensional Spectral Theory

We begin with a brief review; see Chapter 1 of H-J for more details. Let V be finite dimensional and let $L \in \mathcal{L}(V)$. Unless stated otherwise, $\mathbb{F} = \mathbb{C}$.

Definition. $\lambda \in \mathbb{C}$ is an *eigenvalue* of L if $\exists\, v \in V$, $v \neq 0$, $\ni$ $Lv = \lambda v$. The vector v is called an eigenvector associated with the eigenvalue λ.

Thus, if (λ, v) is a eigenvalue-eigenvector pair for L, then span$\{v\}$ is a one-dimensional invariant subspace under L, and L acts on span$\{v\}$ by scalar multiplication by λ. Denote by $E_\lambda = \mathcal{N}(\lambda I - L)$ the λ-eigenspace of L. Every nonzero vector in E_λ is an *eigenvector* of L associated with the eigenvalue λ. Define the *geometric multiplicity* of λ to be $\dim E_\lambda$, i.e., the maximum number of linearly independent eigenvectors associated with λ. The *spectrum* $\sigma(L)$ of L is the set of its eigenvalues, and the *spectral radius* of L is

$$\rho(L) = \max\{|\lambda| : \lambda \in \sigma(L)\}.$$

Clearly $\lambda \in \sigma(L) \Leftrightarrow \lambda I - L$ is singular $\Leftrightarrow \det(\lambda I - L) = 0 \Leftrightarrow p_L(\lambda) = 0$, where $p_L(t) = \det(tI - L)$ is the *characteristic polynomial* of L; p_L is a monic polynomial of degree $n = \dim V$ whose roots are exactly the eigenvalues of L. By the fundamental theorem of algebra, p_L has n roots counting multiplicity; we define the *algebraic multiplicity* of an eigenvalue λ of L to be its multiplicity as a root of p_L.

Facts.

(1) The algebraic multiplicity of any eigenvalue is greater than or equal to its geometric multiplicity.

(2) Eigenvectors corresponding to different eigenvalues are linearly independent; i.e., if $v_i \in E_{\lambda_i}\backslash\{0\}$ for $1 \leq i \leq k$ and $\lambda_i \neq \lambda_j$ for $i \neq j$, then $\{v_1, \ldots, v_k\}$ is linearly independent. Moreover, if $\{v_1, \ldots, v_k\}$ is a set of eigenvectors with the property that for each $\lambda \in \sigma(L)$, the subset of $\{v_1, \ldots, v_k\}$ corresponding to λ (if nonempty) is linearly independent, then $\{v_1, \ldots, v_k\}$ is linearly independent.

Definition. $L \in \mathcal{L}(V)$ is *diagonalizable* if there is a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V consisting of eigenvectors of L. This definition is clearly equivalent to the alternate definition: L is diagonalizable if there is a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V for which the matrix of L with respect to $\mathcal{B}$ is diagonal ($\in \mathbb{C}^{n\times n}$)

Let $\lambda \in \sigma(L)$ and set

$$\begin{aligned} \mathsf{m}_G(\lambda) &= \text{(the geometric multiplicity of } \lambda), \\ \mathsf{m}_A(\lambda) &= \text{(the algebraic multiplicity of } \lambda). \end{aligned}$$

By definition $\mathsf{m}_G(\lambda) \leq \mathsf{m}_A(\lambda)$ for each $\lambda \in \sigma(L)$ and $\sum_{\lambda\in\sigma(L)} \mathsf{m}_A(\lambda) = n = \dim V$. Therefore, $\sum_{\lambda\in\sigma(L)} \mathsf{m}_G(\lambda) \leq n$ with equality iff $\mathsf{m}_G(\lambda) = \mathsf{m}_A(\lambda)$ for all $\lambda \in \sigma(L)$. By Fact 2, $\sum_{\lambda\in\sigma(L)} \mathsf{m}_G(\lambda)$ is the maximum number of linearly independent eigenvectors of L. Thus L is diagonalizable $\Leftrightarrow$ $(\forall\, \lambda \in \sigma(L))$ $\mathsf{m}_G(\lambda) = \mathsf{m}_A(\lambda)$. In particular, since $(\forall\, \lambda \in \sigma(L))$ $\mathsf{m}_G(\lambda) \geq 1$, if L has n distinct eigenvalues, then L is diagonalizable.

We say that a matrix $A \in \mathbb{C}^{n\times n}$ is diagonalizable iff A is similar to a diagonal matrix, i.e., there exists an invertible $S \in C^{n\times n}$ for which $S^{-1}AS = D$ is diagonal. Consider the linear transformation $L : \mathbb{C}^n \to \mathbb{C}^n$ given by $L : x \mapsto Ax$; since matrices similar to A correspond to the matrices of L with respect to different bases, clearly the matrix A is diagonalizable iff L is diagonalizable. Since S is the change of basis matrix and $e_1, \ldots, e_n$ are eigenvectors of D, it follows that the columns of S are linearly independent eigenvectors of A. This is also clear by computing the matrix equality $AS = SD$ column by column.

We will restrict our attention to $\mathbb{C}^n$ with the Euclidean inner product $\langle \cdot, \cdot \rangle$; here $\| \cdot \|$ will denote the norm induced by $\langle \cdot, \cdot \rangle$ (i.e., the ℓ^2-norm on $\mathbb{C}^n$), and we will denote by $\|A\|$ the operator norm induced on $\mathbb{C}^{n\times n}$ (previously denoted $|||A|||_2$). Virtually all the classes of matrices we are about to define generalize to any Hilbert space V, but we must first know that for $y \in V$ and $A \in \mathcal{B}(V)$, $\exists\, A^*y \in V \ni \langle Ax, y\rangle = \langle x, A^*y\rangle$; we will prove this next quarter. So far, we know that we can define the transpose operator $A^* \in \mathcal{B}(V^*)$, so we need to know that we can identify $V^* \cong V$ as we can do in finite dimensions. For now we restrict to $\mathbb{C}^n$.

One can think of many of our operations and sets of matrices in $\mathbb{C}^{n\times n}$ as analogous to corresponding objects in $\mathbb{C}$. For example, the operation $A \mapsto A^H$ is thought of as analogous to conjugation $z \mapsto \bar{z}$ in $\mathbb{C}$. The analogue of a real number is a Hermitian matrix.

Definition. $A \in \mathbb{C}^{n\times n}$ is said to be Hermitian symmetric (or self-adjoint or just Hermitian) if $A = A^H$. $A \in \mathbb{C}^{n\times n}$ is said to be *skew-Hermitian* if $A^H = -A$.

Recall that we have already given a definition of what it means for a sesquilinear form to be Hermitian symmetric. Recall also that there is a $1-1$ correspondence between sesquilinear forms and matrices $A \in \mathbb{C}^{n\times n}$: A corresponds to the form $\langle x, y\rangle_A = \langle Ax, y\rangle$. It is easy to check that A is Hermitian iff the sesquilinear form $\langle \cdot, \cdot\rangle_A$ is Hermitian-symmetric.

Fact: A is Hermitian iff iA is skew-Hermitian (exercise).

The analogue of the imaginary numbers in $\mathbb{C}$ are the skew-Hermitian matrices. Also, any $A \in \mathbb{C}^{n\times n}$ can be written uniquely as $A = B + iC$ where B and C are Hermitian: $B = \frac{1}{2}(A + A^H)$, $C = \frac{1}{2i}(A - A^H)$. Then $A^H = B - iC$. Almost analogous to the $\mathcal{R}e$ and $\mathcal{I}m$ part of a complex number, B is called the *Hermitian part* of A, and iC (not C) is called the *skew-Hermitian part* of A.

Proposition. $A \in \mathbb{C}^{n\times n}$ is Hermitian iff $(\forall\, x \in \mathbb{C}^n)\langle Ax, x\rangle \in \mathbb{R}$.

Proof. If A is Hermitian, $\langle Ax, x\rangle = \frac{1}{2}(\langle Ax, x\rangle + \langle x, Ax\rangle) = \mathcal{R}e\langle Ax, x\rangle \in \mathbb{R}$. Conversely, suppose $(\forall\, x \in \mathbb{C}^n)\langle Ax, x\rangle \in \mathbb{R}$. Write $A = B + iC$ where B, C are Hermitian. Then $\langle Bx, x\rangle \in \mathbb{R}$ and $\langle Cx, x\rangle \in \mathbb{R}$, so $\langle Ax, x\rangle \in \mathbb{R} \Rightarrow \langle Cx, x\rangle = 0$. Since any sesquilinear form $\{x, y\}$ over $\mathbb{C}$ can be recovered from the associated quadratic form $\{x, x\}$ by polarization: $\{x, y\} = \frac{1}{4}[\{x+y, x+y\} - \{x-y, x-y\} + i\{x+iy, x+iy\} - i\{x-iy, x-iy\}]$, we conclude that $\langle Cx, y\rangle = 0\ \forall\, x, y \in \mathbb{C}^n$, and thus $C = 0$, so $A = B$ is Hermitian. □

The analogue of the nonnegative reals are the positive semi-definite matrices.

Definition. $A \in \mathbb{C}^{n\times n}$ is called *positive semi-definite* (or nonnegative) if $(\forall\, x \in \mathbb{C}^n)\langle Ax, x\rangle \geq 0$. By the previous proposition, a positive semi-definite $A \in \mathbb{C}^{n\times n}$ is automatically Hermitian, but one often says Hermitian positive semi-definite anyway.

Caution: If $A \in \mathbb{R}^{n\times n}$ and $(\forall\, x \in \mathbb{R}^n)$ $\langle Ax, x\rangle \geq 0$, A need not be symmetric. For example, if $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, then $\langle Ax, x\rangle = \langle x, x\rangle$ $\forall\, x \in \mathbb{R}^n$.

If $A \in \mathbb{C}^{n\times n}$ then we can think of $A^H A$ as the analogue of $|z|^2$ for $z \in \mathbb{C}$. Observe that $A^H A$ is positive semi-definite: $\langle A^H Ax, x\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0$. In fact, $\|A^H A\| = \|A\|^2$ since

$$\|A\| = \|A^H\|,\ \|A^H A\| \leq \|A^H\| \cdot \|A\| = \|A\|^2$$

and

$$\begin{aligned} \|A^H A\| &= \sup_{\|x\|=1} \|A^H Ax\| \\ &= \sup_{\|x\|=1} \sup_{\|y\|=1} |\langle A^H Ax, y\rangle| \\ &\geq \sup_{\|x\|=1} \langle A^H Ax, x\rangle \\ &= \sup_{\|x\|=1} \|Ax\|^2 \ = \ \|A\|^2 \,. \end{aligned}$$

The analogue of complex numbers of modulus 1 are the unitary matrices.

Definition. $A \in \mathbb{C}^{n\times n}$ is *unitary* if $A^H A = I$. Since injectivity is equivalent to surjectivity for $A \in \mathbb{C}^{n\times n}$, it follows that $A^H = A^{-1}$ and $AA^H = I$ (each of these is equivalent to $A^H A = I$).

Proposition. For $A \in \mathbb{C}^{n\times n}$, the following conditions are all equivalent:

(1) A is unitary.

(2) The columns of A form an orthonormal basis of $\mathbb{C}^n$.

(3) The rows of A form an orthonormal basis of $\mathbb{C}^n$.

(4) A preserves the Euclidean norm: $(\forall\, x \in \mathbb{C}^n)$ $\|Ax\| = \|x\|$.

(5) A preserves the Euclidean inner product: $(\forall\, x, y \in \mathbb{C}^n)$ $\langle Ax, Ay\rangle = \langle x, y\rangle$.

Proof Sketch. Let $a_1, \ldots, a_n$ be the columns of A. Clearly $A^H A = I \Leftrightarrow a_i^H a_j = \delta_{ij}$. So (1) $\Leftrightarrow$ (2). Similarly (1) $\Leftrightarrow AA^H = I \Leftrightarrow$ (3). Since $\|Ax\|^2 = \langle Ax, Ax\rangle = \langle A^H Ax, x\rangle$ and $A^H A$ is Hermitian, (4) $\Leftrightarrow \langle (A^H A - I)x, x\rangle = 0\, \forall\, x \in \mathbb{C}^n \Leftrightarrow A^H A = I \Leftrightarrow$ (1). Finally, clearly (5) $\Rightarrow$ (4), and (4) $\Rightarrow$ (5) by polarization. □

Normal matrices don't really have an analogue in $\mathbb{C}$.

Definition. $A \in \mathbb{C}^{n\times n}$ is *normal* if $AA^H = A^H A$.

Proposition. For $A \in \mathbb{C}^{n\times n}$, the following conditions are equivalent:

(1) A is normal.

(2) The Hermitian and skew-Hermitian parts of A commute, i.e., if $A = B + iC$ with B, C Hermitian, $BC = CB$.

(3) $(\forall\, x \in \mathbb{C}^n)\ \|Ax\| = \|A^H x\|$.

Proof Sketch. Clearly (1) $\Leftrightarrow$ (2) (exercise). Since $\|Ax\|^2 = \langle A^H Ax, x\rangle$ and $\|A^H x\|^2 = \langle AA^H x, x\rangle$, and since $A^H A$ and AA^H are Hermitian,

$$(3) \quad \Leftrightarrow \quad (\forall\, x \in \mathbb{C}^n)\ \langle (A^H A - AA^H)x, x\rangle = 0 \quad \Leftrightarrow \quad (1).$$

□

Observe that Hermitian, skew-Hermitian, and unitary matrices are all normal.

The above definitions can all be specialized to the real case. Real Hermitian matrices are (real) *symmetric* matrices: $A^T = A$. Every $A \in \mathbb{R}^{n\times n}$ can be written uniquely as $A = B + C$ where $B = B^T$ is symmetric and $C = -C^T$ is skew-symmetric: $B = \frac{1}{2}(A + A^T)$ is called the *symmetric part* of A; $C = \frac{1}{2}(A - A^T)$ is the *skew-symmetric part*. Real unitary matrices are called *orthogonal matrices*, characterized by $A^T A = I$ or $A^T = A^{-1}$. Since $(\forall\, A \in \mathbb{R}^{n\times n})(\forall\, x \in \mathbb{R}^n)\langle Ax, x\rangle \in \mathbb{R}$, there is no characterization of symmetric matrices analogous to that given above for Hermitian matrices. Also unlike the complex case, the values of the quadratic form $\langle Ax, x\rangle$ for $x \in \mathbb{R}^n$ only determine the symmetric part of A, not A itself (the real polarization identity $\{x, y\} = \frac{1}{4}(\{x + y, x + y\} - \{x - y, x - y\})$ is valid only for *symmetric* bilinear forms $\{x, y\}$ over $\mathbb{R}^n$). Consequently, the definition of real positive semi-definite matrices includes symmetry in the definition, together with $(\forall\, x \in \mathbb{R}^n)\ \langle Ax, x\rangle \geq 0$. (This is standard, but not universal. In some mathematical settings, symmetry is not assumed automatically. This is particularly the case in monotone operator theory and optimization theory where it is essential to the theory and the applications that positive definite matrices and operators are **not** assumed to be symmetric.)

The analogy with the complex numbers is particularly clear when considering the eigenvalues of matrices in various classes. For example, consider the characteristic polynomial of a matrix $A \in \mathbb{C}^{n\times n}$, $P_A(t)$. Since $\overline{P_A(t)} = P_{A^H}(\bar{t})$, we have $\lambda \in \sigma(A) \Leftrightarrow \bar{\lambda} \in \sigma(A^H)$. If A is Hermitian, then all eigenvalues of A are real: if x is an eigenvector associated with λ, then $\lambda\langle x, x\rangle = \langle Ax, x\rangle = \langle x, Ax\rangle = \bar{\lambda}\langle x, x\rangle$, so $\lambda = \bar{\lambda}$. Also eigenvectors corresponding to different eigenvalues are orthogonal: if $Ax = \lambda x$ and $Ay = \mu y$, then $\lambda\langle x, y\rangle = \langle Ax, y\rangle = \langle x, Ay\rangle = \mu\langle x, y\rangle$, so $\langle x, y\rangle = 0$ if $\lambda \neq \mu$. Any eigenvalue λ of a unitary matrix satisfies $|\lambda| = 1$ since $|\lambda| \cdot \|x\| = \|Ax\| = \|x\|$. Again, eigenvectors corresponding to different eigenvalues of a unitary matrix are orthogonal: if $Ax = \lambda x$ and $Ay = \mu y$, then $\lambda\langle x, y\rangle = \langle Ax, y\rangle = \langle x, A^{-1}y\rangle = \langle x, \mu^{-1}y\rangle = \bar{\mu}^{-1}\langle x, y\rangle = \mu\langle x, y\rangle$.

Matrices which are both Hermitian and unitary, i.e., $A = A^H = A^{-1}$, satisfy $A^2 = I$. The linear transformations determined by such matrices can be thought of as generalizations of reflections: One example is $A = -I$, corresponding to reflections about the origin. Householder transformations are of the form $I - \frac{1}{\langle y,y\rangle} yy^H$ where $y \in \mathbb{C}^n \backslash \{0\}$: they correspond to reflection about the hyperplane orthogonal to y, as $x \mapsto x - 2\frac{\langle x,y\rangle}{\langle y,y\rangle}y$; $\frac{\langle x,y\rangle}{\langle y,y\rangle}y$ is the orthogonal projection onto span$\{y\}$; $x - \frac{\langle x,y\rangle}{\langle y,y\rangle}y$ is the projection onto $\{y\}^\perp$.

Unitary Equivalence

Similar matrices represent the same linear transformation. There is a special case of similarity which is of particular importance.

Definition. We say that $A, B \in \mathbb{C}^{n\times n}$ are *unitarily equivalent* (or unitarily similar) if there is a unitary matrix $U \in \mathbb{C}^{n\times n} \ni B = U^H AU$, i.e., A and B are similar via a unitary similarity transformation (recall: $U^H = U^{-1}$).

Unitary equivalence is important for several reasons. One is that the Hermitian transpose of a matrix is much easier to compute than an inverse, so unitary similarity is computationally advantageous. Another is that, with respect to the operator norm $\|\cdot\|$ on $\mathbb{C}^{n\times n}$ induced by the Euclidean norm on $\mathbb{C}^n$, a unitary matrix U is perfectly conditioned: $(\forall\, x \in \mathbb{C}^n)$ $\|Ux\| = \|x\| = \|U^H x\|$ implies $\|U\| = \|U^H\| = 1$, so $\kappa(U) = \|U\| \cdot \|U^{-1}\| = \|U\| \cdot \|U^H\| = 1$; moreover, unitary similarity preserves the condition number of a matrix relative to $\|\cdot\|$: $\kappa(U^H AU) = \|U^H AU\| \cdot \|U^H A^{-1} U\| \le \kappa(A)$ and likewise $\kappa(A) \le \kappa(U^H AU)$. (In general, for any submultiplicative norm on $\mathbb{C}^{n\times n}$, we obtain the often crude estimate $\kappa(S^{-1}AS) = \|S^{-1}AS\| \cdot \|S^{-1}A^{-1}S\| \le \|S^{-1}\|^2 \|A\| \cdot \|A^{-1}\| \cdot \|S\|^2 = \kappa(S)^2 \kappa(A)$, indicating that similarity transformations can drastically change the condition number of A if the transition matrix S is poorly conditioned; note also that $\kappa(A) \le \kappa(S)^2 \kappa(S^{-1}AS)$.) Another basic reason is that unitary similarity preserves the Euclidean operator norm $\|\cdot\|$ and the Frobenius norm $\|\cdot\|_F$ of a matrix.

Proposition. Let $U \in \mathbb{C}^{n\times n}$ be unitary, and $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times k}$. Then

(1) In the operator norms induced by the Euclidean norms, $\|AU\| = \|A\|$ and $\|UB\| = \|B\|$.

(2) In the Frobenius norms, $\|AU\|_F = \|A\|_F$ and $\|UB\|_F = \|B\|_F$.

So multiplication by a unitary matrix on either side preserves $\|\cdot\|$ and $\|\cdot\|_F$.

Proof Sketch. (1) $(\forall\, x \in \mathbb{C}^k)$ $\|UBx\| = \|Bx\|$, so $\|UB\| = \|B\|$. Likewise, since U^H is also unitary, $\|AU\| = \|(AU)^H\| = \|U^H A^H\| = \|A^H\| = \|A\|$. (2) Let $b_1, \ldots, b_k$ be the columns of B. Then $\|UB\|_F^2 = \sum_{j=1}^k \|Ub_j\|_2^2 = \sum_{j=1}^k \|b_j\|_2^2 = \|B\|_F^2$. Likewise, since U^H is also unitary, $\|AU\|_F = \|U^H A^H\|_F = \|A^H\|_F = \|A\|_F$. $\square$

Observe that $\|U\|_F = \sqrt{n}$.

Schur Unitary Triangularization Theorem

Any matrix $A \in \mathbb{C}^{n\times n}$ is unitarily equivalent to an upper triangular matrix T. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A in any prescribed order, then one can choose a unitary similarity transformation so that the diagonal entries of T are $\lambda_1, \ldots, \lambda_n$ in that order.

Proof Sketch (see also pp. 79–80 of H-J). By induction on n. Obvious for $n = 1$. Assume true for $n - 1$. Given $A \in \mathbb{C}^{n\times n}$ and an ordering $\lambda_1, \ldots, \lambda_n$ of its eigenvalues, choose an eigenvector x for λ_1 with Euclidean norm $\|x\| = 1$. Extend $\{x\}$ to a basis of $\mathbb{C}^n$ and apply the Gram-Schmidt procedure to obtain an orthonormal basis $\{x, u_2, \ldots, u_n\}$ of $\mathbb{C}^n$. Let

$U_1 = [x, u_2, \cdots, u_n] \in \mathbb{C}^{n\times n}$ be the unitary matrix whose columns are $x, u_2, \ldots, u_n$. Since $Ax = \lambda_1 x$, $U_1^H A U_1 = \begin{bmatrix} \lambda_1 & y_1^H \\ 0 & B \end{bmatrix}$ for some $y_1 \in \mathbb{C}^{n-1}$, $B \in \mathbb{C}^{(n-1)\times(n-1)}$. Since similar matrices have the same characteristic polynomial,

$$\begin{aligned} p_A(t) &= \det\left(tI - \begin{bmatrix} \lambda_1 & y_1^H \\ 0 & B \end{bmatrix}\right) \\ &= (t-\lambda_1)\det(tI - B) \\ &= (t-\lambda_1)p_B(t), \end{aligned}$$

so the eigenvalues of B are $\lambda_2, \ldots, \lambda_n$. By the induction hypothesis, $\exists$ a unitary $\widetilde{U} \in \mathbb{C}^{(n-1)\times(n-1)}$ and upper triangular $\widetilde{T} \in \mathbb{C}^{(n-1)\times(n-1)} \ni \widetilde{U}^H B \widetilde{U} = \widetilde{T}$ and the diagonal entries on $\widetilde{T}$ are $\lambda_2, \ldots, \lambda_n$ in that order. Let $U_2 = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{U} \end{bmatrix} \in \mathbb{C}^{n\times n}$. Then U_2 is unitary, and

$$U_2^H U_1^H A U_1 U_2 = \begin{bmatrix} \lambda_1 & y_1^H \widetilde{U} \\ 0 & \widetilde{U}^H B \widetilde{U} \end{bmatrix} = \begin{bmatrix} \lambda_1 & y_1^H \widetilde{U} \\ 0 & \widetilde{T} \end{bmatrix} \equiv T.$$

Since $U \equiv U_1 U_2$ is unitary and $U^H A U = T$, the statement is true for n as well. □

Note: The basic iterative step that reduces the dimension by 1 is called a deflation. The deflation trick is used to derive a number of important matrix factorizations.

Fact. Unitary equivalence preserves the classes of Hermitian, skew-Hermitian, and normal matrices: e.g., if $A^H = A$, then $(U^H A U)^H = U^H A^H U = U^H A U$ is also Hermitian; if $A^H A = A A^H$, then $(U^H A U)^H (U^H A U) = U^H A^H A U = U^H A A^H U = (U^H A U)(U^H A U)^H$ is normal.

Spectral Theorem. *Let $A \in \mathbb{C}^{n\times n}$ be normal. Then A is unitarily diagonalizable, i.e., A is unitarily similar to a diagonal matrix; so there is an orthonormal basis of eigenvectors.*

Proof Sketch. By Schur Triangularization Theorem, $\exists$ unitary $U \ni U^H A U = T$ is upper triangular. So A is normal $\Rightarrow$ T is normal: $T^H T = T T^H$. By equating the diagonal entries of $T^H T$ and $T T^H$, we show T is diagonal. The $(1,1)$ entry of $T^H T$ is $|t_{11}|^2$; that of $T T^H$ is $\sum_{j=1}^n |t_{1j}|^2$. Since $|t_{11}|^2 = \sum_{j=1}^n |t_{1j}|^2$, it must be the case that so $t_{1j} = 0$ for $j \geq 2$. Now the $(2,2)$ entry of $T^H T$ is $|t_{22}|^2$; that of $T T^H$ is $\sum_{j=2}^n |t_{2j}|^2$; so again it must be the case that $t_{2j} = 0$ for $j \geq 3$. Continuing with the remaining rows yields the result. □

Cayley-Hamilton Theorem

The Schur Triangularization Theorem gives a quick proof of:

Theorem. (Cayley-Hamilton) *Every matrix $A \in \mathbb{C}^{n\times n}$ satisfies its characteristic polynomial: $p_A(A) = 0$.*

Proof. By Schur, $\exists$ unitary $U \in \mathbb{C}^{n\times n}$ and upper triangular $T \in \mathbb{C}^{n\times n} \ni U^H A U = T$, where the diagonal entries of T are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A (in some order). Since $A = U T U^H$, $A^k = U T^k U^H$, so $p_A(A) = U p_A(T) U^H$. Writing $p_A(t)$ as

$$p_A(t) = (t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n)$$

gives

$$p_A(T) = (T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I).$$

Since $T - \lambda_j I$ is upper triangular with its jj entry being zero, it follows easily that $p_A(T) = 0$ (accumulate the product from the left, in which case one shows by induction on k that the first k columns of $(T - \lambda_1 I) \cdots (T - \lambda_k I)$ are zero). □

Rayleigh Quotients and the Courant-Fischer Minimax Theorem

For $A \in \mathbb{C}^{n \times n}$ and $x \in \mathbb{C}^n \backslash \{0\}$, define the *Rayleigh quotient of* x (for A) to be

$$r_A(x) = \frac{x^H A x}{x^H x}.$$

(This is not a standard notation; often $\rho(x)$ is used, but we avoid this notation to prevent possible confusion with the spectral radius $\rho(A)$.) Rayleigh quotients are most useful for Hermitian matrices $A \in \mathbb{C}^{n \times n}$.

Proposition. Let $A \in \mathbb{C}^{n \times n}$ be Hermitian with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$.

(1) For $x \in \mathbb{C}^n$ with Euclidean norm $\|x\| = 1$, $r_A(x) = x^H A x$.

(2) For $x \in \mathbb{C}^n \backslash \{0\}$ and $\alpha \in \mathbb{C} \backslash \{0\}$, $r_A(\alpha x) = r_A(x)$. In particular, $r_A(x) = r_A\left(\frac{x}{\|x\|}\right)$, so $\{r_A(x) : x \in \mathbb{C}^n \backslash \{0\}\} = \{r_A(x) : \|x\| = 1\}$.

(3) Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A corresponding to $\lambda_1, \ldots, \lambda_n$; let $U = [u_1 \cdots u_n] \in \mathbb{C}^{n \times n}$, so U is unitary and

$$U^H A U = \Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n).$$

Given $x \in \mathbb{C}^n \backslash \{0\}$, let $y = U^H x$, so $x = Uy = y_1 u_1 + \cdots + y_n u_n$. Then

$$r_A(x) = r_\Lambda(y) = \frac{\sum_{i=1}^n \lambda_i |y_i|^2}{\sum_{i=1}^n |y_i|^2}.$$

(4) For $x \neq 0$, $\lambda_1 \leq r_A(x) \leq \lambda_n$. Moreover, $\min_{x \neq 0} r_A(x) = \lambda_1$ and $\max_{x \neq 0} r_A(x) = \lambda_n$.

(5) In the context of (3), given $x \in \mathbb{C}^n \backslash \{0\}$ and setting $y = U^H x$,

$$\frac{\|Ax\|^2}{\|x\|^2} = \frac{\sum_{i=1}^n \lambda_i^2 |y_i|^2}{\sum_{i=1}^n |y_i|^2}.$$

(6) The Euclidean operator norm of A satisfies $\|A\| = \rho(A)$, the spectral radius of A.

Caution. (6) is not necessarily true $\forall\, A \in \mathbb{C}^{n \times n}$.

Exercise. Show that $\|A\| = \rho(A)$ where the norm is the Euclidean operator norm if A is normal.

Proof Sketch. (1) and (2) follow immediately from the definition. For (3),

$$r_A(x) = \frac{x^H A x}{x^H x} = (y^H U^H A U y)/(y^H U^H U y) = \frac{y^H \Lambda y}{y^H y} = r_\Lambda(y).$$

For (4), clearly $\lambda_1 \leq r_\Lambda(y) \leq \lambda_n$; if $x = u_j$, then $r_A(x) = \lambda_j$, so λ_1 and λ_n are taken on. For (5), since $A^H = A$,

$$\|Ax\|^2/\|x\|^2 = \frac{x^H A^2 x}{x^H x} = r_{A^2}(x) = r_{\Lambda^2}(y) = \frac{\sum_{i=1}^n \lambda_i^2 |y_i|^2}{\sum_{i=1}^n |y_i|^2}.$$

For (6), clearly $r_{\Lambda^2}(y) \leq \rho(A)^2$, so $\|A\| \leq \rho(A)$. Since $\frac{\|Au_j\|}{\|u_j\|} = |\lambda_j|$, $\|A\| \geq \rho(A)$. Thus $\|A\| = \rho(A)$. (Note here that $\rho(A) = \max\{|\lambda_1|, |\lambda_n|\}$.) □

Corollary. If $A \in \mathbb{C}^{m\times n}$, then $\|A\| = \sqrt{\rho(A^H A)}$ (in the operator norm induced by Euclidean norms).

Proof. For $x \neq 0$, $\|Ax\|^2/\|x\|^2 = (x^H A^H A x)/x^H x = r_{A^H A}(x)$. Since $A^H A$ is positive semidefinite, its eigenvalues are nonnegative, so $\max_{x\neq 0} \frac{\|Ax\|^2}{\|x\|^2} = \max_{x\neq 0} r_{A^H A}(x) = \rho(A^H A)$. □

There is a very useful extension of part (4) of the Proposition above.

Courant-Fischer Minimax Theorem. *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian. In what follows, S_k will denote an arbitrary subspace of $\mathbb{C}^n$ of dimension k, and $\min_{S_k}$ and $\max_{S_k}$ denote taking the min or max over all subspaces of $\mathbb{C}^n$ of dimension k.*

(1) For $1 \leq k \leq n$, $\quad \min_{S_k} \max_{x\neq 0, x\in S_k} r_A(x) = \lambda_k$ *(minimax)*

(2) For $1 \leq k \leq n$, $\quad \max_{S_{n-k+1}} \min_{x\neq 0, x\in S_{n-k+1}} r_A(x) = \lambda_k$ *(maximin)*

where $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of A.

Proof. Let $u_1, \ldots, u_n$ be orthonormal eigenvectors of A corresponding to $\lambda_1, \ldots, \lambda_n$. Let $U = [u_1 \cdots u_n] \in \mathbb{C}^{n\times n}$, so $U^H A U = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, and for $x \in \mathbb{C}^n$, let $y = U^H x$, so $x = Uy = y_1 u_1 + \cdots + y_n u_n$. To prove (1), let $W = \operatorname{span}\{u_k, \ldots, u_n\}$, so $\dim W = n - k + 1$. If $\dim S_k = k$, then by dimension arguments $\exists\, x \in S_k \cap W\backslash\{0\}$, so $r_A(x) = \sum_{i=k}^n \lambda_i |y_i|^2 / \sum_{i=k}^n |y_i|^2 \geq \lambda_k$. Thus $\forall\, S_k$, $\max_{x\neq 0, x\in S_k} r_A(x) \geq \lambda_k$. But for $x \in \operatorname{span}\{u_1, \ldots, u_k\}\backslash\{0\}$, $r_A(x) = \sum_{i=1}^k \lambda_i |y_i|^2 / \sum_{i=1}^k |y_i|^2 \leq \lambda_k$. Thus $\min\max = \lambda_k$. The proof of (2) is similar. Let $W = \operatorname{span}\{u_1, \ldots, u_k\}$, so $\dim W = k$. If $\dim S_{n-k+1} = n - k + 1$, then $\exists\, x \in S_{n-k+1} \cap W\backslash\{0\}$, and $r_A(x) \leq \lambda_k$. Thus $\forall\, S_{n-k+1}\ \min_{x\neq 0, x\in S_{n-k+1}} r_A(x) \leq \lambda_k$. But for $S_{n-k+1} = \operatorname{span}\{u_k, \ldots, u_n\}$, the min is λ_k, so $\max\min = \lambda_k$. □

Remark. (1) for $k = 1$ and (2) for $k = n$ give part (4) of the previous Proposition.

Non-Unitary Similarity Transformations

Despite the advantages of unitary equivalence, there are limitations. Not every diagonalizable matrix is unitarily diagonalizable. For example, consider an upper-triangular matrix T with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. We know that T is diagonalizable. However, T is not unitarily similar to a diagonal matrix unless it is already diagonal. This is because unitary equivalence preserves the Frobenius norm: $\|T\|_F^2 = \sum_{i=1}^n |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2$, but any diagonal matrix similar to T has Frobenius norm $\sum_{i=1}^n |\lambda_i|^2$. In order to diagonalize T it is necessary to use non-unitary similar transformations.

Proposition. Let $A \in \mathbb{C}^{n\times n}$, and let $\lambda_1, \ldots, \lambda_k$ be the *distinct* eigenvalues of A, with multiplicities $m_1, \ldots, m_k$, respectively (so $m_1 + \cdots + m_k = n$). Then A is similar to a block diagonal matrix of the form

$$\begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ \bigcirc & & & T_k \end{bmatrix},$$

where each $T_i \in \mathbb{C}^{m_i \times m_i}$ is upper triangular with λ_i as each of its diagonal entries.

Proof. By Schur, A is similar to an upper triangular T with diagonal entries ordered $\overbrace{\lambda_1, \ldots, \lambda_1}^{m_1}, \overbrace{\lambda_2, \ldots, \lambda_2}^{m_2}, \cdots, \overbrace{\lambda_k, \ldots, \lambda_k}^{m_k}$. We use a strategy as in Gaussian Elimination (but must be sure to do similarity transformations). Consider the matrices $E_{rs} \in \mathbb{C}^{n\times n}$ having 1 is the (r, s)-entry and 0 elsewhere. Left multiplication of T by E_{rs} moves the sth row of T to the rth row and zeros out all other elements, that is, the elements of the matrix $E_{rs}T$ are all zero except for those in the rth row which is just the sth row of T. Therefore, left multiplication of T by the matrix $(I - \alpha E_{rs})$ corresponds to subtracting α times the sth row of T from the rth row of T. This is just one of the elementary row operations used in Gaussian elimination. Note in particular that if $r < s$, then this operation introduces no new non-zero entries below the main diagonal of T, that is, $E_{rs}T$ is still upper triangular (as is $(I - \alpha E_{rs})$).

Similarly, right multiplication of T by E_{rs} moves the rth column of T to the sth column and zeros out all other entries in the matrix, that is, the elements of the matrix TE_{rs} are all zero except for those in the sth column which is just the rth column of T. Therefore, right multiplication of T by $(I + \alpha E_{rs})$ corresponds to adding α times the rth column of T to the sth column of T. In particular, if $r < s$, then this operation introduces no new non-zero entries below the main diagonal of T, that is, TE_{rs} is still upper triangular.

Because of the properties described above, the matrices $(I \pm \alpha E_{rs})$ are sometimes referred to as *Gaussian elimination matrices*. Note that $E_{rs}^2 = 0$ whenever $r \neq s$, and so

$$(I - \alpha E_{rs})(I + \alpha E_{rs}) = I - \alpha E_{rs} + \alpha E_{rs} - \alpha E_{rs}^2 = I.$$

That is, $(I + \alpha E_{rs})^{-1} = (I - \alpha E_{rs})$, which makes sense since the inverse of adding α times the sth row to the rth row is to subtract it.

Now consider the similarity transformation

$$T \mapsto (I + \alpha E_{rs})^{-1} T (I + \alpha E_{rs}) = (I - \alpha E_{rs}) T (I + \alpha E_{rs})$$

with $\alpha \in \mathbb{C}$ and $r < s$. Since T is upper triangular (as are $I \pm \alpha E_{rs}$ for $r < s$), it follows that this similarity transformation only changes T in the s^{th} column above (and including) the r^{th} row, and in the r^{th} row to the right of (and including) the s^{th} column, and that t_{rs} gets replaced by $t_{rs} + \alpha(t_{rr} - t_{ss})$. So if $t_{rr} \neq t_{ss}$ it is possible to choose α to make $t_{rs} = 0$. Using these observations, it is easy to see that such transformations can be performed successively without destroying previously created zeroes to zero out all entries except those in the desired block diagonal form (work backwards row by row starting with row $n - m_k$; in each row, zero out entries from left to right). □

Jordan Form

Let $T \in \mathbb{C}^{n\times n}$ be an upper triangular matrix in block diagonal form

$$T = \begin{bmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_k \end{bmatrix}$$

as in the previous Proposition, i.e., $T_i \in \mathbb{C}^{m_i \times m_i}$ satisfies $T_i = \lambda_i I + N_i$ where $N_i \in \mathbb{C}^{m_i \times m_i}$ is strictly upper triangular, and $\lambda_1, \ldots, \lambda_k$ are distinct. Then for $1 \leq i \leq k$, $N_i^{m_i} = 0$, so N is nilpotent. Recall that any nilpotent operator is a direct sum of shift operators in an appropriate basis, so the matrix N_i is similar to a direct sum of shift matrices

$$S_\ell = \begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}$$

of varying sizes ℓ. Thus each T_i is similar to a direct sum of *Jordan blocks*

$$J_\ell(\lambda) = \lambda I_\ell + S_\ell = \begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix} \in \mathbb{C}^{\ell\times\ell}$$

of varying sizes ℓ (with $\lambda = \lambda_i$).

Definition. A matrix J is in *Jordan normal form* if it is the direct sum of finitely many Jordan blocks (with, of course, possibly different values of λ and ℓ).

The previous Proposition, together with our results on the structure of nilpotent operators as discussed above, establishes the following Theorem.

Theorem. *Every matrix $A \in \mathbb{C}^{n\times n}$ is similar to a matrix in Jordan normal form.*

Remarks.

(1) The Jordan form of A is not quite unique since the blocks may be arbitrarily reordered by a similarity transformation. As we will see, this is the only nonuniqueness: the number of blocks of each size for each eigenvalue λ is determined by A.

(2) For $\lambda \in \sigma(A)$ and $j \geq 1$, let $b_j(\lambda)$ denote the number of $j \times j$ blocks associated with λ in *some* Jordan matrix similar to A, and let $r(\lambda) = \max\{j : b_j(\lambda) > 0\}$ be the size of the largest block associated with λ. Let $k_j(\lambda) = \dim(\mathcal{N}(A - \lambda I)^j)$. Then from our remarks on nilpotent operators,

$$0 < k_1(\lambda) < k_2(\lambda) < \cdots < k_{r(\lambda)}(\lambda) = k_{r(\lambda)+1}(\lambda) = \cdots = m(\lambda),$$

where $m(\lambda)$ is the algebraic multiplicity of λ. By considering the form of powers of shift matrices, one can easily show that

$$b_j(\lambda) + b_{j+1}(\lambda) + \cdots + b_{r(\lambda)}(\lambda) = k_j(\lambda) - k_{j-1}(\lambda),$$

i.e., the number of blocks of size $\geq j$ associated with λ is $k_j(\lambda) - k_{j-1}(\lambda)$. (In particular, for $j = 1$, the number of Jordan blocks associated with λ is $k_1(\lambda) =$ the geometric multiplicity of λ.) Thus,

$$b_j(\lambda) = -k_{j+1}(\lambda) + 2k_j(\lambda) - k_{j-1}(\lambda),$$

which is completely determined by A. Since $k_j(\lambda)$ is invariant under similarity transformations, we conclude:

Proposition. (a) The Jordan form of A is unique up to reordering of the Jordan blocks. (b) Two matrices in Jordan form are similar iff they can be obtained from each other by reordering the blocks.

Remarks.

(3) Knowing the algebraic and geometric multiplicities of each eigenvalue of A is not sufficient to determine the Jordan form (unless the algebraic multiplicity of each eigenvalue is at most one greater than its geometric multiplicity. *Exercise.* Why is it determined in this case?)

For example,

$$N_1 = \begin{bmatrix} 0 & 1 & & \bigcirc \\ & 0 & 1 & \\ & & 0 & \\ \bigcirc & & & 0 \end{bmatrix} \quad \text{and} \quad N_2 = \begin{bmatrix} 0 & 1 & & \bigcirc \\ 0 & 0 & & \\ & & 0 & 1 \\ \bigcirc & & 0 & 0 \end{bmatrix}$$

are not similar as $N_1^2 \neq 0 = N_2^2$, but both have 0 as the only eigenvalue with algebraic multiplicity 4 and geometric multiplicity 2.

(4) The expression for $b_j(\lambda)$ in remark (2) above can also be given in terms of $r_j(\lambda) \equiv \operatorname{rank}((A - \lambda I)^j) = \dim(\mathcal{R}(A - \lambda I)^j) = n - k_j(\lambda)$: $b_j = r_{j+1} - 2r_j + r_{j-1}$.

(5) A necessary and sufficient condition for two matrices in $\mathbb{C}^{n\times n}$ to be similar is that they are both similar to the same Jordan normal form matrix.

Spectral Decomposition

There is a useful invariant formulation (i.e., basis-free) of some of the above. Let $L \in \mathcal{L}(V)$ where $\dim V = n < \infty$ (and $\mathbb{F} = \mathbb{C}$). Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of L, with algebraic multiplicities $m_1, \ldots, m_k$. Define the *generalized eigenspaces* $\widetilde{E}_i$ of L to be $\widetilde{E}_i = \mathcal{N}(L-\lambda_i I)^{m_i}$. (The eigenspaces are $E_{\lambda_i} = \mathcal{N}(L-\lambda_i I)$. Vectors in $\widetilde{E}_i \backslash E_{\lambda_i}$ are sometimes called *generalized eigenvectors*). Then

$$\dim \widetilde{E}_i = m_i \ (1 \le i \le k) \quad \text{and} \quad V = \bigoplus_{i=1}^{k} \widetilde{E}_i.$$

This follows easily upon choosing a basis for V in which L is represented as a block-diagonal upper triangular matrix as above. Let P_i $(1 \le i \le k)$ be the projections associated with this decomposition of V, and define $D = \sum_{i=1}^k \lambda_i P_i$. Then clearly by using the same basis that represents L as a block-diagonal upper triangular matrix, the matrix for D is diagonal. Using this same basis, the matrix of $N \equiv L - D$ is strictly upper triangular, and thus N is nilpotent (in fact $N^m = 0$ where $m = \max m_i$); moreover, $N = \sum_{i=1}^k N_i$ where $N_i = P_i N P_i$; also $L\widetilde{E}_i \subset \widetilde{E}_i$, and $LD = DL$ since D is a multiple of the identity on each of the L-invariant subspaces $\widetilde{E}_i$, and thus also $ND = DN$. We have proved:

Theorem. *Any $L \in \mathcal{L}(V)$ can be written as $L = D + N$ where D is diagonalizable, N is nilpotent, and $DN = ND$. If P_i is the projection onto the λ_i-generalized eigenspace and $N_i = P_i N P_i$, then $D = \sum_{i=1}^k \lambda_i P_i$ and $N = \sum_{i=1}^k N_i$. Moreover,*

$$LP_i = P_i L = P_i L P_i = \lambda_i P_i + N_i \ (1 \le i \le k),$$

$$P_i P_j = \delta_{ij} P_i \text{ and } P_i N_j = N_j P_i = \delta_{ij} N_j \ (1 \le i \le k)(1 \le j \le k),$$

and

$$N_i N_j = N_j N_i = 0 \quad (1 \le i < j \le k),$$

where $\delta_{ij} = 0$ if $i \ne j$ and $\delta_{ij} = 1$ if $i = j$.

Note: D and N are uniquely determined by L, but we will not prove this here.

If V has an inner product $\langle \cdot, \cdot \rangle$, and L is normal, then we know that L is diagonalizable, so $N = 0$. In this case we know that eigenvectors corresponding to different eigenvalues are orthogonal, so the subspaces $\widetilde{E}_i$ $(= E_{\lambda_i}$ here) are mutually orthogonal in V. The associated projections P_i are orthogonal projections (as $E_{\lambda_i}^{\perp} = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_{i-1}} \oplus E_{\lambda_{i+1}} \oplus \cdots \oplus E_{\lambda_k}$).

Recall that any $P \in \mathcal{L}(V)$ satisfying $P^2 = P$ is a projection: one has $V = \mathcal{R}(P) \oplus \mathcal{N}(P)$, and P is the projection of V onto $\mathcal{R}(P)$ along $\mathcal{N}(P)$. Recall also that P is called an *orthogonal projection* if $\mathcal{R}(P) \perp \mathcal{N}(P)$.

Proposition. A projection P is orthogonal iff it is self-adjoint (i.e., P is Hermitian: $P^* = P$, where P^* is the adjoint of P with respect to the inner product $\langle \cdot, \cdot \rangle$).

Proof. Let $P \in \mathcal{L}(V)$ be a projection. If $P^* = P$, then $\langle Px, y \rangle = \langle x, Py \rangle \ \forall x, y \in V$. So $y \in \mathcal{N}(P)$ iff $(\forall x \in V)$ $\langle Px, y \rangle = \langle x, Py \rangle = 0$ iff $y \in \mathcal{R}(P)^{\perp}$, so P is an orthogonal projection. Conversely, suppose $\mathcal{R}(P) \perp \mathcal{N}(P)$. We must show that $\langle Px, y \rangle = \langle x, Py \rangle$ for all $x, y \in V$. Since $V = \mathcal{R}(P) \oplus \mathcal{N}(P)$, it suffices to check this separately in the four cases $x, y \in \mathcal{R}(P), \mathcal{N}(P)$. Each of these cases is straightforward since $Pv = v$ for $v \in \mathbb{R}(P)$ and $Pv = 0$ for $v \in \mathcal{N}(P)$. □

Jordan form depends discontinuously on A

Ignoring the reordering question, the Jordan form of A is discontinuous at every matrix A except those with distinct eigenvalues. For example, when $\epsilon = 0$, the Jordan form of $\begin{pmatrix} \epsilon & 1 \\ 0 & 0 \end{pmatrix}$ is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, but for $\epsilon \neq 0$, the Jordan form is $\begin{pmatrix} \epsilon & 0 \\ 0 & 0 \end{pmatrix}$. So small perturbations in A can significantly change the Jordan form. For this reason, the Jordan form is almost never used for numerical computation.

Jordan Form over $\mathbb{R}$

The previous results do not hold for real matrices: for example, in general a real matrix is not similar to a real upper-triangular matrix via a real similarity transformation; if it were, then its eigenvalues would be the real diagonal entries, but a real matrix need not have only real eigenvalues. However, nonreal eigenvalues are the only obstruction to carrying out our previous arguments. Precisely, if $A \in \mathbb{R}^{n\times n}$ has real eigenvalues, then A is orthogonally similar to a real upper triangular matrix, and A can be put into block diagonal and Jordan form using real similarity transformations, by following the same arguments as before. If A does have some nonreal eigenvalues, then there are substitute normal forms which can be obtained via real similarity transformations.

Recall that nonreal eigenvalues of a real matrix $A \in \mathbb{R}^{n\times n}$ come in complex conjugate pairs: if $\lambda = a + ib$ (with $a, b \in \mathbb{R}$, $b \neq 0$) is an eigenvalue of A, then since $p_A(t)$ has real coefficients, $0 = \overline{p_A(\lambda)} = p_A(\bar{\lambda})$, so $\bar{\lambda} = a - ib$ is also an eigenvalue. If $u + iv$ (with $u, v \in \mathbb{R}^n$) is an eigenvector of A for λ, then $A(u - iv) = A\overline{(u + iv)} = \overline{A(u + iv)} = \overline{\lambda(u + iv)} = \bar{\lambda}(u - iv)$, so $u - iv$ is an eigenvector of A for $\bar{\lambda}$. It follows that $u + iv$ and $u - iv$ (being eigenvectors for different eigenvalues) are linearly independent over $\mathbb{C}$, and thus $u = \frac{1}{2}(u + iv) + \frac{1}{2}(u - iv)$ and $v = \frac{1}{2i}(u + iv) - \frac{1}{2i}(u - iv)$ are linearly independent over $\mathbb{C}$, and consequently also over $\mathbb{R}$. Since $A(u + iv) = (a + ib)(u + iv) = (au - bv) + i(bu + av)$, $Au = au - bv$ and $Av = bu + av$. Thus span$\{u, v\}$ is a 2-dimensional real invariant subspace of $\mathbb{R}^n$ for A, and the matrix of A restricted to the subspaces span$\{u, v\}$ with respect to the basis $\{u, v\}$ is $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ (observe that this 2×2 matrix has eigenvalues $\lambda, \bar{\lambda}$).

Over $\mathbb{R}$, the best one can generally do is to have such 2×2 diagonal blocks instead of upper triangular matrices with $\lambda, \bar{\lambda}$ on the diagonal. For example, the real Jordan blocks for $\lambda, \bar{\lambda}$ are

$$J_\ell(\lambda, \bar{\lambda}) = \begin{bmatrix} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & & 0 \\ & \ddots & \ddots & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \bigcirc & & & \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{2\ell \times 2\ell}.$$

The real Jordan form of $A \in \mathbb{R}^{n\times n}$ is a direct sum of such blocks, with the usual Jordan blocks for the real eigenvalues. See H-J for details.