CHAPTER 2

Unitary similarity and unitary equivalence

2.0 Introduction

†In Chapter 1, we made an initial study of similarity of $A \in M_n$ via a general nonsingular matrix S, that is, the transformation $A \to S^{-1}AS$. For certain very special nonsingular matrices, called *unitary matrices*, the inverse of S has a simple form: $S^{-1} = S^*$. Similarity via a unitary matrix U, $A \to U^*AU$, is not only conceptually simpler than general similarity (the conjugate transpose is much easier to compute than the inverse), but it also has superior stability properties in numerical computations. A fundamental property of unitary similarity is that every $A \in M_n$ is unitarily similar to an upper triangular matrix whose diagonal entries are the eigenvalues of A. This triangular form can be further refined under general similarity; we study the latter in Chapter 3.

The transformation $A \to S^*AS$, in which S is nonsingular but not necessarily unitary, is called **congruence*; we study it in Chapter 4. Notice that similarity by a unitary matrix is both a similarity and a *congruence.

For $A \in M_{n,m}$, the transformation $A \to UAV$, in which $U \in M_m$ and $V \in M_n$ are both unitary, is called *unitary equivalence*. The upper triangular form achievable under unitary similarity can be greatly refined under unitary equivalence and generalized to non-square matrices: every $A \in M_{n,m}$ is unitarily equivalent to a nonnegative diagonal matrix whose diagonal entries (the *singular values* of A) are of great importance.

2.1 Unitary matrices and the QR factorization

2.1.1 Definition. A set of vectors $\{x_1, \ldots, x_k\} \subset \mathbf{C}^n$ is *orthogonal* if $x_i^* x_j = 0$ for all $i \neq j$, $i, j = 1, \ldots, k$. If, in addition, $x_i^* x_i = 1$ for all $i = 1, \ldots, k$ (that is, the vectors are *normalized*), then the set is *orthonormal*.

Exercise. If $\{y_1, \ldots, y_k\}$ is an orthogonal set of nonzero vectors, show that the set $\{x_1, \ldots, x_k\}$ defined by $x_i = (y_i^* y_i)^{-1/2} y_i$, $i = 1, \ldots, k$, is an orthonormal set.

2.1.2 Theorem. Every orthonormal set of vectors in $\mathbf{C}^n$ is linearly independent.

Proof: Suppose that $\{x_1, \ldots, x_k\}$ is an orthonormal set, and suppose that $0 = \alpha_1 x_1 + \cdots + \alpha_k x_k$. Then $0 = (\alpha_1 x_1 + \cdots + \alpha_k x_k)^*(\alpha_1 x_1 + \cdots + \alpha_k x_k) = \Sigma_{i,j} \bar{\alpha}_i \alpha_j x_i^* x_j = \Sigma_{i=1}^k |\alpha_i|^2 x_i^* x_i = \Sigma_{i=1}^k |\alpha_i|^2$ because the vectors x_i are orthogonal and normalized. Thus, all $\alpha_i = 0$ and hence $\{x_1, \ldots, x_k\}$ is a linearly independent set. $\square$

Exercise. Show that every orthogonal set of nonzero vectors in $\mathbf{C}^n$ is linearly independent.

Exercise. If $\{x_1, \ldots, x_k\} \in \mathbf{C}^n$ is an orthogonal set, show that either $k \leq n$ or at least $k - n$ of the vectors x_i are equal to zero.

An independent set need not be orthonormal, of course, but one can apply the Gram–Schmidt orthonormalization procedure (0.6.4) to it and obtain an orthonormal set with the same span as the original set.

Exercise. Show that any nonzero subspace of $\mathbf{R}^n$ or $\mathbf{C}^n$ has an orthonormal basis (0.6.5).

2.1.3 Definition. A matrix $U \in M_n$ is *unitary* if $U^*U = I$. If, in addition, $U \in M_n(\mathbf{R})$, U is *real orthogonal*.

Exercise. Show that $U \in M_n$ and $V \in M_m$ are unitary if and only if $U \oplus V \in M_{n+m}$ is unitary.

Exercise. Verify that the matrices Q, U, and V in Problems 19, 20, and 21 in (1.3) are unitary.

The unitary matrices in M_n form a remarkable and important set. We list some of the basic equivalent conditions for U to be unitary in (2.1.4).

2.1.4 Theorem. If $U \in M_n$, the following are equivalent:

(a) U is unitary;

(b) U is nonsingular and $U^* = U^{-1}$;
(c) $UU^* = I$;
(d) U^* is unitary;
(e) The columns of U form an orthonormal set;
(f) The rows of U form an orthonormal set; and
(g) For all $x \in \mathbf{C}^n$, $\|x\|_2 = \|Ux\|_2$, that is, x and Ux have the same Euclidean norm.

Proof: (a) implies (b) since U^{-1} (when it exists) is the unique matrix, left multiplication by which produces I (0.5); the definition of unitary says that U^* is such a matrix. Since $BA = I$ if and only if $AB = I$ (for $A, B \in M_n$ (0.5)), (b) implies (c). Since $(U^*)^* = U$, (c) implies that U^* is unitary; that is, (c) implies (d). The converse of each of these implications is similarly observed, so (a)–(d) are equivalent.

Partition $U = [u_1 \ \dots \ u_n]$ according to its columns. Then $U^*U = I$ means that $u_i^* u_i = 1$ for all $i = 1, \dots, n$ and $u_i^* u_j = 0$ for all $i \neq j$. Thus, $U^*U = I$ is another way of saying that the columns of U are orthonormal, and hence (a) is equivalent to (e). Similarly, (d) and (f) are equivalent.

If (a) holds and $y = Ux$, then $y^*y = x^*U^*Ux = x^*Ix = x^*x$, so (a) implies (g). To prove the converse, let $U^*U = A = [a_{ij}]$, let $z, w \in \mathbf{C}$ be given, and take $x = z + w$ in (g). Then $x^*x = z^*z + w^*w + 2\operatorname{Re} z^*w$ and $y^*y = x^*Ax = z^*Az + w^*Aw + 2\operatorname{Re} z^*Aw$; (g) ensures that $z^*z = z^*Az$ and $w^*w = w^*Aw$, and hence $\operatorname{Re} z^*w = \operatorname{Re} z^*Aw$ for any z and w. Take $z = e_p$ and $w = ie_q$ and compute $\operatorname{Re} ie_p^T e_q = 0 = \operatorname{Re} ie_p^T Ae_q = \operatorname{Re} ia_{pq} = -\operatorname{Im} a_{pq}$, so every entry of A is real. Finally, take $z = e_p$ and $w = e_q$ and compute $e_p^T e_q = \operatorname{Re} e_p^T e_q = \operatorname{Re} e_p^T Ae_q = a_{pq}$, which tells us that $A = I$ and U is unitary. $\square$

2.1.5 Definition. A linear transformation $T : \mathbf{C}^n \to \mathbf{C}^m$ is called a *Euclidean isometry* if $x^*x = (Tx)^*(Tx)$ for all $x \in \mathbf{C}^n$. Theorem (2.1.4) says that a square complex matrix $U \in M_n$ is a Euclidean isometry (via $U : x \to Ux$) if and only if it is unitary. See (5.2) for other kinds of isometries.

Exercise. Let $T(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$, in which θ is a real parameter. (a) Show that a given $U \in M_2(\mathbf{R})$ is real orthogonal if and only if either $U = T(\theta)$ or $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} T(\theta)$ for some $\theta \in \mathbf{R}$. (b) Show that a given $U \in M_2(\mathbf{R})$ is real orthogonal if and only if either $U = T(\theta)$ or $U = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} T(\theta)$ for some $\theta \in \mathbf{R}$. These are two different presentations, involving a parameter θ, of the 2-by-2 real orthogonal matrices. Interpret them geometrically.

2.1.6 Observation. If $U, V \in M_n$ are unitary (respectively, real orthogonal), then UV is also unitary (respectively, real orthogonal).

Exercise. Use (b) of (2.1.4) to prove (2.1.6).

2.1.7 Observation. The set of unitary (respectively, real orthogonal) matrices in M_n forms a group. This group is generally referred to as the n-by-n *unitary* (respectively, *orthogonal*) *group*, a subgroup of $GL(n, \mathbf{C})$ (0.5).

Exercise. A *group* is a set that is *closed* under a single *associative* binary operation ("multiplication") and is such that the *identity* for and *inverses* under the operation are contained in the set. Verify (2.1.7). *Hint:* Use (2.1.6) for closure; matrix multiplication is associative; $I \in M_n$ is unitary; and $U^* = U^{-1}$ is again unitary.

The set (group) of unitary matrices in M_n has another very important property. Notions of "convergence" and "limit" of a sequence of matrices will be presented precisely in Chapter 5, but can be understood here in terms of "convergence" and "limit" of entries. The defining identity $U^*U = I$ means that every column of U has Euclidean norm 1, and hence no entry of $U = [u_{ij}]$ can have absolute value greater than 1. If we think of the set of unitary matrices as a subset of $\mathbf{C}^{n^2}$, this says it is a *bounded* subset. If $U_k \equiv [u_{ij}^{(k)}]$ is an infinite sequence of unitary matrices, $k = 1, 2, \ldots$ such that $\lim_{k\to\infty} u_{ij}^{(k)} \equiv u_{ij}$ exists for all $i, j = 1, 2, \ldots, n$, then from the identity $U_k^* U_k = I$ for all $k = 1, 2, \ldots$ we see that $\lim_{k\to\infty} U_k^* U_k = U^*U = I$, in which $U = [u_{ij}]$. Thus, the limit matrix U is also unitary. This says that the set of unitary matrices is a *closed* subset of $\mathbf{C}^{n^2}$.

Since a closed and bounded subset of a finite dimensional Euclidean space is a *compact* set (see Appendix E), we conclude that the set (group) of unitary matrices in M_n is compact. For our purposes, the most important consequence of this observation is the following *selection principle* for unitary matrices.

2.1.8 Lemma. Let $U_1, U_2, \ldots \in M_n$ be a given infinite sequence of unitary matrices. There exists an infinite subsequence $U_{k_1}, U_{k_2}, \ldots$, $1 \le k_1 < k_2 < \cdots$, such that all of the entries of U_{k_i} converge (as sequences of complex numbers) to the entries of a unitary matrix as $i \to \infty$.

Proof: All that is required here is the fact that from any infinite sequence in a compact set one may always select a convergent subsequence. We have already observed that if a sequence of unitary matrices converges to some matrix, then the limit matrix must be unitary. $\square$

The unitary limit guaranteed by the lemma need not be unique; it can depend upon the subsequence chosen.

Exercise. Consider the sequence of unitary matrices $U_k = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^k, k = 1, 2, \ldots$. Show that there are two possible limits of subsequences.

Exercise. Explain why the selection principle (2.1.8) applies as well to the (real) orthogonal group; that is, an infinite sequence of real orthogonal matrices has an infinite subsequence that converges to a real orthogonal matrix.

A unitary matrix U has the property that U^{-1} *equals* U^*. One way to generalize the notion of a unitary matrix is to require that U^{-1} be *similar* to U^*. The set of such matrices is easily characterized as the range of the mapping $A \to A^{-1}A^*$ for all nonsingular $A \in M_n$.

2.1.9 Theorem. Let $A \in M_n$ be nonsingular. Then A^{-1} is similar to A^* if and only if there is a nonsingular $B \in M_n$ such that $A = B^{-1}B^*$.

Proof: If $A = B^{-1}B^*$ for some nonsingular $B \in M_n$, then $A^{-1} = (B^*)^{-1}B$ and $B^*A^{-1}(B^*)^{-1} = B(B^*)^{-1} = (B^{-1}B^*)^* = A^*$, so A^{-1} is similar to A^* via the similarity matrix B^*. Conversely, if A^{-1} is similar to A^*, then there is a nonsingular $S \in M_n$ such that $SA^{-1}S^{-1} = A^*$ and hence $S = A^*SA$. Set $S_\theta \equiv e^{i\theta}S$ for $\theta \in \mathbf{R}$, so that $S_\theta = A^*S_\theta A$ and $S_\theta^* = A^*S_\theta^*A$. Adding these two identities gives $H_\theta = A^*H_\theta A$, in which $H_\theta \equiv S_\theta + S_\theta^*$ is Hermitian. If H_θ were singular, there would be a nonzero $x \in \mathbf{C}^n$ such that $0 = H_\theta x = S_\theta x + S_\theta^* x$, so $-x = S_\theta^{-1}S_\theta^* x = e^{-2i\theta}S^{-1}S^*x$ and $S^{-1}S^*x = -e^{2i\theta}x$. Choose a value of $\theta = \theta_0 \in [0, 2\pi)$ such that $-e^{2i\theta_0}$ is not an eigenvalue of $S^{-1}S^*$; the resulting Hermitian matrix $H \equiv H_{\theta_0}$ is nonsingular and has the property that $H = A^*HA$.

Now choose any complex α such that $|\alpha| = 1$ and α is not an eigenvalue of A^*. Set $B \equiv \beta(\alpha I - A^*)H$, in which the complex parameter $\beta \neq 0$ is to be chosen, and observe that B is nonsingular. We want to have $A = B^{-1}B^*$, or $BA = B^*$. Compute $B^* = H(\bar{\beta}\bar{\alpha}I - \bar{\beta}A)$, and $BA = \beta(\alpha I - A^*)HA = \beta(\alpha HA - A^*HA) = \beta(\alpha HA - H) = H(\alpha\beta A - \beta I)$. We shall be done if we can select a nonzero β such that $\beta = -\bar{\beta}\bar{\alpha}$, but if $\alpha = e^{i\psi}$, then $\beta = e^{i(\pi - \psi)/2}$ will do. $\square$

If a unitary matrix is presented as a 2-by-2 block matrix, then the ranks of its off-diagonal blocks are equal; the ranks of its diagonal blocks are related by a simple formula.

2.1.10 Lemma. Let a unitary $U \in M_n$ be partitioned as $U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$, in which $U_{11} \in M_k$. Then $\operatorname{rank} U_{12} = \operatorname{rank} U_{21}$ and $\operatorname{rank} U_{22} = \operatorname{rank} U_{11} + n - 2k$. In particular, $U_{12} = 0$ if and only if $U_{21} = 0$, in which case U_{11} and U_{22} are unitary.

Proof: The two assertions about rank follow immediately from the law of complementary nullities (0.7.5) using the fact that $U^{-1} = \left[\begin{smallmatrix} U_{11}^* & U_{21}^* \\ U_{12}^* & U_{22}^* \end{smallmatrix}\right]$. $\square$

Plane rotations and Householder matrices are special (and very simple) unitary matrices that play an important role in establishing some basic matrix factorizations.

2.1.11 Example. plane rotations. Let

$$U(\theta;i,j) = \begin{array}{cc}
\left[\begin{array}{cccccccccc}
1 & & & & & & & & & \\
 & \ddots & & & & 0 & & & & \\
 & & 1 & & & & & & & \\
 & & & \cos\theta & 0 & \ldots & 0 & -\sin\theta & & \\
 & & & 0 & 1 & & & & & \\
 & 0 & & \vdots & & \ddots & & & 0 & \\
 & & & 0 & 0 & & 1 & & & \\
 & & & \sin\theta & 0 & \ldots & 0 & \cos\theta & & \\
 & & & \vdots & & 0 & & & \ddots & \\
 & & & & & & & & & 1
\end{array}\right] &
\begin{array}{l}
\\ \\ \\ \text{--- row } i \\ \\ \\ \\ \text{--- row } j \\ \\ \\
\end{array} \\
\begin{array}{cc} \text{column } i & \qquad\qquad \text{column } j \end{array} &
\end{array}$$

This is simply the identity matrix, with the i, i and j, j entries replaced by $\cos\theta$ and the i, j entry (respectively j, i entry) replaced by $-\sin\theta$ (respectively, $\sin\theta$).

Exercise. Verify that $U(\theta; i, j) \in M_n(\mathbf{R})$ is real orthogonal for any pair of indices $1 \le i < j \le n$ and any parameter $0 \le \theta < 2\pi$. The matrix $U(\theta; i, j)$ carries out a rotation (through an angle θ) in the i, j coordinate plane of $\mathbf{R}^n$. Left multiplication by $U(\theta; i, j)$ affects only rows i and j of the matrix multiplied; right multiplication by $U(\theta; i, j)$ affects only columns i and j of the matrix multiplied.

2.1.12 Example. Householder matrices. Let $w \in \mathbf{C}^n$ be a nonzero vector.

The *Householder matrix* $U_w \in M_n$ is defined by $U_w = I - 2(w^*w)^{-1}ww^*$. If w is a unit vector, then $U_w = I - 2ww^*$.

Exercise. Show that a Householder matrix U_w is both unitary and Hermitian; if $w \in \mathbf{R}^n$ then U_w is real orthogonal and symmetric.

Exercise. Show that a Householder matrix U_w acts as the identity on the subspace $w^\perp$ and that it acts as a reflection on the one-dimensional subspace spanned by w; that is, $U_w x = x$ if $x \perp w$ and $U_w w = -w$.

Exercise. Use (0.8.5.11) to show that $\det U_w = -1$ for all n. Thus, for all n and every nonzero $w \in \mathbf{R}^n$, the Householder matrix $U_w \in M_n(\mathbf{R})$ is a real orthogonal matrix that is never a *proper rotation matrix* (a real orthogonal matrix whose determinant is $+1$).

Exercise. Use (1.2.8) to show that the eigenvalues of a Householder matrix are always $-1, 1, \ldots, 1$ and explain why its determinant is always -1.

Householder matrices provide a simple way to construct a unitary matrix that takes a given vector into any other vector that has the same Euclidean norm.

2.1.13 Theorem. Let $x, y \in \mathbf{C}^n$ be given and suppose $\|x\|_2 = \|y\|_2$. If $y = e^{i\theta}x$ for some real θ, let $U(y,x) = e^{i\theta}I_n$; otherwise let $\phi \in [0, 2\pi)$ be such that $x^*y = e^{i\phi}|x^*y|$ (take $\phi = 0$ if $x^*y = 0$), let $w = e^{i\phi}x - y$, and let $U(y,x) = e^{i\phi}U_w$, in which U_w is the Householder matrix $U_w = I - 2(w^*w)^{-1}ww^*$. Then $U(y,x)$ is unitary and essentially Hermitian, $U(y,x)x = y$, and $U(y,x)z \perp y$ whenever $z \perp x$. If x and y are real, then $U(y,x)$ is real orthogonal: $U(y,x) = I$ if $y = x$, and $U(y,x)$ is the real Householder matrix U_{x-y} otherwise.

Proof: The assertions are readily verified if x and y are linearly dependent, that is, if $y = e^{i\theta}x$ for some real θ. If x and y are linearly independent, the Cauchy-Schwarz inequality (0.6.3) ensures that $x^*x \neq |x^*y|$. Compute

$$\begin{aligned} w^*w &= (e^{i\phi}x - y)^*(e^{i\phi}x - y) = x^*x - e^{-i\phi}x^*y - e^{i\phi}y^*x + y^*y \\ &= 2(x^*x - \operatorname{Re}(e^{-i\phi}x^*y)) = 2(x^*x - |x^*y|) \end{aligned}$$

and

$$w^*x = e^{-i\phi}x^*x - y^*x = e^{-i\phi}x^*x - e^{-i\phi}|y^*x| = e^{-i\phi}(x^*x - |x^*y|)$$

and, finally,

$$e^{i\phi}U_w x = e^{i\phi}(x - 2(w^*w)^{-1}ww^*x) = e^{i\phi}(x - (e^{i\phi}x - y)e^{-i\phi}) = y$$

If z is orthogonal to x, then $w^*z = -y^*z$ and

$$\begin{aligned} y^*U(y,x)z &= e^{i\phi}\left(y^*z - \frac{1}{\|x\|_2^2 - |x^*y|}\left(e^{i\phi}y^*x - \|y\|_2^2\right)(-y^*x)\right) \\ &= e^{i\phi}\left(y^*z + (-y^*x)\right) = 0 \end{aligned}$$

Since U_w is unitary and Hermitian, $U(y,x) = (e^{i\phi}I)U_w$ is unitary (as a product of two unitary matrices) and essentially Hermitian (0.2.5). □

Exercise. Let $y \in \mathbf{C}^n$ be a given unit vector and let e_1 be the first column of the n-by-n identity matrix. Construct $U(y, e_1)$ using the recipe in the preceding theorem and verify that its first column is y (which it should be, since $y = U(y, e_1)e_1$).

Exercise. Let $x \in \mathbf{C}^n$ be a given nonzero vector. Explain why the matrix $U(\|x\|_2\, e_1, x)$ constructed in the preceding theorem is an essentially Hermitian unitary matrix that takes x into $\|x\|_2\, e_1$.

The following *QR factorization* of a complex or real matrix is of considerable theoretical and computational importance.

2.1.14 Theorem. (QR factorization) Suppose $A \in M_{n,m}$ and $n \geq m$. Then
(a) There is a $Q \in M_{n,m}$ with orthonormal columns and an upper triangular $R \in M_m$ with non-negative main diagonal entries such that $A = QR$.
(b) If $\operatorname{rank} A = m$, then the factors Q and R in (a) are uniquely determined and the main diagonal entries of R are all positive.
(c) If $m = n$, then the factor Q in (a) is unitary.
(d) If A is real, then both of the factors Q and R in (a) may be taken to be real.

Proof: Let $a_1 \in \mathbf{C}^n$ be the first column of A, let $r_1 = \|a_1\|_2$, and let U_1 be a unitary matrix such that $U_1 a_1 = r_1 e_1$. Theorem (2.1.13) gives an explicit construction for such a matrix, which is either a unitary scalar matrix or the product of a unitary scalar matrix and a Householder matrix. Partition

$$U_1 A = \begin{bmatrix} r_1 & \bigstar \\ 0 & A_2 \end{bmatrix}$$

in which $A_2 \in M_{n-1,m-1}$. Let $a_2 \in \mathbf{C}^{n-1}$ be the first column of A_2 and let $r_2 = \|a_2\|_2$. Use (2.1.13) again to construct a unitary $V_2 \in M_{n-1}$ such that $V_2 a_2 = r_2 e_1$ and let $U_2 = [I_1] \oplus V_2$. Then

$$U_2 U_1 A = \begin{bmatrix} r_1 & & \bigstar \\ 0 & r_2 & \\ 0 & 0 & A_3 \end{bmatrix}$$

Repeat this construction m times to obtain

$$U_m U_{m-1} \cdots U_2 U_1 A = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

in which $R \in M_m$ is upper triangular. Its main diagonal entries are $r_1, \ldots, r_m$ are all nonnegative. Let $U = U_m U_{m-1} \cdots U_2 U_1$. Partition $U^* = U_1^* U_2^* \cdots U_{m-1}^* U_m^* = [Q \;\; Q_2]$, in which $Q \in M_{n,m}$ has orthonormal columns since it comprises the first m columns of a unitary matrix. Then $A = QR$, as desired. If A has full column rank, then R is nonsingular, so its main diagonal entries are all positive.

Suppose that $\operatorname{rank} A = m$ and $A = QR = \tilde{Q}\tilde{R}$, in which R and $\tilde{R}$ are upper triangular and have positive main diagonal entries, and Q and $\tilde{Q}$ have orthonormal columns. Then $A^*A = R^*(Q^*Q)R = R^*IR = R^*R$ and also $A^*A = \tilde{R}^*\tilde{R}$, so $R^*R = \tilde{R}^*\tilde{R}$ and $\tilde{R}^{-*}R^* = \tilde{R}R^{-1}$. This says that a lower triangular matrix equals an upper triangular matrix, so both must be diagonal: $\tilde{R}R^{-1} = D$ is diagonal, and it must have positive main diagonal entries because the main diagonal entries of both $\tilde{R}$ and R^{-1} are positive. But $\tilde{R} = DR$ implies that $D = \tilde{R}R^{-1} = \tilde{R}^{-*}R^* = (DR)^{-*}R^* = D^{-1}R^{-*}R^* = D^{-1}$, so $D^2 = I$ and hence $D = I$. We conclude that $\tilde{R} = R$ and hence $\tilde{Q} = Q$.

The assertion in (c) follows from the fact that a square matrix with orthonormal columns is unitary. The final assertion (d) follows from the construction in (a) and the assurance in (2.1.13) that the unitary matrices U_i may all be chosen to be real. □

Exercise. Show that any $B \in M_n$ of the form $B = A^*A$, $A \in M_n$, may be written as $B = LL^*$, in which $L \in M_n$ is lower triangular and has nonnegative diagonal entries. Explain why this factorization is unique if A is nonsingular. This is called the *Cholesky factorization* of B; every positive definite matrix may be factored in this way (see Chapter 7).

For square matrices $A \in M_n$, there are some easy variants of the QR factorization that can be useful. Let K be the (real orthogonal and symmetric) n-by-n reversal matrix (0.9.5.1), which has the pleasant property that $K^2 = I$. Moreover, KRK is lower triangular if R is upper triangular (the main diagonal entries are the same, but the order is reversed), and of course KLK is upper triangular if L is lower triangular. If we factor $KAK = QR$ as in (2.1.14), then $A = (KQK)(KRK) = Q_1 L$, in which $Q_1 = KQK$ is unitary and L is lower triangular with nonnegative main diagonal entries; we call this a QL *factorization* of A. Now let $A^* = QL$ be a QL factorization of A^*, and observe that $A = L^*Q^*$, which is an RQ *factorization* of A. Finally, factoring

$KA^*K = QL$ gives $A = (KLK)^*(BQB)^*$, which is an *LQ factorization* of A.

2.1.15 Corollary. Let $A \in M_n$ be given. Then there are unitary matrices Q_1, Q_2, Q_3, lower triangular matrices L_2, L_3 with nonnegative main diagonal entries, and an upper triangular matrix R_2 with nonnegative main diagonal entries such that $A = Q_1L_1 = R_2Q_2 = L_3Q_3$. If A is nonsingular, then the respective unitary and triangular factors are uniquely determined and the main diagonal entries of the triangular factors are all positive. If A is real, then all of the factors $Q_1, Q_2, Q_3, L_2, L_3, R_2$ may be chosen to be real.

Problems

1. If $U \in M_n$ is unitary, show that $|\det U| = 1$.

2. Let $U \in M_n$ be unitary and let λ be a given eigenvalue of U. Show that (a) $|\lambda| = 1$ and (b) x is a (right) eigenvector of U corresponding to λ if and only if x is a left eigenvector of U corresponding to λ. *Hint:* Use (2.1.4g) and Problem 1 in (1.1).

3. Given real parameters $\theta_1, \theta_2, \ldots, \theta_n$, show that $U = \operatorname{diag}(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_n})$ is unitary. Show that every diagonal unitary matrix has this form.

4. Characterize the diagonal real orthogonal matrices.

5. Show that the permutation matrices (0.9.5) in M_n are a subgroup (a subset that is itself a group) of the group of real orthogonal matrices. How many different permutation matrices are there in M_n?

6. Give a presentation in terms of parameters of the 3-by-3 orthogonal group. Two presentations of the 2-by-2 orthogonal group are given in (2.1).

7. Suppose $A, B \in M_n$ and $AB = I$. Provide details for the following argument that $BA = I$: Every $y \in \mathbf{C}^n$ can be represented as $y = A(By)$, so $\operatorname{rank} A = n$ and hence $\dim(\text{nullspace}(A)) = 0$ (0.2.3.1). Compute $A(AB - BA) = A(I - BA) = A - (AB)A = A - A = 0$, so $AB - BA = 0$.

8. A matrix $A \in M_n$ is *complex orthogonal* if $A^TA = I$. A real orthogonal matrix is unitary, but a nonreal orthogonal matrix need not be unitary. (a) Let $K = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in M_2(\mathbf{R})$. Show that $A(t) = (\cosh t)I + (i \sinh t)K \in M_2$ is complex orthogonal for all $t \in \mathbf{R}$, but that $A(t)$ is unitary only for $t = 0$. The hyperbolic functions are defined by $\cosh t = (e^t + e^{-t})/2$, $\sinh t = (e^t - e^{-t})/2$. (b) Show that, unlike the unitary matrices, the set of complex orthogonal matrices is not a bounded set, and it is therefore not a compact set. (c) Show that the set of complex orthogonal matrices of a given size forms a group. The smaller (and compact) group of *real* orthogonal matrices of a given size is often called the *orthogonal group*. (d) If $A \in M_n$ is complex orthogonal, show that $|\det A| = 1$; consider $A(t)$ in (a) to show that A can have eigenvalues λ with $|\lambda| \neq 1$. (e) If $A \in M_n$ is complex orthogonal, show that $\bar{A}$, A^T, and A^* are all complex orthogonal and nonsingular. Do the rows or columns of A form an orthogonal set? (f) Characterize the diagonal complex orthogonal matrices. Compare with Problem 4. (g) Show that $A \in M_n$ is both complex orthogonal and unitary if and only if it is real orthogonal.

9. If $U \in M_n$ is unitary, show that $\bar{U}$, U^T, and U^* are all unitary.

10. If $U \in M_n$ is unitary, show that $x, y \in \mathbf{C}^n$ are orthogonal if and only if *Ux* and *Uy* are orthogonal.

11. A nonsingular matrix $A \in M_n$ is *skew orthogonal* if $A^{-1} = -A^T$. Show that A is skew-orthogonal if and only if $\pm iA$ is orthogonal. More generally, if $\theta \in \mathbf{R}$, show that $A^{-1} = e^{i\theta} A^T$ if and only if $e^{i\theta/2} A$ is orthogonal. What is this for $\theta = \pi$? for $\theta = 0$?

12. Show that if $A \in M_n$ is similar to a unitary matrix, then A^{-1} is similar to A^*.

13. Consider diag $(2, \frac{1}{2}) \in M_2$ and show that the set of matrices that are similar to unitary matrices is a proper subset of the set of matrices A for which A^{-1} is similar to A^*.

14. Show that the intersection of the group of unitary matrices in M_n with the group of complex orthogonal matrices in M_n is the group of real orthogonal matrices in M_n. *Hint:* $U^{-1} = U^T = U^*$.

15. If $U \in M_n$ is unitary, $\alpha \subset \{1, \ldots, n\}$, and $U[\alpha, \alpha^c] = 0$, (0.7.1) show that $U[\alpha^c, \alpha] = 0$, and $U[\alpha]$ and $U[\alpha^c]$ are unitary.

16. Let $x, y \in \mathbf{R}^n$ be given linearly independent unit vectors and let $w = x + y$. Consider the *Palais matrix* $P_{x,y} = I - 2(w^T w)^{-1} w w^T + 2 y x^T$. Show that: (a) $P_{x,y} = (I - 2(w^T w)^{-1} w w^T)(I - 2 x x^T) = U_w U_x$ is a product of two real Householder matrices, so it is a real orthogonal matrix; (b) $\det P_{x,y} = +1$, so $P_{x,y}$ is always a proper rotation matrix; (c) $P_{x,y} x = y$ and $P_{x,y} y = -x + 2(x^T y) y$; (d) $P_{x,y} z = z$ if $z \in \mathbf{R}^n$, $z \perp x$, and $z \perp y$; (e) $P_{x,y}$ acts as the identity on the $(n-2)$-dimensional subspace $(\operatorname{span}\{x, y\})^{\perp}$ and it is a proper rotation on the 2-dimensional subspace $\operatorname{span}\{x, y\}$ that takes x into y; (f) If $n = 3$, explain why $P_{x,y}$ is the unique proper rotation that takes x into y and leaves fixed their vector cross product $x \times y$; (g) the eigenvalues of $P_{x,y}$ are $x^T y \pm i(1 - (x^T y)^2)^{1/2} = e^{\pm i\theta}, 1, \ldots, 1$, in which $\cos\theta = x^T y$.

Hint: (1.3.23) the eigenvalues of $[w \;\; x]^T[-(w^Tw)^{-1}w \;\; y] \in M_2(\mathbf{R})$ are $\frac{1}{2}(x^Ty - 1 \pm i(1-(x^Ty)^2)^{1/2})$.

17. Suppose that $A \in M_{n,m}$, $n \geq m$, and $\operatorname{rank} A = m$. Describe the steps of the Gram-Schmidt process applied to the columns of A, proceeding from left to right. Explain why this process produces, column-by-column, an explicit matrix $Q \in M_{n,m}$ with orthonormal columns and an explicit upper triangular matrix $R \in M_m$ such that $Q = AR$. How is this factorization related to the one in (2.1.14)?

18. Let $A \in M_n$ be factored as $A = QR$ as in (2.1.14), partition $A = [a_1 \; \ldots \; a_n]$ and $Q = [q_1 \; \ldots \; q_n]$ according to their columns and let $R = [r_{ij}]_{i,j=1}^n$. (a) Explain why $\{q_1, \ldots, q_k\}$ is an orthonormal basis for $\operatorname{span}\{a_1, \ldots, a_k\}$ for each $k = 1, \ldots, n$. (b) Show that r_{kk} is the Euclidean distance from a_k to $\operatorname{span}\{a_1, \ldots, a_{k-1}\}$ for each $k = 2, \ldots, n$.

19. Let $X = [x_1 \; \ldots \; x_m] \in M_{n,m}$, suppose $\operatorname{rank} X = m$, and factor $X = QR$ as in (2.1.14). Let $Y = QR^{-*} = [y_1 \; \ldots \; y_m]$. Show that the columns of Y are a basis for the subspace $\mathcal{S} = \operatorname{span}\{x_1, \ldots, x_m\}$ and that $Y^*X = I_m$, so $y_i^*x_j = 0$ if $i \neq j$ and each $y_i^*x_i = 1$. One says that $\{y_1, \ldots, y_m\}$ is the basis of $\mathcal{S}$ that is *dual* (*reciprocal*) to the basis $\{x_1, \ldots, x_m\}$.

20. If $U \in M_n$ is unitary, show that $\operatorname{adj} U = \det(U)U^*$.

21. Explain why Lemma 2.1.10 remains true if "unitary" is replaced with "complex orthogonal."

22. Suppose that $X, Y \in M_{n,m}$ have orthonormal columns. Show that X and Y have the same range (column space) if and only if there is a unitary $U \in M_m$ such that $X = YU$. *Hint:* (0.2.7).

23. Let $A \in M_n$, let $A = QR$ be a QR factorization, let $R = [r_{ij}]$, and partition both A and R according to their columns: $A = [a_1 \; \ldots \; a_n]$ and $R = [r_1 \; \ldots \; r_n]$. Explain why $\|a_i\|_2 = \|r_i\|_2$ for each $i = 1, \ldots, n$, $|\det A| = \det R = r_{11} \cdots r_{nn}$, and $r_{ii} \leq \|r_i\|_2$ for each $i = 1, \ldots, n$. Conclude that $|\det A| \leq \prod_{i=1}^n \|a_i\|_2$. This is known as *Hadamard's inequality*.

Further Reading. For more information about matrices that satisfy the conditions of (2.1.9), see C. R. DePrima and C. R. Johnson, The Range of $A^{-1}A^*$ in $GL(n, \mathbf{C})$, *Linear Algebra Appl.* 9 (1974) 209–222.

2.2 Unitary similarity

Since $U^* = U^{-1}$ for unitary U, the transformation on M_n given by $A \to U^*AU$ is a similarity transformation if U is unitary. This special type of similarity is called *unitary similarity*.

2.2.1 Definition. Let $A, B \in M_n$ be given. We say that A is *unitarily similar* to B if there is a unitary $U \in M_n$ such that $A = UBU^*$. If U may be taken to be real (and hence is real orthogonal), then A is said to be *real orthogonally similar* to B. We say that A is *unitarily diagonalizable* if it is unitarily similar to a diagonal matrix; A is *real orthogonally diagonalizable* if it is real orthogonally similar to a diagonal matrix.

Exercise. Show that unitary similarity is an equivalence relation.

2.2.2 Theorem. Let $U \in M_n$ and $V \in M_m$ be unitary, let $A = [a_{ij}] \in M_{n,m}$ and $B = [b_{ij}] \in M_{n,m}$, and suppose $A = UBV$. Then $\sum_{i,j=1}^{n,m} |b_{ij}|^2 = \sum_{i,j=1}^{n,m} |a_{ij}|^2$. In particular, this identity is satisfied if $m = n$ and $V = U^*$, that is, if A is unitarily similar to B.

Proof: It suffices to check that $\operatorname{tr} B^*B = \operatorname{tr} A^*A$. (0.2.5) Compute $\operatorname{tr} A^*A = \operatorname{tr}(UBV)^*(UBV) = \operatorname{tr}(V^*B^*U^*UBV) = \operatorname{tr} V^*B^*BV = \operatorname{tr} B^*BVV^* = \operatorname{tr} B^*B$. $\square$

Exercise. Show that the matrices $\begin{bmatrix} 3 & 1 \\ -2 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$ are similar but not unitarily similar.

Unitary similarity implies similarity, but not conversely. The unitary similarity equivalence relation partitions M_n into *finer* equivalence classes than the similarity equivalence relation. Unitary similarity, like similarity, corresponds to a change of basis, but of a special type—it corresponds to a change from one *orthonormal* basis to another.

Exercise. Using the notation of (2.1.11), explain why only rows and columns i and j are changed under real orthogonal similarity via the plane rotation $U(\theta; i, j)$.

Exercise. Using the notation of (2.1.13), explain why $U(y,x)^*AU(y,x) = U_w^*AU_w$ for any $A \in M_n$, that is, a unitary similarity via an essentially Hermitian unitary matrix of the form $U(y,x)$ is a unitary similarity via a Householder matrix. Unitary (or real orthogonal) similarity via a Householder matrix is often called a *Householder transformation*.

For computational or theoretical reasons, it is often convenient to transform a given matrix by unitary similarity into another matrix with a special form. Here are two examples.

2.2.3 Example. Suppose $A = [a_{ij}] \in M_n$ is given. We claim that there is a unitary $U \in M_n$ such that all the main diagonal entries of $U^*AU = B = [b_{ij}]$ are equal; if A is real, then U may be taken to be real orthogonal. If this claim is true, then $\operatorname{tr} A = \operatorname{tr} B = nb_{11}$, so every main diagonal entry of B is equal to the average of the main diagonal entries of A.

Begin by considering the complex case and $n = 2$. Since we can replace $A \in M_2$ by $A - (\frac{1}{2}\operatorname{tr} A)I$, there is no loss of generality to assume that $\operatorname{tr} A = 0$, in which case the two eigenvalues of A are $\pm\lambda$ for some $\lambda \in \mathbf{C}$. We wish to determine a unit vector u such that $u^*Au = 0$. If $\lambda = 0$, let u be any unit vector such that $Au = 0$. If $\lambda \neq 0$, let w and z be any unit eigenvectors associated with the distinct eigenvalues $\pm\lambda$. Let $x(\theta) = e^{i\theta}w + z$, which is nonzero for all $\theta \in \mathbf{R}$ since w and z are linearly independent. Compute $x(\theta)^*Ax(\theta) = \lambda(e^{i\theta}w + z)^*(e^{i\theta}w - z) = 2i\lambda \operatorname{Im}(e^{i\theta}z^*w)$. If $z^*w = e^{i\phi}|z^*w|$,then $x(-\phi)^*Ax(-\phi) = 0$. Let $u = x(-\phi)/\left\|x(-\phi)\right\|_2$. Now let $v \in \mathbf{C}^2$ be any unit vector that is orthogonal to u and let $U = [u\ v]$. Then U is unitary and $(U^*AU)_{11} = u^*Au = 0$. But $\operatorname{tr}(U^*AU) = 0$, so $(U^*AU)_{22} = 0$ as well.

Now suppose $n = 2$ and A is real. If the diagonal entries of $A = [a_{ij}]$ are not equal, consider the plane rotation matrix $U_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. A calculation reveals that the diagonal entries of $U_\theta A U_\theta^T$ are equal if $(\cos^2\theta - \sin^2\theta)(a_{11} - a_{22}) = 2\sin\theta\cos\theta(a_{12} + a_{21})$, so equal diagonal entries are achieved if $\theta \in (0, \pi/2)$ is chosen so that $\cot 2\theta = (a_{12} + a_{21})/(a_{11} - a_{22})$.

We have now shown that any 2-by-2 complex matrix A is unitarily similar to a matrix with both diagonal entries equal to the average of the diagonal entries of A; if A is real, the similarity may be taken to be real orthogonal.

Now suppose $n > 2$ and define $f(A) = \max\{|a_{ii} - a_{jj}| : i, j = 1, 2, \ldots, n\}$. If $f(A) > 0$, let $A_2 = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}$ for a pair of indices i, j for which $f(A) = |a_{ii} - a_{jj}|$ (there could be several pairs of indices for which this maximum positive separation is attained; choose any one of them). Let $U_2 \in M_2$ be unitary, real if A is real, and such that $U_2^* A_2 U_2$ has both main diagonal entries equal to $\frac{1}{2}(a_{ii} + a_{jj})$. Construct $U(i, j) \in M_n$ from U_2 in the same way that $U(\theta; i, j)$ was constructed from a 2-by-2 plane rotation in (2.1.11). The unitary similarity $U(i, j)^*AU(i, j)$ affects only entries in rows and columns i and j, so it leaves unchanged every main diagonal entry of A except the entries in positions i and j, which it replaces with the average $\frac{1}{2}(a_{ii} + a_{jj})$. For any $k \neq i, j$ the triangle inequality ensures that

$$\begin{aligned} |a_{kk} - \frac{1}{2}(a_{ii} + a_{jj})| &= |\frac{1}{2}(a_{kk} - a_{ii}) + \frac{1}{2}(a_{kk} - a_{jj})| \\ &\leq \frac{1}{2}|a_{kk} - a_{ii}| + \frac{1}{2}|a_{kk} - a_{jj}| \\ &\leq \frac{1}{2}f(A) + \frac{1}{2}f(A) = f(A) \end{aligned}$$

with equality only if the scalars $a_{kk} - a_{ii}$ and $a_{kk} - a_{jj}$ both lie on the same ray in the complex plane and $|a_{kk} - a_{ii}| = |a_{kk} - a_{jj}|$. These two conditions imply that $a_{ii} = a_{jj}$, so it follows that $|a_{kk} - \frac{1}{2}(a_{ii} + a_{jj})| < f(A)$ for all $k \neq i, j$. Thus, the unitary similarity we have just constructed reduces by one the finitely many pairs of indices k, ℓ for which $f(A) = |a_{kk} - a_{\ell\ell}|$. Repeat the construction, if necessary, to deal with any such remaining pairs and achieve a unitary U (real if A is real) such that $f(U^*AU) < f(A)$.

Finally, consider the compact set $R(A) = \{U^*AU : U \in M_n \text{ is unitary}\}$. Since f is a continuous nonnegative-valued function on $R(A)$, it achieves its minimum value there, that is, there is some $B \in R(A)$ such that $f(A) \geq f(B) \geq 0$ for all $A \in R(A)$. If $f(B) > 0$, we have just seen that there is a unitary U (real if A is real) such that $f(B) > f(U^*BU)$. This contradiction shows that $f(B) = 0$, so all the diagonal entries of B are equal.

2.2.4 Example. Suppose $A = [a_{ij}] \in M_n$ is given. The following construction shows that A is unitarily similar to an upper Hessenberg matrix with nonnegative entries in its first subdiagonal. Let a_1 be the first column of A, partitioned as $a_1^T = [a_{11}\ \xi^T]$ with $\xi \in \mathbf{C}^{n-1}$. Let $U_1 = I_{n-1}$ if $\xi = 0$; otherwise, use (2.1.13) to construct $U_1 = U(\|\xi\|_2\, e_1, \xi) \in M_{n-1}$, a unitary matrix that takes ξ into a positive multiple of e_1. Form the unitary matrix $V_1 = I_1 \oplus U_1$ and observe that the first column of V_1A is the vector $[a_{11}\ \ \|\xi\|_2\ \ 0]^T$. Moreover, $\mathcal{A}_1 = (V_1A)V_1^*$ has the same first column as V_1A and is unitarily similar to A. Partition it as

$$\mathcal{A}_1 = \begin{bmatrix} a_{11} & \bigstar \\ \begin{bmatrix} \|\xi\|_2 \\ 0 \end{bmatrix} & A_2 \end{bmatrix}, \quad A_2 \in M_{n-1}$$

Use (2.1.13) again to form, in the same way, a unitary matrix U_2 that takes the first column of A_2 into a vector whose entries below the second are all zero and whose second entry is nonnegative. Let $V_2 = I_2 \oplus U_2$ and let $\mathcal{A}_2 = V_2AV_2^*$. This similarity does not affect the first column of $\mathcal{A}_1$. After at most $n-1$ steps, this construction produces an upper Hessenberg matrix $\mathcal{A}_{n-1}$ that is unitarily similar to A and has nonnegative subdiagonal entries.

Exercise. If A is Hermitian or skew-Hermitian, explain why the construction in the preceding example produces a tridiagonal Hermitian or skew-Hermitian matrix that is unitarily similar to A.

Theorem (2.2.2) provides a necessary but not sufficient condition for two given matrices to be unitarily similar. It can be augmented with additional identities that collectively do provide necessary and sufficient conditions. A key role is played by the following simple notion. Let s, t be two given noncommuting variables. We refer to any finite formal product of nonnegative powers of s, t

$$W(s,t) = s^{m_1}t^{n_1}s^{m_2}t^{n_2}\cdots s^{m_k}t^{n_k}, \quad m_1, n_1, \ldots, m_k, n_k \geq 0 \qquad (2.2.5)$$

as a *word in s* and t. The *degree* of the word $W(s,t)$ is the nonnegative integer $m_1 + n_1 + m_2 + n_2 + \cdots + m_k + n_k$, that is, the sum of all the exponents in the word. If $A \in M_n$ is given, we define a *word in A and* A^* as

$$W(A, A^*) = A^{m_1}(A^*)^{n_1}A^{m_2}(A^*)^{n_2}\cdots A^{m_k}(A^*)^{n_k}$$

Since the powers of A and A^* need not commute, it may not be possible to simplify the expression of $W(A, A^*)$ by rearranging the terms in the product.

Suppose A is unitarily similar to $B \in M_n$, so that $A = UBU^*$ for some unitary $U \in M_n$. For any word $W(s,t)$ we have

$$\begin{aligned}
W(A, A^*) &= (UBU^*)^{m_1}(UB^*U^*)^{n_1}\cdots(UBU^*)^{m_k}(UB^*U^*)^{n_k}\\
&= UB^{m_1}U^*U(B^*)^{n_1}U^*\cdots UB^{m_k}U^*U(B^*)^{n_k}U^*\\
&= UB^{m_1}(B^*)^{n_k}\cdots B^{m_k}(B^*)^{n_k}U^*\\
&= UW(B, B^*)U^*
\end{aligned}$$

so $W(A, A^*)$ is unitarily similar to $W(B, B^*)$. Thus, $\operatorname{tr} W(A, A^*) = \operatorname{tr} W(B, B^*)$. If we take the word $W(s,t) =$ *ts*, we obtain the identity in (2.2.2).

If one considers all possible words $W(s,t)$, this observation gives infinitely many necessary conditions for two matrices to be unitarily similar. A theorem of W. Specht, which we state without proof, guarantees that these necessary conditions are also sufficient.

2.2.6 Theorem. Two matrices $A, B \in M_n$ are unitarily similar if and only if

$$\operatorname{tr} W(A, A^*) = \operatorname{tr} W(B, B^*) \qquad (2.2.7)$$

for every word $W(s,t)$ in two noncommuting variables.

Specht's theorem can be used to show that two matrices are *not* unitarily similar by exhibiting a specific word that violates (2.2.7). However, except in special situations (see Problem 6), it may be useless in showing that two given matrices *are* unitarily similar because infinitely many conditions must be verified. Fortunately, a refinement of Specht's theorem says that it suffices to check the trace identities (2.2.7) for only finitely many words, which gives a practical criterion to assess unitary similarity of matrices of small size.

2.2.8 Theorem. Two matrices $A, B \in M_n$ are unitarily similar if and only if $\operatorname{tr} W(A, A^*) = \operatorname{tr} W(B, B^*)$ for every word $W(s,t)$ in two noncommuting variables whose degree is at most

$$n\sqrt{\frac{2n^2}{n-1}+\frac{1}{4}}+\frac{n}{2}-2$$

For $n = 2$, it suffices to verify (2.2.7) for the three words $W(s,t) = s, s^2$, and st. For $n = 3$, it suffices to verify (2.2.7) for the seven words $W(s,t) = s, s^2, st, s^3, s^2t, s^2t^2$, and s^2t^2st.

Problems

1. Let $A = [a_{ij}] \in M_n(\mathbf{R})$ be symmetric but *not* diagonal, and suppose that indices i, j with $i < j$ are chosen so that $|a_{ij}|$ is as large as possible. Define θ by $\cot 2\theta = (a_{jj} - a_{ii})/2a_{ij}$, let $U(\theta; i, j)$ be the plane rotation (2.1.11), and let $B = U(\theta; i, j)AU(\theta; i, j)^T = [b_{pq}]$. Show that $b_{ij} = 0$ and use (2.2.2) to show that $\sum_{p\neq q} |b_{pq}|^2 < \sum_{p\neq q} |a_{pq}|^2$. Indeed, it is not necessary to compute θ; just take $\cos\theta = a_{jj}(a_{ij}^2 + a_{jj}^2)^{-1/2}$ and $\sin\theta = a_{ij}(a_{ij}^2 + a_{jj}^2)^{-1/2}$. Show that repeated real orthogonal similarities via plane rotations (chosen in the same way for B and its successors) strictly decrease the sums of the squares of the *off-diagonal* entries while preserving the sums of the squares of *all* the entries; at each step, the computed matrix is (in this sense) more nearly diagonal than at the step before. This is the method of Jacobi for calculating the eigenvalues of a real symmetric matrix. It produces a sequence of matrices that converges to a real diagonal matrix. Why must the diagonal entries of the limit be the eigenvalues of A? How can the corresponding eigenvectors be obtained?

2. The eigenvalue calculation method of Givens for real matrices also uses plane rotations, but in a different way. For $n \geq 3$, provide details for the following argument showing that every $A = [a_{ij}] \in M_n(\mathbf{R})$ is real orthogonally similar to a real lower Hessenberg matrix, which is necessarily tridiagonal if A is symmetric; see (0.9.9) and (0.9.10). Choose a plane rotation $U_{1,3}$ of the form $U(\theta; 1, 3)$, as in the preceding problem, so that the 1,3 entry of $U_{1,3}^* A U_{1,3}$ is 0. Choose another plane rotation of the form $U_{1,4} = U(\theta; 1, 4)$ so that the 1,4 entry of $U_{1,4}^*(U_{1,3}^* A U_{1,3})U_{1,4}$ is 0; continue in this way to zero out the rest of the first row with a sequence of real orthogonal similarities. Then start on the second row beginning with the 2,4 entry and zero out the 2,4, 2,5, …, 2,n entries. Explain why this process does not disturb previously manufactured 0 entries, and why it preserves symmetry if A is symmetric. Proceeding in this way through row $n-3$ produces a lower Hessenberg matrix after finitely many real orthogonal similarities via plane rotations; that matrix is tridiagonal if A is symmetric. However, the eigenvalues of A are not displayed as in Jacobi's method; they must be obtained from a further calculation.

3. Show that every $A \in M_2$ is unitarily similar to its transpose. *Hint:* Consider the three words $W(s,t) = s, s^2, st$.

4. Let

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

Any matrix is similar to its transpose (3.2.3), but A is not unitarily similar to A^T. For which of the seven words listed in (2.2.8) do A and $B = A^T$ fail the test (2.2.7)?

5. If $A \in M_n$ and there is a unitary $U \in M_n$ such that $A^* = UAU^*$, show that $A + A^* = U(A + A^*)U^*$, that is, U commutes with $A + A^*$. Apply this observation to the 3-by-3 matrix in the preceding problem and conclude that if it is unitarily similar to its transpose, then any such unitary similarity must be diagonal. Show that no diagonal unitary similarity can take this matrix into its transpose.

6. Let $A \in M_n$ and $B, C \in M_m$ be given. Use either (2.2.6) or (2.2.8) to show that B and C are unitarily similar if and only if any one of the following conditions holds:
(a) $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ and $\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$ are unitarily similar.
(b) $B \oplus \cdots \oplus B$ and $C \oplus \cdots \oplus C$ are unitarily similar if both direct sums contain the same number of terms.
(c) $A \oplus B \oplus \cdots \oplus B$ and $A \oplus C \oplus \cdots \oplus C$ are unitarily similar if both direct sums contain the same number of terms.

7. Give an example of two 2-by-2 matrices that satisfy the identity (2.2.2) but are not unitarily similar. Explain why.

8. Let $A, B \in M_2$ and let $C = AB - BA$. Use Example 2.2.3 to show that $C^2 = \lambda I$ for some scalar λ. Hint: $\operatorname{tr} C = ?$; $\begin{bmatrix} 0 & b \\ a & 0 \end{bmatrix}^2 = ?$

9. Let $A \in M_n$ and suppose $\operatorname{tr} A = 0$. Use Example 2.2.3 to show that A can be written as a sum of two nilpotent matrices. Conversely, if A can be written as a sum of nilpotent matrices, explain why $\operatorname{tr} A = 0$. *Hint:* Write $A = UBU^*$, in which $B = [b_{ij}]$ has zero main diagonal entries. Then write $B = B_L + B_R$, in which $B_L = [\beta_{ij}]$, $\beta_{ij} = b_{ij}$ if $i \geq j$ and $\beta_{ij} = 0$ if $j > i$.

10. Let $n \geq 2$ be a given integer and define $\omega = e^{2\pi i/n}$. (a) Explain why $\sum_{k=0}^{n-1} \omega^{k\ell} = 0$ unless $\ell = mn$ for some $m = 0, \pm 1, \pm 2, \ldots$, in which case the sum is equal to n. (b) Let $F_n = n^{-1/2}[\omega^{(i-1)(j-1)}]_{i,j=1}^n$ denote the n-by-n *Fourier matrix*. Show that F_n is symmetric, unitary, and coninvolutory: $F_n F_n^* = F_n \overline{F_n} = I$. (c) Let C_n denote the basic circulant permutation matrix (0.9.6.2). Explain why C_n is unitary (real orthogonal). (d) Let $D = \operatorname{diag}(1, \omega, \omega^2, \ldots, \omega^{n-1})$ and show that $C_n F_n = F_n D$, so $C_n = F_n D F_n^*$ and $C_n^k = F_n D^k F_n^*$ for all $k = 1, 2, \ldots$. (e) Let A denote the circulant matrix (0.9.6.1), expressed as the sum in (0.9.6.3). Explain why $A = F_n \Lambda F_n^*$, in which $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, each $\lambda_\ell = \sum_{k=0}^{n-1} a_{k+1}\omega^{k(\ell-1)}$, and the diagonal entries of Λ are the entries of the vector $n^{1/2}F_n^* A e_1$. Thus, the Fourier matrix provides an explicit unitary diagonalization for every circulant matrix. (f) Write $F_n = \mathcal{C}_n + i\mathcal{S}_n$, in which $\mathcal{C}_n$ and $\mathcal{S}_n$ are real. What are the entries of $\mathcal{C}_n$ and $\mathcal{S}_n$? Let $H_n = \mathcal{C}_n + \mathcal{S}_n$ denote the n-by-n *Hartley matrix*. (g) Show that $\mathcal{C}_n^2 + \mathcal{S}_n^2 = I$, $\mathcal{C}_n\mathcal{S}_n = \mathcal{S}_n\mathcal{C}_n = 0$, H_n is symmetric, and H_n is real orthogonal. (h) Let K_n denote the reversal matrix (0.9.5.1). Show that $\mathcal{C}_n K_n = K_n \mathcal{C}_n = \mathcal{C}_n$, $\mathcal{S}_n K_n = K_n \mathcal{S}_n = -\mathcal{S}_n$, and $H_n K_n = K_n H_n$, so $\mathcal{C}_n$, $\mathcal{S}_n$, and H_n are centrosymmetric. It is known that $H_n A H_n = \Lambda$ is diagonal for any matrix of the form $A = E + K_n F$, in which E and F are real circulant matrices, $E = E^T$, and $F = -F^T$; the diagonal entries of Λ are the entries of the vector $n^{1/2} H_n A e_1$. In particular, the Hartley matrix provides an explicit real orthogonal diagonalization for every real symmetric circulant matrix.

Further Readings and Notes. For the original proof of (2.2.6), see W. Specht, Zur Theorie der Matrizen II, *Jahresber. Deutsch. Math.-Verein.* 50 (1940) 19–23; there is a modern proof in [Kap]. For a survey of the issues addressed in (2.2.8), see D. Đoković and C. Johnson, Unitarily Achievable Zero Patterns and Traces of Words in A and A^*, *Linear Algebra Appl.* 421 (2007) 63-68.

2.3 Unitary triangularizations

Perhaps the most fundamentally useful fact of elementary matrix theory is a theorem attributed to I. Schur: any square complex matrix A is unitarily similar to a triangular matrix whose diagonal entries are the eigenvalues of A. The proof involves a sequential deflation by unitary similarity.

2.3.1 Theorem. (Schur) Let $A \in M_n$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ in any prescribed order and let x be a unit vector such that $Ax = \lambda_1 x$. Then there is a unitary $U = [x\ u_2\ \ldots\ u_n] \in M_n$ such that $U^*AU = T = [t_{ij}]$ is upper triangular with diagonal entries $t_{ii} = \lambda_i, i = 1, \ldots, n$. That is, every square complex matrix A is unitarily similar to an upper triangular matrix whose diagonal entries are the eigenvalues of A in any prescribed order. Furthermore, if $A \in M_n(\mathbf{R})$ and if all its eigenvalues are real, then U may be chosen to be real orthogonal.

Proof: Let x be a normalized eigenvector of A associated with the eigenvalue λ_1, that is, $x^*x = 1$ and $Ax = \lambda_1 x$. Let $U_1 = [x\ u_2\ \ldots\ u_n]$ be any unitary matrix whose first column is x. For example, one may take $U_1 = U(x, e_1)$ as in (2.1.13) or see Problem 1. Then

$$\begin{aligned}
U_1^* A U_1 &= U_1^* \begin{bmatrix} Ax & Au_2 & \ldots & Au_n \end{bmatrix} = U_1^* \begin{bmatrix} \lambda_1 x & Au_2 & \ldots & Au_n \end{bmatrix} \\
&= \begin{bmatrix} x^* \\ u_2^* \\ \vdots \\ u_n^* \end{bmatrix} \begin{bmatrix} \lambda_1 x & Au_2 & \ldots & Au_n \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 x^* x & x^* A u_2 & \ldots & x^* A u_n \\ \lambda_1 u_2^* x & & & \\ \vdots & & A_1 & \\ \lambda_1 u_n^* x & & & \end{bmatrix} = \begin{bmatrix} \lambda_1 & \bigstar \\ 0 & A_1 \end{bmatrix}
\end{aligned}$$

because the columns of U_1 are orthonormal. The eigenvalues of the submatrix $A_1 = [u_i^* A u_j]_{i,j=2}^n \in M_{n-1}$ are $\lambda_2, \ldots, \lambda_n$. If $n = 2$, we have achieved the desired unitary triangularization. If not, let $\xi \in \mathbf{C}^{n-1}$ be a normalized eigenvector of A_1 corresponding to λ_2, and perform the preceding reduction on A_1. If $U_2 \in M_{n-1}$ is any unitary matrix whose first column is ξ, then we have seen that

$$U_2^* A_1 U_2 = \begin{bmatrix} \lambda_2 & \bigstar \\ 0 & A_2 \end{bmatrix}$$

Let $V_2 = [1] \oplus U_2$ and compute the unitary similarity

$$(U_1 V_2)^* A U_1 V_2 = V_2^* U_1^* A U_1 V_2 = \left[\begin{array}{cc:c} \lambda_1 & * & * \\ 0 & \lambda_2 & \\ \hdashline \multicolumn{2}{c:}{\mathbf{0}} & A_2 \end{array}\right]$$

Continue this reduction to produce unitary matrices $U_i \in M_{n-i+1}, i = 1, \ldots, n-1$ and unitary matrices $V_i \in M_n, i = 2, \ldots, n-2$. The matrix

$$U = U_1 V_2 V_3 \cdots V_{n-2}$$

is unitary and U^*AU is upper triangular.

If all the eigenvalues of $A \in M_n(\mathbf{R})$ are real, then all of the eigenvectors and unitary matrices in the preceding algorithm can be chosen to be real (Problem 3 in (1.1) and (2.1.13)). □

Exercise. Follow the proof of (2.3.1) to see that *upper triangular* can be replaced by *lower triangular* in the statement of the theorem with, of course, a different unitary similarity.

Exercise. If the eigenvector x in the proof of (2.3.1) is also a left eigenvector of A, we know that $x^*A = \lambda x^*$. (1.4.7a) Explain why $U_1^*AU_1 = \left[\begin{smallmatrix} \lambda_1 & 0 \\ 0 & A_1 \end{smallmatrix}\right]$. If *every* right eigenvector of A is also a left eigenvector, explain why the upper triangular matrix T constructed in (2.3.1) is actually a diagonal matrix.

2.3.2 Example. If the eigenvalues of A are re-ordered and the corresponding upper triangularization (2.3.1) is performed, the entries of T above the main diagonal can look very different. Consider

$$T_1 = \begin{bmatrix} 1 & 1 & 4 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix},\ T_2 = \begin{bmatrix} 2 & -1 & 3\sqrt{2} \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 3 \end{bmatrix},\ U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$$

Explain why U is unitary and $T_2 = UT_1U^*$.

Exercise. If $A = [a_{ij}] \in M_n$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ and is unitarily similar to an upper triangular matrix $T = [t_{ij}] \in M_n$, the diagonal entries of T are the eigenvalues of A in some order. Apply (2.2.2) to A and T to show that

$$\sum_{i=1}^{n} |\lambda_i|^2 = \sum_{i,j=1}^{n} |a_{ij}|^2 - \sum_{i<j} |t_{ij}|^2 \le \sum_{i,j=1}^{n} |a_{ij}|^2 \tag{2.3.2a}$$

with equality if and only if T is diagonal.

Exercise. If $A = [a_{ij}]$ and $B = [b_{ij}] \in M_2$ have the same eigenvalues and if $\sum_{i,j=1}^{2} |a_{ij}|^2 = \sum_{i,j=1}^{2} |b_{ij}|^2$, use the criterion in (2.2.8) to show that A and B are unitarily similar. However, consider

$$A = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 5 \\ 0 & 0 & 3 \end{bmatrix} \tag{2.3.2b}$$

which have the same eigenvalues and the same sums of squared entries. Use the criterion in (2.2.8) or the exercise following (2.4.5.1) to show that A and B are not unitarily similar. Nevertheless, A and B are similar. Why?

It is a useful adjunct to (2.3.1) that a commuting family of matrices may be simultaneously upper triangularized.

2.3.3 Theorem. Let $\mathcal{F} \subseteq M_n$ be a commuting family. There is a unitary $U \in M_n$ such that U^*AU is upper triangular for every $A \in \mathcal{F}$.

Proof: Return to the proof of (2.3.1). Exploiting (1.3.19) at each step of the proof in which a choice of an eigenvector (and unitary matrix) is made, choose an eigenvector that is common to every $A \in \mathcal{F}$ and choose a single unitary matrix that has this common eigenvector as its first column; it deflates (via unitary similarity) every matrix in $\mathcal{F}$ in the same way. Similarity preserves commutativity, and a partitioned multiplication calculation reveals that, if two matrices of the form

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$$

commute, then A_{22} and B_{22} commute also. Thus, the commuting family property is inherited by the submatrix A_i at each reduction step in the proof of (2.3.1). We conclude that all ingredients in the U of (2.3.1) may be chosen in the same way for all members of a commuting family, thus verifying (2.3.3). $\square$

In (2.3.1) we are permitted to specify the main diagonal of T (that is, we may specify in advance the order in which the eigenvalues of A appear as the deflation progresses), but (2.3.3) makes no such claim—even for a single matrix in $\mathcal{F}$. At each stage of the deflation, the common eigenvector used is associated with *some* eigenvalue of each matrix in $\mathcal{F}$, but we may not be able to specify which one. We simply take the eigenvalues as they come, using (1.3.19).

If a real matrix A has any non-real eigenvalues, there is no hope of reducing it to upper triangular form T by a *real* similarity because the diagonal entries of T would be the eigenvalues of A. However, we can always reduce A to a real quasi-triangular form by a real similarity as well as by a real orthogonal similarity.

2.3.4 Theorem. Suppose that $A \in M_n(\mathbf{R})$ has p complex conjugate pairs of non-real eigenvalues $\lambda_1 = a_1 + ib_1, \overline{\lambda_1} = a_1 - ib_1 \ldots, \lambda_p = a_p + ib_p, \overline{\lambda_p} = a_p - ib_p$ in which all $a_j, b_j \in \mathbf{R}$ and all $b_j \neq 0$, and, if $2p < n$, an additional $n - 2p$ real eigenvalues $\mu_1, \ldots, \mu_{n-2p}$. Then there is a nonsingular

$S \in M_n(\mathbf{R})$ such that

$$S^{-1}AS = \begin{bmatrix} A_1 & & & \bigstar \\ & A_2 & & \\ & & \ddots & \\ \mathbf{0} & & & A_{n-p} \end{bmatrix} \tag{2.3.5}$$

is real and block upper triangular, and each diagonal block is either 1-by-1 or 2-by-2. There are p real diagonal blocks of the form $\left[\begin{smallmatrix} a_j & b_j \\ -b_j & a_j \end{smallmatrix}\right]$, one for each conjugate pair of non-real eigenvalues $\lambda_j, \overline{\lambda_j} = a_j \pm ib_j$. There are $n - 2p$ diagonal blocks of the form $[\mu_j]$, one for each of the real eigenvalues $\mu_1, \ldots, \mu_{n-2p}$. The p 2-by-2 diagonal blocks and the $n - 2p$ 1-by-1 diagonal blocks may appear in (2.3.5) in any prescribed order. The real similarity S may be taken to be a real orthogonal matrix Q; in this case the 2-by-2 diagonal blocks of $Q^T AQ$ have the form $R_j \left[\begin{smallmatrix} a_j & b_j \\ -b_j & a_j \end{smallmatrix}\right] R_j^{-1}$, in which each R_j is a nonsingular real upper triangular matrix.

Proof: The proof of (2.3.1) shows how to deflate A by a sequence of real orthogonal similarities corresponding to its real eigenvalues, if any. Problem 33 in (1.3) describes the deflation step corresponding to a complex conjugate pair of non-real eigenvalues; repeating this deflation p times achieves the form (2.3.5), whose 2-by-2 diagonal blocks have the asserted form (which reveals their corresponding conjugate pair of eigenvalues). It remains to consider how the 2-by-2 diagonal blocks would be modified if we were to use only real orthogonal similarities in the 2-by-2 deflations. If $\lambda = a + ib$ is a non-real eigenvalue of A with associated eigenvector $x = u + iv, u, v \in \mathbf{R}^n$, we have seen that $\{u, v\}$ is linearly independent and $A[u\ v] = [u\ v] \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$. If $\{u, v\}$ is not an orthonormal set, use the QR factorization (2.1.14) to write $[u\ v] = Q_2 R_2$, in which $Q_2 = [q_1\ q_2] \in M_{n,2}(\mathbf{R})$ has orthonormal columns and $R_2 \in M_2(\mathbf{R})$ is nonsingular and upper triangular. Then $A[u\ v] = AQ_2R_2 = Q_2R_2 \left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$, so

$$AQ_2 = Q_2R_2 \begin{bmatrix} a & b \\ -b & a \end{bmatrix} R_2^{-1}$$

If we let S be a real orthogonal matrix whose first two columns are q_1 and q_2, we obtain a deflation of the asserted form. Of course, if $\{u, v\}$ is orthonormal, the QR factorization is unnecessary and we can take S to be a real orthogonal matrix whose first two columns are u and v. $\square$

There is also a real version of (2.3.3).

2.3.6 Theorem. Let $\mathcal{F} \subseteq M_n(\mathbf{R})$ be a commuting family. There is a real orthogonal $Q \subseteq M_n(\mathbf{R})$ such that $Q^T AQ$ has the form (2.3.5) for every $A \in \mathcal{F}$.

Exercise. Modify the proof of (2.3.3) to prove (2.3.6) as follows: First deflate all members of $\mathcal{F}$ using all the common real eigenvectors. Then consider the common non-real eigenvectors and deflate two columns at a time as in the proof of (2.3.4). Notice that different members of $\mathcal{F}$ may have different numbers of 2-by-2 diagonal blocks after the common real orthogonal similarity, but if one member has a 2-by-2 block in a certain position and another member does not, then commutativity requires that the latter must have a pair of equal 1-by-1 blocks there.

Problems

1. Let $x \in \mathbf{C}^n$ be a given unit vector and write $x = [x_1 \ y^T]^T$, in which $x_1 \in \mathbf{C}$ and $y \in \mathbf{C}^{n-1}$. Choose $\theta \in \mathbf{R}$ such that $e^{i\theta}x_1 \geq 0$ and define $z = e^{i\theta}x = [z_1 \ \zeta^T]^T$, in which $z_1 \in \mathbf{R}$ is nonnegative and $\zeta \in \mathbf{C}^{n-1}$. Consider the Hermitian matrix

$$V_x = \left[\begin{array}{c|c} z_1 & \zeta^* \\ \hline \zeta & -I + \frac{1}{1+z_1}\zeta\zeta^* \end{array}\right]$$

Use partitioned multiplication to compute $V_x^* V_x = V_x^2$. Conclude that $U = e^{-i\theta}V_x = [x \ u_2 \ \ldots \ u_n]$ is a unitary matrix whose first column is the given vector x.

2. If $x \in \mathbf{R}^n$ is a given unit vector, show how to streamline the construction described in Problem 1 to produce a real orthogonal matrix $Q \in M_n(\mathbf{R})$ whose first column is x. Prove that your construction works.

3. Let $A \in M_n(\mathbf{R})$. Explain why the nonreal eigenvalues of A (if any) must occur in conjugate pairs.

4 Consider the family $\mathcal{F} = \left\{ \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \right\}$ and show that the hypothesis of commutativity in (2.3.3), while sufficient to imply simultaneous unitary upper triangularizability of $\mathcal{F}$, is not necessary.

5. Let $\mathcal{F} = \{A_1, \ldots, A_k\} \subset M_n$ be a given family, and let $\mathcal{G} = \{A_iA_j : i, j = 1, 2, \ldots, k\}$ be the family of all pair-wise products of matrices in $\mathcal{F}$. If $\mathcal{G}$ is commutative, it is known that $\mathcal{F}$ can be simultaneously unitarily upper triangularized if and only if every eigenvalue of every commutator $A_iA_j - A_jA_i$ is zero. Show that assuming commutativity of $\mathcal{G}$ is a weaker hypothesis than assuming commutativity of $\mathcal{F}$. Show that the family $\mathcal{F}$ in Problem 4 has a corresponding $\mathcal{G}$ that is commutative, and that it also satisfies the zero eigenvalue condition.

6. Let $A, B \in M_n$ be given, and suppose A and B are simultaneously similar to upper triangular matrices; that is, $S^{-1}AS$ and $S^{-1}BS$ are both upper triangular for some nonsingular $S \in M_n$. Show that every eigenvalue of $AB - BA$ must be zero. *Hint:* If $\Delta_1, \Delta_2 \in M_n$ are both upper triangular, what is the main diagonal of $\Delta_1\Delta_2 - \Delta_2\Delta_1$?

7. If a given $A \in M_n$ can be written as $A = Q\Delta Q^T$, in which $Q \in M_n$ is complex orthogonal and $\Delta \in M_n$ is upper triangular, show that A has at least one eigenvector $x \in \mathbf{C}^n$ such that $x^T x \neq 0$. Consider $A = \begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$ to show that not every $A \in M_n$ can be upper triangularized by a complex orthogonal similarity.

8. Let $Q \in M_n$ be complex orthogonal, and suppose $x \in \mathbf{C}^n$ is an eigenvector of Q associated with an eigenvalue $\lambda \neq \pm 1$. Show that $x^T x = 0$. See Problem 8(a) in (2.1) for an example of a family of 2-by-2 complex orthogonal matrices with both eigenvalues different from ± 1. Show that none of these matrices can be reduced to upper triangular form by orthogonal similarity. *Hint:* $Qx = \lambda x$ $\Rightarrow (Qx)^T Qx = \lambda^2 x^T x$.

9. Let $\lambda, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $A \in M_n$, suppose x is a nonzero vector such that $Ax = \lambda x$, and let $y \in \mathbf{C}^n$ and $\alpha \in \mathbf{C}$ be given. Provide details for the following argument to show that the eigenvalues of the bordered matrix $\mathcal{A} = \begin{bmatrix} \alpha & y^* \\ x & A \end{bmatrix} \in M_{n+1}$ are the two eigenvalues of $\begin{bmatrix} \alpha & y^*x \\ 1 & \lambda \end{bmatrix}$ together with $\lambda_2, \ldots, \lambda_n$. Form a unitary U whose first column is $x/\|x\|_2$, let $V = [1] \oplus U$, and show that $V^*\mathcal{A}V = \begin{bmatrix} B & \bigstar \\ 0 & C \end{bmatrix}$, in which $B = \begin{bmatrix} \alpha & y^*x/\|x\|_2 \\ \|x\|_2 & \lambda \end{bmatrix} \in M_2$ and $C \in M_{n-2}$ has eigenvalues $\lambda_2, \ldots, \lambda_n$. Consider a similarity of B via $\operatorname{diag}(1, \|x\|_2^{-1})$. If $y \perp x$, conclude that the eigenvalues of $\mathcal{A}$ are $\alpha, \lambda, \lambda_2, \ldots, \lambda_n$. Explain why the eigenvalues of $\begin{bmatrix} \alpha & y^* \\ x & A \end{bmatrix}$ and $\begin{bmatrix} A & x \\ y^* & \alpha \end{bmatrix}$ are the same.

10. Let $A = [a_{ij}] \in M_n$ and let $c = \max\{|a_{ij}| : 1 \leq i, j \leq n\}$. Show that $|\det A| \leq c^n n^{n/2}$ as follows: Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A. Use the arithmetic-geometric mean inequality and (2.3.2a) to explain why $|\det A|^2 = |\lambda_1 \cdots \lambda_n|^2 \leq ((|\lambda_1|^2 + \cdots + |\lambda_n|^2)/n)^n \leq (\sum_{i,j=1}^n |a_{ij}|^2/n)^n \leq (nc^2)^n$.

11. Use (2.3.1) to prove that if all the eigenvalues of $A \in M_n$ are zero, then $A^n = 0$. *Hint*: If $T \in M_n$ is strictly upper triangular, what does T^2 look like? T^3? T^{n-1}? T^n?

Further Reading. See (3.4.3.1) for a refinement of the upper triangularization (2.3.1). For a proof of the stronger form of (2.3.3) asserted in Problem 5, see Y. P. Hong and R. A. Horn, On Simultaneous Reduction of Families of Matrices to Triangular or Diagonal Form by Unitary Congruences, *Linear and Multilinear Algebra* 17 (1985) 271–288.

2.4 Some consequences of Schur's triangularization theorem

A bounty of results can be harvested from Schur's unitary triangularization theorem.

2.4.1 The trace and determinant Suppose $A \in M_n$ has eigenvalues $\lambda_1, \ldots, \lambda_n$. In (1.2) we used the characteristic polynomial to show that $\sum_{i=1}^n \lambda_i = \operatorname{tr} A$, $\sum_{i=1}^n \Pi_{j\neq i}^n \lambda_j = \operatorname{tr}(\operatorname{adj} A)$, and $\det A = \Pi_{i=1}^n \lambda_i$, but these identities and others follow simply from inspection of the triangular form in (2.3.1).

For any nonsingular $S \in M_n$ we have $\operatorname{tr}(S^{-1}AS) = \operatorname{tr}(ASS^{-1}) = \operatorname{tr} A$; $\operatorname{tr}(\operatorname{adj}(S^{-1}AS)) = \operatorname{tr}((\operatorname{adj} S)(\operatorname{adj} A)(\operatorname{adj} S^{-1})) = \operatorname{tr}((\operatorname{adj} S)(\operatorname{adj} A)(\operatorname{adj} S)^{-1}) = \operatorname{tr}(\operatorname{adj} A)$; and $\det(S^{-1}AS) = (\det S^{-1})(\det A)(\det S) = (\det S)^{-1}(\det A)(\det S) = \det A$. Thus, $\operatorname{tr} A$, $\operatorname{tr}(\operatorname{adj} A)$, and $\det A$ can be evaluated using any matrix that is similar to A. The upper triangular matrix $T = [t_{ij}]$ in Schur's theorem (2.3.1) is convenient for this purpose, since its main diagonal entries $t_{11}, \ldots, t_{nn}$ are the eigenvalues of A, $\operatorname{tr} T = \sum_{i=1}^n t_{ii}$, $\det T = \Pi_{i=1}^n t_{ii}$, and the main diagonal entries of $\operatorname{adj} T$ are $\Pi_{j\neq 1}^n t_{jj}, \ldots, \Pi_{j\neq n}^n t_{jj}$.

2.4.2 The eigenvalues of a polynomial in *A* Suppose $A \in M_n$ has eigenvalues $\lambda_1, \ldots, \lambda_n$ and let $p(t)$ be a given polynomial. We showed in (1.1.6) that $p(\lambda_i)$ is an eigenvalue of $p(A)$ for *each* $i = 1, \ldots, n$ and that if μ is an eigenvalue of $p(A)$, then there is *some* $i \in \{1, \ldots, n\}$ such that $\mu = p(\lambda_i)$. These observations identify the the distinct eigenvalues of $p(A)$ (that is, its *spectrum* (1.1.4)), but not their multiplicities. Schur's theorem (2.3.1) reveals the multiplicities.

Let $A = UTU^*$, in which U is unitary and $T = [t_{ij}]$ is upper triangular with main diagonal entries $t_{11} = \lambda_1, t_{22} = \lambda_2, \ldots, t_{nn} = \lambda_n$. Then $p(A) = p(UTU^*) = Up(T)U^*$ (Problem 2 in (1.3)). The main diagonal entries of $p(T)$ are $p(\lambda_1), p(\lambda_2), \ldots, p(\lambda_n)$, so these are the eigenvalues (including multiplicities) of $p(T)$ and hence also of $p(A)$. In particular, for each $k = 1, 2, \ldots$ the eigenvalues of A^k are $\lambda_1^k, \ldots, \lambda_n^k$ and

$$\operatorname{tr} A^k = \lambda_1^k + \cdots + \lambda_n^k \tag{2.4.2.1}$$

Exercise. If $T \in M_n$ is strictly upper triangular, show that: all of the entries in the main diagonal and the first $p-1$ superdiagonals of T^p are zero, $p = 1, \ldots, n$; in particular, $T^n = 0$.

Suppose that $A \in M_n$. We know (Problem 6 in (1.1)) that if $A^k = 0$ for some positive integer k, then $\sigma(A) = \{0\}$, so the characteristic polynomial of A is $p_A(t) = t^n$ We can now prove the converse, and a little more. If $\sigma(A) = \{0\}$, then there is a unitary U and a *strictly* upper triangular T such that $A = UTU^*$; the preceding exercise tells us that $T^n = 0$, so $A^n = UT^nU^* = 0$. Thus, the following are equivalent for $A \in M_n$: (a) A is nilpotent; (b) $A^n = 0$; and (c) $\sigma(A) = \{0\}$.

2.4.3 The Cayley-Hamilton theorem The fact that every matrix satisfies its own characteristic equation follows from Schur's theorem and a simple observation about multiplication of triangular matrices with special patterns of zero entries.

2.4.3.1 Lemma. Suppose that $R = [r_{ij}], T = [t_{ij}] \in M_n$ are upper triangular and that $r_{ij} = 0$, $1 \le i, j \le k < n$, and $t_{k+1,k+1} = 0$. Let $S = [s_{ij}] = RT$. Then $s_{ij} = 0, 1 \le i, j \le k+1$.

Proof: The hypotheses describe block matrices R and T of the form

$$R = \begin{bmatrix} 0_k & R_{12} \\ 0 & R_{22} \end{bmatrix}, \qquad T = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}, \quad T_{11} \in M_k$$

in which R_{22}, T_{11}, and T_{22} are upper triangular and the first column of T_{22} is zero. The product RT is necessarily upper triangular. We must show that it has a zero upper-left principal submatrix of size $k+1$. Partition $T_{22} = [0\ Z]$ to reveal its first column and perform a block multiplication

$$RT = \begin{bmatrix} 0_k T_{11} + R_{12} 0 & 0_k T_{12} + R_{12}[0\ Z] \\ 0 T_{11} + R_{22} 0 & 0 T_{12} + R_{22}[0\ Z] \end{bmatrix} = \begin{bmatrix} 0_k & [0\ R_{12} Z] \\ 0 & [0\ R_{22} Z] \end{bmatrix}$$

which reveals the desired zero upper-left principal submatrix of size $k+1$. □

2.4.3.2 Theorem. (Cayley–Hamilton). Let $p_A(t)$ be the characteristic polynomial of $A \in M_n$. Then $p_A(A) = 0$

Proof: Factor $p_A(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$ as in (1.2.6) and use (2.3.1) to write A as $A = UTU^*$, in which U is unitary, T is upper triangular, and the main diagonal entries of T are $\lambda_1, \ldots, \lambda_n$. Compute

$$\begin{aligned} p_A(A) &= p_A(UTU^*) = Up_A(T)U^* \\ &= U[(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_n I)]U^* \end{aligned}$$

It suffices to show that $p_A(T) = 0$. The upper left 1-by-1 block of $T - \lambda_1 I$ is 0, and the $2, 2$ entry of $T - \lambda_2 I$ is 0, so the preceding lemma ensures that the upper left 2-by-2 principal submatrix of $(T - \lambda_1 I)(T - \lambda_2 I)$ is 0. Suppose the upper left *k*-by-*k* principal submatrix of $(T - \lambda_1 I) \cdots (T - \lambda_k I)$ is zero. The $k+1, k+1$ entry of $(T - \lambda_{k+1} I)$ is 0, so invoking the lemma again, we know that the upper left principal submatrix of $(T - \lambda_1 I) \cdots (T - \lambda_{k+1} I)$ of size $k+1$ is 0. By induction, we conclude that $((T - \lambda_1 I) \cdots (T - \lambda_{n-1} I))(T - \lambda_n I) = 0$. □

Exercise. What is wrong with the following argument? "Since $p_A(\lambda_i) = 0$ for every eigenvalue λ_i of $A \in M_n$, and since the eigenvalues of $p_A(A)$ are $p_A(\lambda_1), \ldots, p_A(\lambda_n)$, all eigenvalues of $p_A(A)$ are 0. Therefore, $p_A(A) = 0$." Give an explicit example to illustrate the fallacy in the argument.

Exercise. What is wrong with the following argument? "Since $p_A(t) = \det(tI - A)$, we have $p_A(A) = \det(AI - A) = \det(A - A) = \det 0 = 0$. Therefore, $p_A(A) = 0$."

The Cayley–Hamilton theorem is often paraphrased as "every square matrix satisfies its own characteristic equation," (1.2.3) but this must be understood carefully: The scalar polynomial $p_A(t)$ is first computed as $p_A(t) = \det(tI - A)$, and one then forms the matrix $p_A(A)$ from the characteristic polynomial.

We have proved the Cayley–Hamilton theorem for matrices with complex entries, and hence it must hold for matrices whose entries come from any subfield of the complex numbers (the reals or the rationals, for example). In fact, the Cayley–Hamilton theorem is a completely formal result that holds for matrices whose entries come from any field or, more generally, any commutative ring. See Problem 3.

One important use of the Cayley–Hamilton theorem is to write powers A^k of $A \in M_n$, for $k \geq n$, as linear combinations of $I, A, A^2, \ldots, A^{n-1}$.

2.4.3.3 Example. Let $A = \begin{bmatrix} 3 & 1 \\ -2 & 0 \end{bmatrix}$. Then $p_A(t) = t^2 - 3t + 2$, so $A^2 - 3A + 2I = 0$. Thus, $A^2 = 3A - 2I$; $A^3 = A(A^2) = 3A^2 - 2A = 3(3A - 2I) - 2A = 7A - 6I$; $A^4 = 7A^2 - 6A = 15A - 14I$, and so on. We can also express negative powers of the nonsingular matrix A as linear combinations of A and I. Write $A^2 - 3A + 2I = 0$ as $2I = -A^2 + 3A = A(-A + 3I)$, or $I = A[\frac{1}{2}(-A + 3I)]$. Thus, $A^{-1} = -\frac{1}{2}A + \frac{3}{2}I = \begin{bmatrix} 0 & -1/2 \\ 1 & 3/2 \end{bmatrix}$, $A^{-2} =$

$(-\frac{1}{2}A + \frac{3}{2}I)^2 = \frac{1}{4}A^2 - \frac{3}{2}A + \frac{9}{4}I = \frac{1}{4}(3A - 2I) - \frac{3}{2}A + \frac{9}{4}I = -\frac{3}{4}A + \frac{7}{4}I$, and so on.

2.4.3.4 Corollary. Suppose $A \in M_n$ is nonsingular and let $p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$. Let $q(t) = \frac{-1}{a_0}(t^{n-1} + a_{n-1}t^{n-2} + \cdots + a_2 t + a_1)$. Then $A^{-1} = q(A)$ is a polynomial in A.

Proof: Write $p_A(A) = 0$ as $A(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2 A + a_1 I) = -a_0 I$, that is, $Aq(A) = I$. □

Exercise. If $A, B \in M_n$ are similar and $g(t)$ is any given polynomial, show that $g(A)$ is similar to $g(B)$, and that any polynomial equation satisfied by A is satisfied by B. Give some thought to the converse: Satisfaction of the same polynomial equations implies similarity—true or false?

2.4.3.5 Example. We have shown that each $A \in M_n$ satisfies a polynomial equation of degree n, for example, its characteristic equation. It is possible for $A \in M_n$ to satisfy a polynomial equation of degree less than n, however. Consider

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \in M_3$$

The characteristic polynomial is $p_A(t) = (t-1)^3$ and indeed $(A - I)^3 = 0$. But $(A - I)^2 = 0$ so A satisfies a polynomial equation of degree 2. There is no polynomial $h(t) = t + a_0$ of degree 1 such that $h(A) = 0$ since $h(A) = A + a_0 I \neq 0$ for all $a_0 \in \mathbf{C}$.

Exercise. Suppose that a diagonalizable matrix $A \in M_n$ has $d \leq n$ distinct eigenvalues $\lambda_1, \ldots, \lambda_d$. Let $q(t) = (t - \lambda_1) \cdots (t - \lambda_d)$. Show that $q(A) = 0$, so A satisfies a polynomial equation of degree d. Why is there no polynomial $g(t)$ of degree strictly less than d such that $g(A) = 0$? Consider the matrix in the preceding example to show that the minimum degree of a polynomial equation satisfied by a nondiagonalizable matrix can be strictly larger than the number of its distinct eigenvalues.

2.4.4 Sylvester's theorem on linear matrix equations The equation $AX - XA = 0$ associated with commutativity is a special case of the linear matrix equation $AX - XB = C$, often called *Sylvester's equation.* The following theorem gives a necessary and sufficient condition for Sylvester's equation to have a unique solution X for every given C. It relies on the Cayley-Hamilton theorem, and on the observation that if $AX = XB$, then $A^2X = A(AX) =$

$A(XB) = (AX)B = (XB)B = XB^2$, $A^3X = A(A^2X) = A(XB^2) = (AX)B^2 = XB^3$, etc. Thus, with the usual understanding that A^0 denotes the identity matrix, we have

$$\left(\sum_{k=0}^{m} a_k A^k\right) X = \sum_{k=0}^{m} a_k A^k X = \sum_{k=0}^{m} a_k X B^k = X \left(\sum_{k=0}^{m} a_k B^k\right)$$

that is, $AX - XB = 0$ implies that $g(A)X - Xg(B) = 0$ for any polynomial $g(t)$.

2.4.4.1 Theorem. (Sylvester) Let $A \in M_n$ and $B \in M_m$ be given. The equation $AX - XB = C$ has a unique solution $X \in M_{n,m}$ for each given $C \in M_{n,m}$ if and only if $\sigma(A) \cap \sigma(B) = \varnothing$, that is, if and only if A and B have no eigenvalue in common. In particular, if $\sigma(A) \cap \sigma(B) = \varnothing$ then the only X such that $AX - XB = 0$ is $X = 0$. If A and B are real, then $AX - XB = C$ has a unique solution $X \in M_{n,m}(\mathbf{R})$ for each given $C \in M_{n,m}(\mathbf{R})$.

Proof: Consider the linear transformation $T : M_{n,m} \to M_{n,m}$ defined by $T(X) = AX - XB$. In order to ensure that the equation $T(X) = C$ has a unique solution X for every given $C \in M_{n,m}$ it suffices to show that the only solution of $T(X) = 0$ is $X = 0$. (0.5) If $AX - XB = 0$, we know from the preceding discussion that $p_B(A)X - Xp_B(B) = 0$. The Cayley-Hamilton theorem ensures that $p_B(B) = 0$, so $p_B(A)X = 0$.

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of B, so $p_B(t) = (t - \lambda_1) \cdots (t - \lambda_n)$ and $p_B(A) = (A - \lambda_1 I) \cdots (A - \lambda_n I)$. If $\sigma(A) \cap \sigma(B) = \varnothing$ then each factor $A - \lambda_j I$ is nonsingular, $p_B(A)$ is nonsingular, and the only solution of $p_B(A)X = 0$ is $X = 0$. Conversely, if $p_B(A)X = 0$ has a nontrivial solution, then $p_B(A)$ must be nonsingular, some factor $A - \lambda_j I$ is singular, and some λ_j is an eigenvalue of A.

If A and B are real, consider the linear transformation $T : M_{n,m}(\mathbf{R}) \to M_{n,m}(\mathbf{R})$ defined by $T(X) = AX - XB$. The same argument shows that the real matrix $p_B(A)$ is nonsingular if and only if $\sigma(A) \cap \sigma(B) = \varnothing$ (even if some of the eigenvalues λ_i of B are not real). $\square$

Sylvester's theorem is often used in the following form.

2.4.4.2 Corollary. Suppose that $B, C \in M_n$ are block diagonal and conformally partitioned as $A = A_1 \oplus \cdots \oplus A_k$ and $B = B_1 \oplus \cdots \oplus B_k$, and suppose that $\sigma(B_i) \cap \sigma(C_j) = \varnothing$ whenever $i \neq j$. If $A \in M_n$ and $AB = CA$, then $A = A_1 \oplus \cdots \oplus A_k$ is block diagonal conformal to B and C, and $A_i B_i = C_i A_i$ for each $i = 1, \ldots, k$.

Proof: Partition $A = [A_{ij}]$ conformally to B and C. Then $AB = CA$ if and only if $A_{ij}B_j = C_iA_{ij}$. If $i \neq j$, then (2.4.4.1) ensures that $A_{ij} = 0$. □

2.4.5 Uniqueness in Schur's triangularization theorem For a given $A \in M_n$, the upper triangular form T described in (2.3.1) that can be achieved by unitary similarity need not be unique. That is, different upper triangular matrices with the same main diagonals can be unitarily similar to A. However, if A has d distinct eigenvalues $\lambda_1, \ldots, \lambda_d$ in a prescribed order with respective multiplicities $n_1, \ldots, n_d$, and if we let $\Lambda = \lambda_1 I_{n_1} \oplus \cdots \oplus \lambda_d I_{n_d}$, then any two upper triangular matrices T and T' that are unitarily similar to A and have the same main diagonal as Λ are related in a simple way: there is a block diagonal unitary matrix $W = W_1 \oplus \cdots \oplus W_d$ with each $W_i \in M_{n_i}$ such that $T' = WTW^*$.

2.4.5.1 Theorem. Suppose that $d, n_1, \ldots, n_d$ are given positive integers with $d \geq 2$, let $n_1 + \cdots + n_d = n$, and let $\Lambda = \lambda_1 I_{n_1} \oplus \cdots \oplus \lambda_d I_{n_d}$. Let $T = [T_{ij}]_{i,j=1}^d$ and $T' = [T'_{ij}]_{i,j=1}^d$ be upper triangular matrices that are partitioned conformally to Λ and have the same main diagonal as Λ. Let a given $W = [W_{ij}]_{i,j=1}^d \in M_n$ be partitioned conformally to Λ and suppose that $TW = WT'$. Then $W_{ij} = 0$ if $i > j$, that is, W is block upper triangular conformal to Λ. If W is unitary, then $T' = WTW^*$, $W = W_{11} \oplus \cdots \oplus W_{dd}$ is block diagonal, and each block W_{ii} is unitary. Conversely, if $W_i \in M_{n_i}$ is unitary for each $i = 1, \ldots, d$ and $T_{ij} = W_i T'_{ij} W_j^*$ for all $i, j = 1, \ldots, d$, $i \leq j$, then T is unitarily similar to T' (via $W = W_1 \oplus \cdots \oplus W_d$).

Proof: Begin with the identity $WT = T'W$. Because T and T' are block upper triangular, the $d, 1$ block of $T'W$ is $T'_{dd}W_{d1}$ and the $d, 1$ block of WT is $W_{d1}T_{11}$. Since $\sigma(T'_{dd}) = \{\lambda_d\} \neq \{\lambda_1\} = \sigma(T_{11})$, (2.4.4.1) ensures that $W_{d1} = 0$ is the only solution to $T'_{dd}W_{d1} = W_{d1}T_{11}$. If $d = 2$, we stop at this point. If $d > 2$, then the $d, 2$ block of $T'W$ is $T'_{dd}W_{d2}$, the $d, 2$ block of WT is $W_{d1}T_{12} + W_{d2}T_{22} = W_{d2}T_{22}$, and $T'_{dd}W_{d2} = W_{d2}T_{22}$. Again, (2.4.4.1) ensures that $W_{d2} = 0$ since $\sigma(T'_{dd}) = \{\lambda_d\} \neq \{\lambda_2\} = \sigma(T_{22})$. Proceeding in this way across the d^{th} block row of $T'W = WT$, we see that all of the blocks $W_{d1}, \ldots, W_{d,d-1}$ are zero. Now equate the $(d-1), k$ blocks of $T'W = WT$ in turn for $k = 1, \ldots, d-1$ and conclude in the same way that they are all zero. Working our way up the block rows of $T'W = WT$, we conclude that $W_{ij} = 0$ for all $i > j$, which means that W is block upper triangular conformal to Λ.

Now assume that W is unitary. Partition $W = \begin{bmatrix} W_{11} & X \\ 0 & \hat{W} \end{bmatrix}$ and conclude from (2.1.10) that W_{11} and $\hat{W}$ are unitary, and $X = 0$. Since $\hat{W}$ is also block upper triangular and unitary, an induction leads to the conclusion that $W = W_{11} \oplus \cdots \oplus W_{dd}$.

The converse assertion follows from a computation. $\square$

Exercise. Suppose $T = [t_{ij}], T' = [t'_{ij}] \in M_n$ are upper triangular and have the same main diagonals, and suppose the main diagonal entries are distinct. Explain why T and T' are unitarily similar if and only if there are $\theta_1, \ldots, \theta_n \in \mathbf{R}$ such that $t_{jk} = e^{i(\theta_j - \theta_k)} t'_{jk}$ whenever $k > j$. Describe the set of all matrices that have the same main diagonal as the matrix T_1 in Example 2.3.2 and are unitarily similar to it.

2.4.6 Every square matrix is block diagonalizable The following application and extension of Schur's theorem (2.3.1) is an important step toward the Jordan canonical form, which we discuss in the next chapter.

2.4.6.1 Theorem. Let the distinct eigenvalues of $A \in M_n$ be $\lambda_1, \ldots, \lambda_d$, with respective multiplicities $n_1, \ldots, n_d$. Theorem (2.3.1) ensures that A is unitarily similar to a d-by-d block upper triangular matrix $T = [T_{ij}]_{i,j=1}^d$ in which each block T_{ij} is n_i-by-n_j, $T_{ij} = 0$ if $i > j$, and each diagonal block T_{ii} is upper triangular with diagonal entries λ_i, that is, each $T_{ii} = \lambda_i I_{n_i} + R_i$ and $R_i \in M_{n_i}$ is strictly upper triangular. Then A is similar to

$$\begin{bmatrix} T_{11} & & & \mathbf{0} \\ & T_{22} & & \\ & & \ddots & \\ \mathbf{0} & & & T_{dd} \end{bmatrix} \tag{2.4.6.2}$$

If $A \in M_n(\mathbf{R})$ and if all its eigenvalues are real, then the unitary similarity that reduces A to the special upper triangular form T and the similarity matrix that reduces T to the block diagonal form (2.4.6.2) may both be taken to be real.

Proof: Partition T as

$$T = \begin{bmatrix} T_{11} & Y \\ 0 & S_2 \end{bmatrix}$$

in which $S_2 = [T_{ij}]_{i,j=2}^d$. Notice that the only eigenvalue of T_{11} is λ_1 and that the eigenvalues of S_2 are $\lambda_2, \ldots, \lambda_d$. Sylvester's theorem (2.4.4.1) ensures that the equation $T_{11}X - XS = -Y$ has a solution X; use it to construct

$$M = \begin{bmatrix} I_{n_1} & X \\ 0 & I \end{bmatrix} \text{ and its inverse } M^{-1} = \begin{bmatrix} I_{n_1} & -X \\ 0 & I \end{bmatrix}$$

Then

$$
\begin{aligned}
M^{-1}TM &= \begin{bmatrix} I_{n_1} & -X \\ 0 & I \end{bmatrix} \begin{bmatrix} T_{11} & Y \\ 0 & S_2 \end{bmatrix} \begin{bmatrix} I_{n_1} & X \\ 0 & I \end{bmatrix} \\
&= \begin{bmatrix} T_{11} & T_{11}X - XS + Y \\ 0 & S_2 \end{bmatrix} = \begin{bmatrix} T_{11} & 0 \\ 0 & S_2 \end{bmatrix}
\end{aligned}
$$

If $d = 2$, this is the desired block diagonalization. If $d > 2$, repeat this reduction process to show that S_2 is similar to $T_{22} \oplus S_3$ in which $S_3 = [T_{ij}]_{i,j=3}^{d}$. After $d-1$ reductions, we find that T is similar to $T_{11} \oplus \cdots \oplus T_{dd}$.

If A is real and has real eigenvalues, then it is real orthogonally similar to a real block upper triangular matrix of the form just considered. Each of the reduction steps can be carried out with a real similarity. □

Exercise. Suppose that $A \in M_n$ is unitarily similar to a d-by-d block upper triangular matrix $T = [T_{ij}]_{i,j=1}^{d}$. If any block T_{ij} with $j > i$ is nonzero, use (2.2.2) to explain why T is not *unitarily* similar to $T_{11} \oplus \cdots \oplus T_{dd}$.

There are two extensions of the preceding theorem that, for commuting families and simultaneous (but not necessarily unitary) similarity, significantly refine the block structure achieved in (2.3.3).

2.4.6.3 Theorem. Let $\mathcal{F} \subset M_n$ be a commuting family, let A_0 be any given matrix in $\mathcal{F}$, and suppose that A_0 has d distinct eigenvalues $\lambda_1, \ldots, \lambda_d$, with respective multiplicities $n_1, \ldots, n_d$. Then there is a nonsingular $S \in M_n$ such that (a) $\hat{A}_0 = S^{-1}A_0S = T_1 \oplus \cdots \oplus T_d$, in which each $T_i \in M_{n_i}$ is upper triangular and all its diagonal entries are λ_i; and (b) for every $A \in \mathcal{F}$, $S^{-1}AS$ is upper triangular and block diagonal conformal to $\hat{A}_0$.

Proof: First use (2.4.6.1) to choose a nonsingular S_0 such that $S_0^{-1}A_0S_0 = R_1 \oplus \cdots \oplus R_d = \tilde{A}_0$, in which each $R_i \in M_{n_i}$ has λ_i as its only eigenvalue. Let $S_0^{-1}\mathcal{F}S_0 = \{S_0^{-1}AS_0 : A \in \mathcal{F}\}$, which is also a commuting family. Partition any given $B \in S_0^{-1}\mathcal{F}S_0$ as $B = [B_{ij}]_{i,j=1}^{d}$, conformal to $\tilde{A}_0$. Then $[R_iB_{ij}] = \tilde{A}_0B = B\tilde{A}_0 = [B_{ij}R_j]$, so $R_iB_{ij} = B_{ij}R_j$ for all $i, j = 1, \ldots, d$. Sylvester's theorem (2.4.4.1) now ensures that $B_{ij} = 0$ for all $i \neq j$ since R_i and R_j have no eigenvalues in common. Thus, $S_0^{-1}\mathcal{F}S_0$ is a commuting family of block diagonal matrices that are all conformal to $\tilde{A}_0$. For each $i = 1, \ldots, d$, consider the family $\mathcal{F}_i \subset M_{n_i}$ consisting of the i^{th} diagonal block of every matrix in $S_0^{-1}\mathcal{F}S_0$; notice that $R_i \in \mathcal{F}_i$ for each $i = 1, \ldots, d$. Each $\mathcal{F}_i$ is a commuting family, so (2.3.3) ensures that there is a unitary $U_i \in M_{n_i}$ such that $U_i^*\mathcal{F}_iU_i$ is an upper triangular family. The main diagonal entries of $U_i^*R_iU_i$ are its eigenvalues, which are all equal to λ_i. Let

$U = U_1 \oplus \cdots \oplus U_d$ and observe that $S = S_0 U$ accomplishes the asserted reduction, in which $T_i = U_i^* R_i U_i$. $\square$

2.4.6.4 Corollary. Let $\mathcal{F} \subset M_n$ be a commuting family. There is a nonsingular $S \in M_n$ such that, for every $A \in \mathcal{F}$, $S^{-1}AS$ is upper triangular and block diagonal, and each diagonal block has exactly one eigenvalue.

Proof: If every matrix in $\mathcal{F}$ has only one eigenvalue, apply (2.3.3) and stop. If some matrix in $\mathcal{F}$ has at least two distinct eigenvalues, let $A_0 \in \mathcal{F}$ be any matrix that has the maximum number of distinct eigenvalues among all matrices in $\mathcal{F}$. Construct a simultaneous block diagonal upper triangularization as in the preceding theorem, and observe that the size of every diagonal block obtained is strictly smaller than the size of A_0. Associated with each diagonal block of the reduced form of A_0 is a commuting family. Among the members of that family, either (a) each matrix has only one eigenvalue (no further reduction required), or (b) some matrix has at least two distinct eigenvalues, in which case we choose any matrix that has the maximum number of distinct eigenvalues and reduce again to obtain a set of strictly smaller diagonal blocks. Recursively repeat this reduction, which must terminate in finitely many steps, until no member of any commuting family has more than one eigenvalue. $\square$

2.4.7 Every square matrix is almost diagonalizable Another use of Schur's result is to make it clear that every matrix is "almost" diagonalizable in two possible interpretations of the phrase. The first says that arbitrarily close to a given matrix there is a diagonalizable matrix, and the second says that any given matrix is similar to an upper triangular matrix whose off-diagonal entries are arbitrarily small.

2.4.7.1 Theorem. Let $A = [a_{ij}] \in M_n$. For each $\epsilon > 0$, there exists a matrix $A(\epsilon) = [a_{ij}(\epsilon)] \in M_n$ that has n distinct eigenvalues (and is therefore diagonalizable) and is such that $\sum_{i,j=1}^{n} |a_{ij} - a_{ij}(\epsilon)|^2 < \epsilon$.

Proof: Let $U \in M_n$ be unitary and such that $U^*AU = T$ is upper triangular. Let $E = \operatorname{diag}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$, in which $\varepsilon_1, \ldots, \varepsilon_n$ are chosen so that $|\varepsilon_i| < \left(\frac{\epsilon}{n}\right)^{1/2}$and so that $t_{ii} + \varepsilon_i \neq t_{jj} + \varepsilon_j$ for all $i \neq j$. (Reflect for a moment to see that this can be done.) Then $T + E$ has n distinct eigenvalues: $t_{11} + \varepsilon_1, \ldots, t_{nn} + \varepsilon_n$, and so does $A + UEU^*$, which is similar to $T + E$. Let $A(\epsilon) = A + UEU^*$, so that $A - A(\epsilon) = -UEU^*$ and hence (2.2.2) ensures that $\sum_{i,j} |a_{ij} - a_{ij}(\epsilon)|^2 = \sum_{i=1}^{n} |\varepsilon_i|^2 < n\left(\frac{\epsilon}{n}\right) = \epsilon$. $\square$

Exercise. Show that the condition $\sum_{i,j} |a_{ij} - a_{ij}(\epsilon)|^2 < \epsilon$ in (2.4.6) could be replaced by $\max_{i,j} |a_{ij} - a_{ij}(\epsilon)| < \epsilon$. *Hint:* Apply the theorem with ϵ^2 in place of ϵ and realize that, if a sum of squares is less than ϵ^2, each of the items must be less than ϵ in absolute value.

2.4.7.2 Theorem. Let $A \in M_n$. For each $\epsilon > 0$ there is a nonsingular matrix $S_\epsilon \in M_n$ such that $S_\epsilon^{-1} A S_\epsilon = T_\epsilon = [t_{ij}(\epsilon)]$ is upper triangular and $|t_{ij}(\epsilon)| \leq \epsilon$ for $1 \leq i < j \leq n$.

Proof: First apply Schur's theorem to produce a unitary matrix $U \in M_n$ and an upper triangular matrix $T \in M_n$ such that $U^*AU = T$. Define D_α = diag$(1, \alpha, \alpha^2, \ldots, \alpha^{n-1})$ for a nonzero scalar α and set $t = \max_{i<j} |t_{ij}|$. Assume that $\epsilon < 1$, since it certainly suffices to prove the statement in this case. If $t \leq 1$, let $S_\epsilon = UD_\epsilon$, and, if $t > 1$, let $S_\epsilon = UD_{1/t}D_\epsilon$. In either case, the appropriate S_ϵ substantiates the claim of the theorem. If $t \leq 1$, for example, a calculation reveals that $t_{ij}(\epsilon) = t_{ij}\epsilon^{-i}\epsilon^{j} = t_{ij}\epsilon^{j-i}$, whose absolute value is no more than ϵ^{j-i}, which is, in turn, no more than ϵ if $i < j$. If $t > 1$, on the other hand, the similarity by $D_{1/t}$ simply preprocesses the matrix, producing one in which all off-diagonal entries are no more than 1 in absolute value. □

Exercise. Prove the following variant of (2.4.7.2): If $A \in M_n$ and $\epsilon > 0$, there is a nonsingular $S_\epsilon \in M_n$ such that $S_\epsilon^{-1} A S_\epsilon = T_\epsilon = [t_{ij}(\epsilon)]$ is upper triangular and $\sum_{j>i} |t_{ij}(\epsilon)| \leq \epsilon$. *Hint:* Apply (2.4.7) with $[2/n(n-1)]\epsilon$ in place of ϵ.

2.4.8 Commuting families and simultaneous triangularization We now use the commuting families version (2.3.3) of Schur's theorem to show that the eigenvalues "add" and "multiply"—in some order—for commuting matrices.

2.4.8.1 Theorem. Suppose $A, B \in M_n$ commute. Then there is an ordering $\alpha_1, \ldots, \alpha_n$ of the eigenvalues of A and an ordering $\beta_1, \ldots, \beta_n$ of the eigenvalues of B such that the eigenvalues of $A+B$ are $\alpha_1+\beta_1, \alpha_2+\beta_2, \ldots, \alpha_n+\beta_n$ and the eigenvalues of AB are $\alpha_1\beta_1, \alpha_2\beta_2, \ldots, \alpha_n\beta_n$. In particular, $\sigma(A+B) \subseteq \sigma(A) + \sigma(B)$ and $\sigma(AB) \subseteq \sigma(A)\sigma(B)$.

Proof: Since A and B commute, (2.3.3) ensures that there is a unitary $U \in M_n$ such that $U^*AU = T = [t_{ij}]$ and $U^*BU = R = [r_{ij}]$ are both upper triangular. The main diagonal entries (and hence also the eigenvalues) of the upper triangular matrix $T + R = U^*(A+B)U$ are $t_{11} + r_{11}, \ldots, t_{nn} + r_{nn}$; these are the eigenvalues of $A + B$ since $A + B$ is similar to $T + R$. The main diagonal entries (and hence also the eigenvalues) of the upper triangular matrix $TR = U^*(AB)U$ are $t_{11}r_{11}, \ldots, t_{nn}r_{nn}$; these are the eigenvalues of AB, which is similar to TR. $\square$

2.4.8.2 Example. Even when A and B commute, not every sum of their respective eigenvalues need be an eigenvalue of $A+B$. Consider the diagonal matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$$

and realize that $1+4 = 5 \notin \{4,6\} = \sigma(A+B)$. Thus $\sigma(A+B)$ is contained in, but is not generally equal to, $\sigma(A)+\sigma(B)$ when A and B commute.

2.4.8.3 Example. If A and B do *not* commute, it is difficult to say how $\sigma(A+B)$ is related to $\sigma(A)$ and $\sigma(B)$. In particular, $\sigma(A+B)$ need not be contained in $\sigma(A)+\sigma(B)$. Let

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

Then $\sigma(A+B) = \{-1, 1\}$, while $\sigma(A) = \sigma(B) = \{0\}$.

2.4.8.4 Example. Is there a converse of (2.4.8.1)? If the eigenvalues of A and B add, in some order, must A and B commute? The answer is no, even if the eigenvalues of αA and βB add, in some order, for all scalars α and β. This is an interesting phenomenon, and the characterization of such pairs of matrices is an unsolved problem! Consider the noncommuting matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

for which $\sigma(A) = \sigma(B) = \{0\}$. Moreover, $p_{\alpha A + bB}(t) = t^3$, so $\sigma(\alpha A + bB) = \{0\}$ for all $\alpha, \beta \in \mathbf{C}$ and the eigenvalues add. If A and B were simultaneously upper triangularizable, the proof of (2.4.8.1) shows that the eigenvalues of AB would be products, in some order, of the eigenvalues of A and B. However, $\sigma(AB) = \{-1, 0, 1\}$ is not contained in $\sigma(A)\cdot\sigma(B) = \{0\}$, so A and B are not simultaneously triangularizable.

2.4.8.5 Corollary. Suppose that $A, B \in M_n$ commute, $\sigma(A) = \{\alpha_1, \ldots, \alpha_{d_1}\}$, and $\sigma(B) = \{\beta_1, \ldots, \beta_{d_2}\}$. If $\alpha_i \neq -\beta_j$ for all i, j, then $A+B$ is nonsingular.

Exercise. Verify (2.4.8.5) using (2.4.8.1).

Exercise. Suppose $T = [t_{ij}]$ and $R = [r_{ij}]$ are n-by-n upper triangular matrices of the same size and let $p(s,t)$ be a polynomial in two noncommuting variables, that is, any linear combination of words in two noncommuting variables. Explain why $p(T, R)$ is upper triangular and its main diagonal entries (its eigenvalues) are $p(t_{11}, r_{11}), \ldots, p(t_{nn}, r_{nn})$.

For complex matrices, simultaneous triangularization and simultaneous unitary triangularization are equivalent concepts.

2.4.8.6 Theorem. Let $A_1, \ldots, A_m \in M_n$ be given. There is a nonsingular $S \in M_n$ such that $S^{-1}A_iS$ is upper triangular for all $i = 1, \ldots, m$ if and only if there is a unitary $U \in M_n$ such that U^*A_iU is upper triangular for all $i = 1, \ldots, m$.

Proof: Use (2.1.14) to write $S = QR$, in which Q is unitary and R is upper triangular. Then $T_i = S^{-1}A_iS = (QR)^{-1}A_i(QR) = R^{-1}(Q^*A_iQ)R$ is upper triangular, so $Q^*A_iQ = RT_iR^{-1}$ is upper triangular, as the product of three upper triangular matrices. □

Simultaneous upper triangularizability of m matrices by similarity is completely characterized by the following theorem of McCoy. It involves a polynomial $p(t_1, \ldots, t_m)$ in m noncommuting variables, which is a linear combination of products of powers of the variables. Since the variables are noncommuting, different powers of the same variables may occur in a given product with products of powers of other variables in between. The key observation is captured in the preceding exercise: if $T_1, \ldots, T_m$ are upper triangular, then so is $p(T_1, \ldots, T_m)$ and the main diagonals of $T_1, \ldots, T_m$ and $p(T_1, \ldots, T_m)$ exhibit specific orderings of their eigenvalues. For each $k = 1, \ldots, n$, the k^{th} main diagonal entry of $p(T_1, \ldots, T_m)$ (an eigenvalue of $p(T_1, \ldots, T_m)$) is the same polynomial in the respective k^{th} main diagonal entries of $T_1, \ldots, T_m$.

2.4.8.7 Theorem. (McCoy) Let $m \geq 2$ and let $A_1, \ldots, A_m \in M_n$ be given. The following statements are equivalent:

(a) For every polynomial $p(t_1, \ldots, t_m)$ in m noncommuting variables and every $k, \ell = 1, \ldots, m$, $p(A_1, \ldots, A_m)(A_kA_\ell - A_\ell A_k)$ is nilpotent.

(b) There is a unitary $U \in M_n$ such that U^*A_iU is upper triangular for each $i = 1, \ldots, m$.

(c) There is an ordering $\lambda_1^{(i)}, \ldots, \lambda_n^{(i)}$ of the eigenvalues of each of the matrices A_i, $i = 1, \ldots, m$, such that for any polynomial $p(t_1, \ldots, t_m)$ in m noncommuting variables, the eigenvalues of $p(A_1, \ldots, A_m)$ are $p(\lambda_i^{(1)}, \ldots, \lambda_i^{(m)})$, $i = 1, \ldots, n$.

Proof: (b) $\Rightarrow$ (c): Let $T_k = U^*A_kU = [t_{ij}^{(k)}]$ be upper triangular and let $\lambda_1^{(k)} = t_{11}^{(k)}, \ldots, \lambda_n^{(k)} = t_{nn}^{(k)}$. Then the eigenvalues of $p(A_1, \ldots, A_m) =$

$p(UT_1U^*, \ldots, UT_mU^*) = Up(T_1, \ldots, T_m)U^*$ are the main diagonal entries of $p(T_1, \ldots, T_m)$, which are $p(\lambda_i^{(1)}, \ldots, \lambda_i^{(m)})$, $i = 1, \ldots, m$.
(c) $\Rightarrow$ (a) For any given polynomial $p(t_1, \ldots, t_m)$ in m noncommuting variables consider the polynomials $q_{k\ell}(t_1, \ldots, t_m) = p(t_1, \ldots, t_m)(t_k t_\ell - t_\ell t_k)$, $k, \ell = 1, \ldots, m$ in m noncommuting variables. The eigenvalues of $q_{k\ell}(A_1, \ldots, A_m)$ are, according to (c), $q_{k\ell}(\lambda_i^{(1)}, \ldots, \lambda_i^{(m)}) = p(\lambda_i^{(1)}, \ldots, \lambda_i^{(m)})(\lambda_i^{(k)}\lambda_i^{(\ell)} - \lambda_i^{(\ell)}\lambda_i^{(k)}) = p(\lambda_i^{(1)}, \ldots, \lambda_i^{(m)}) \cdot 0 = 0$ for all $i = 1, \ldots, n$. Thus, each matrix $p(A_1, \ldots, A_m)(A_k A_\ell - A_\ell A_k)$ is nilpotent. (2.4.2)
(a) $\Rightarrow$ (b): Suppose (see the following lemma) that $A_1, \ldots, A_m$ have a common unit eigenvector x . Subject to this assumption, we proceed by induction as in the proof of (2.3.3). Let U_1 be any unitary matrix that has x as its first column. Use U_1 to deflate each A_i in the same way:

$$\mathcal{A}_i = U_1^* A_i U_1 = \begin{bmatrix} \lambda_1^{(i)} & \bigstar \\ 0 & \tilde{A}_i \end{bmatrix}, \tilde{A}_i \in M_{n-1},\ i = 1, \ldots, m \qquad (2.4.8.8)$$

Let $p(t_1, \ldots, t_m)$ be any given polynomial in m noncommuting variables. Then (a) ensures that the matrix

$$U^* p(A_1, \ldots, A_m)(A_k A_\ell - A_\ell A_k)U = p(\mathcal{A}_1, \ldots, \mathcal{A}_m)(\mathcal{A}_k \mathcal{A}_\ell - \mathcal{A}_\ell \mathcal{A}_k) \qquad (2.4.8.9)$$

is nilpotent for each $k, \ell = 1, \ldots, m$. Partition each of the matrices (2.4.8.9) conformally to (2.4.8.8) and observe that its $1, 1$ entry is zero and its lower right block is $p(\tilde{A}_1, \ldots, \tilde{A}_m)(\tilde{A}_k \tilde{A}_\ell - \tilde{A}_\ell \tilde{A}_k)$, which is necessarily nilpotent. Thus, the matrices $\tilde{A}_1, \ldots, \tilde{A}_m \in M_{n-1}$ inherit property (a) and hence (b) follows by induction, as in (2.3.3). □

We know that commuting matrices always have a common eigenvector (1.3.19). If the matrices $A_1, \ldots, A_m$ in the preceding theorem commute, then the condition (a) is trivially satisfied since $p(A_1, \ldots, A_m)(A_k A_\ell - A_\ell A_k) = 0$ for all $k, \ell = 1, \ldots, m$. The following lemma shows that the condition (a), weaker than commutativity, is sufficient to ensure existence of a common eigenvector.
2.4.8.10 Lemma. Let $A_1, \ldots, A_m \in M_n$ be given. Suppose that for every polynomial $p(t_1, \ldots, t_m)$ in $m \geq 2$ noncommuting variables and every $k, \ell = 1, \ldots, m$, each of the matrices $p(A_1, \ldots, A_m)(A_k A_\ell - A_\ell A_k)$ is nilpotent. Then for each given nonzero vector $x \in \mathbf{C}^n$ there is a polynomial $q(t_1, \ldots, t_m)$ in m noncommuting variables such that $q(A_1, \ldots, A_m)x$ is a common eigenvector of $A_1, \ldots, A_m$.

Proof: We consider only the case $m = 2$, which illustrates all the features of the general case. Let $A, B \in M_n$, let $C = AB - BA$, and assume that

$p(A,B)C$ is nilpotent for every polynomial $p(s,t)$ in two noncommuting variables. Let $x \in \mathbf{C}^n$ be any given nonzero vector. We claim that there is a polynomial $q(s,t)$ in two noncommuting variables such that $q(A,B)x$ is a common eigenvector of A and B.

Begin with (1.1.9) and let $g_1(t)$ be a polynomial such that $\xi_1 = g_1(A)x$ is an eigenvector of A: $A\xi_1 = \lambda\xi_1$.

Case I: Suppose $Cp(B)\xi_1 = 0$ for *every* polynomial $p(t)$, that is,

$$ABp(B)\xi_1 = BAp(B)\xi_1 \text{ for every polynomial } p(t) \tag{2.4.8.11}$$

Using this identity with $p(t) = 1$ shows that $AB\xi_1 = BA\xi_1$. Now proceed by induction: Suppose $AB^k\xi_1 = B^kA\xi_1$ for some $k \geq 1$. Using (2.4.8.11) and the induction hypothesis, we compute

$$\begin{aligned} AB^{k+1}\xi_1 &= AB \cdot B^k\xi_1 = BA \cdot B^k\xi_1 = B \cdot AB^k\xi_1 \\ &= B \cdot B^kA\xi_1 = B^{k+1}A\xi_1 \end{aligned}$$

We conclude that $AB^k\xi_1 = B^kA\xi_1$ for every $k \geq 1$, and hence $Ap(B)\xi_1 = p(B)A\xi_1 = p(B)\lambda\xi_1 = \lambda(p(B)\xi_1)$ for every polynomial $p(t)$. Thus, $p(B)\xi_1$ is an eigenvector of A if it is nonzero. Use (1.1.9) again to choose a polynomial $g_2(t)$ such that $g_2(B)\xi_1 = g_2(B)g_1(A)x$ is an eigenvector of B (necessarily nonzero). Let $q(s,t) = g_2(t)g_1(s)$. We have shown that $q(A,B)x$ is a common eigenvector of A and B, as claimed.

Case II: Suppose there is *some* polynomial $f_1(t)$ such that $Cf_1(B)\xi_1 \neq 0$. Use (1.1.9) to find a polynomial $q_1(t)$ such that $\xi_2 = q_1(A)Cf_1(B)\xi_1$ is an eigenvector of A. If $Cp(B)\xi_2 = 0$ for every polynomial $p(t)$, then Case I permits us to construct the desired common eigenvector; otherwise, let $f_2(t)$ be a polynomial such that $Cf_2(B)\xi_2 \neq 0$ and let $q_2(t)$ be a polynomial such that $\xi_3 = q_2(A)Cf_2(B)\xi_2$ is an eigenvector of A. Continue in this fashion to construct a sequence of eigenvectors

$$\xi_k = q_{k-1}(A)Cf_{k-1}(B)\xi_{k-1},\ k = 2, 3, \ldots \tag{2.4.8.12}$$

of A until either (i) $Cp(B)\xi_k = 0$ for every polynomial $p(t)$, or (ii) $k = n+1$. If (i) occurs for some $k \leq n$, Case I permits us to construct the desired common eigenvector of A and B. If (i) is false for each $k = 1, 2, \ldots, n$, our construction produces $n+1$ vectors $\xi_1, \ldots, \xi_{n+1}$ that must be linearly dependent, so there are $n+1$ scalars $c_1, \ldots, c_{n+1}$, not all zero, such that $c_1\xi_1 + \cdots + c_{n+1}\xi_{n+1} = 0$.

$$\sum_{i=1}^{n+1} c_i\xi_i = 0$$

Let $r = \min\{i : c_i \neq 0\}$. Then

$$\begin{aligned} -c_r\xi_r &= \sum_{i=r}^{n} c_{i+1}\xi_{i+1} = \sum_{i=r}^{n} c_{i+1}q_i(A)Cf_i(B)\xi_i \\ &= c_{r+1}q_r(A)Cf_r(B)\xi_r \\ &\quad + \sum_{i=r}^{n-1} c_{i+2}q_{i+1}(A)Cf_{i+1}(B)\xi_{i+1} \end{aligned} \quad (2.4.8.13)$$

Using (2.4.8.12), the summand in (2.4.8.13) in which $i = r$ can be expanded to

$$c_{r+2}q_{r+1}(A)Cf_{r+1}(B)q_r(A)Cf_r(B)\xi_r$$

In the same fashion, we can use (2.4.8.12) to expand each of the summands in (2.4.8.13) with $i = r+1, r+2, \ldots, n-1$ to an expression of the form $h_i(A,B)Cf_r(B)\xi_r$ in which each $h_i(A,B)$ is a polynomial in A and B. We obtain in this way an identity of the form $-c_r\xi_r = p(A,B)Cf_r(B)\xi_r$, in which $p(s,t)$ is a polynomial in two noncommuting variables. This means that $f_r(B)\xi_r$ is an eigenvector of $p(A,B)C$ associated with the nonzero eigenvalue $-c_r$, in contradiction to the hypothesis that $p(A,B)C$ is nilpotent. This contradiction shows that (i) is true for some $k \leq n$ and hence A and B have a common eigenvector of the asserted form. □

We have stated McCoy's theorem (2.4.8.7) for complex matrices, but if we restate (b) to assert only simultaneous similarity (not simultaneous unitary similarity), then the theorem is valid for matrices and polynomials over any field that contains the eigenvalues of all the matrices $A_1, \ldots, A_m$.

2.4.9 Continuity of eigenvalues Schur's unitary triangularization theorem can be used to prove a basic and widely useful fact: The eigenvalues of a square real or complex matrix depend continuously on its entries. Both aspects of Schur's theorem—*unitary* and *triangular*—play key roles in the proof. The following lemma encapsulates the fundamental principle involved.

2.4.9.1 Lemma. Let an infinite sequence of matrices $A_1, A_2, \ldots \in M_n$ be given and suppose $\lim_{k\to\infty} A_k = A$ (entry-wise convergence). Then there is an infinite sequence of positive integers $k_1 < k_2 < \cdots$ and unitary matrices $U_{k_i} \in M_n$ for $i = 1, 2, \ldots$ such that (a) $T_i \equiv U_{k_i}^* A_{k_i} U_{k_i}$ is upper triangular for all $i = 1, 2, \ldots$; (b) $U \equiv \lim_{i\to\infty} U_{k_i}$ exists and is unitary; (c) $T \equiv U^*AU$ is upper triangular; and (d) $\lim_{i\to\infty} T_i = T$.

Proof: Using (2.3.1), for each $k = 1, 2, \ldots$ let $U_k \in M_n$ be unitary and such that $U_k^* A U_k$ is upper triangular. Lemma (2.1.8) ensures that there is an infinite subsequence $U_{k_1}, U_{k_2}, \ldots$ and a unitary U such that $U_{k_i} \to U$ as $i \to \infty$. Then convergence of each of its three factors ensures that the product $T_i \equiv U_{k_i}^* A_{k_i} U_{k_i}$ converges to a limit $T \equiv U^* A U$, which is upper triangular because each T_i is upper triangular. □

In the preceding argument, the main diagonal of each upper triangular matrix $T, T_1, T_2, \ldots$ is a particular presentation (think of it as an n-vector) of the eigenvalues of $A, A_{k_1}, A_{k_2}, \ldots$, respectively. The entry-wise convergence $T_i \to T$ ensures that among the up to $n!$ different ways of presenting the eigenvalues of each of the matrices $A, A_{k_1}, A_{k_2}, \ldots$ as an n-vector, there is at least one presentation for each matrix such that the respective vectors of eigenvalues converge to a vector whose entries comprise all the eigenvalues of A. It is in this sense, formalized in the following theorem, that the eigenvalues of a square real or complex matrix depend continuously on its entries.

2.4.9.2 Theorem. Let an infinite sequence $A_1, A_2, \ldots \in M_n$ be given and suppose that $\lim_{k\to\infty} A_k = A$ (entry-wise convergence). Let $\lambda(A) = [\lambda_1(A) \ \ldots \ \lambda_n(A)]^T$ and $\lambda(A_k) = [\lambda_1(A_k) \ \ldots \ \lambda_n(A_k)]^T$ be given presentations of the eigenvalues of A and A_k, respectively, for $k = 1, 2, \ldots$. Let $\Pi_n = \{\pi : \pi$ is a permutation of $\{1, 2, \ldots, n\}\}$. Then for each given $\varepsilon > 0$ there exists a positive integer $N = N(\varepsilon)$ such that

$$\min_{\pi \in \Pi_n} \max_{i=1,\ldots,n} \left\{ \left| \lambda_{\pi(i)}(A_k) - \lambda_i(A) \right| \right\} \leq \varepsilon \text{ for all } k \geq N \tag{2.4.9.3}$$

Proof: If the assertion (2.4.9.3) is false, then there is some $\varepsilon_0 > 0$ and an infinite sequence of positive integers $k_1 < k_2 < \cdots$ such that for every $j = 1, 2, \ldots$ we have

$$\max_{i=1,\ldots,n} \left| \lambda_{\pi(i)}(A_{k_j}) - \lambda_i(A) \right| > \varepsilon_0 \text{ for every } \pi \in \Pi_n \tag{2.4.9.4}$$

However, (2.4.9.1) ensures that there is an infinite sub-subsequence $k_1 \leq k_{j_1} < k_{j_2} < \cdots$, unitary matrices $U, U_{k_{j_1}}, U_{k_{j_2}}, \ldots$, and upper triangular matrices $T \equiv U^* A U$ and $T_p \equiv U_{k_{j_p}}^* A_{k_{j_p}} U_{k_{j_p}}$ for $p = 1, 2, \ldots$ such that all of the entries of T_p (in particular, the main diagonal entries) converge to the corresponding entries of T as $p \to \infty$. Since the vectors of main diagonal entries of $T, T_1, T_2, \ldots$ are obtained, respectively, from the given presentations of eigenvalues $\lambda(A), \lambda(A_{k_{j_1}}), \lambda(A_{k_{j_2}}), \ldots$ by permuting their entries, the entry-wise convergence we have observed contradicts (2.4.9.4) and proves the theorem. □

2.4.10 Eigenvalues of a rank one perturbation It is often useful to know that any one eigenvalue of a matrix can be shifted arbitrarily by a rank one perturbation, without disturbing the rest of the eigenvalues.

2.4.10.1 Theorem (A. Brauer) Suppose $A \in M_n$ has eigenvalues $\lambda, \lambda_2, \ldots, \lambda_n$, and let x be a nonzero vector such that $Ax = \lambda x$. Then for any $v \in \mathbf{C}^n$ the eigenvalues of $A + xv^*$ are $\lambda + v^*x, \lambda_2, \ldots, \lambda_n$.

Proof: Let $\xi = x / \|x\|_2$ and let $U = [\xi \; u_2 \; \ldots \; u_n]$ be unitary. Then the proof of (2.3.1) shows that

$$U^*AU = \begin{bmatrix} \lambda & \bigstar \\ 0 & A_1 \end{bmatrix}$$

in which $A_1 \in M_{n-1}$ has eigenvalues $\lambda_2, \ldots, \lambda_n$. Also,

$$\begin{aligned} U^*xv^*U &= \begin{bmatrix} \xi^*x \\ u_2^*x \\ \vdots \\ u_n^*x \end{bmatrix} v^*U = \begin{bmatrix} \|x\|_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} v^*\xi & v^*u_2 & \cdots & v^*u_n \end{bmatrix} \\ &= \begin{bmatrix} \|x\|_2 \, v^*\xi & \bigstar \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} v^*x & \bigstar \\ 0 & 0 \end{bmatrix} \end{aligned}$$

Therefore,

$$U^*(A + xv^*)U = \begin{bmatrix} \lambda + v^*x & \bigstar \\ 0 & A_1 \end{bmatrix}$$

has eigenvalues $\lambda + v^*x, \lambda_2, \ldots, \lambda_n$. □

For a different approach to this result, see (1.2.8).

2.4.11 The complete principle of biorthogonality The *principle of biorthogonality* says that left and right eigenvectors corresponding to different eigenvalues are orthogonal; see (1.4.7). We now address all the possibilities for left and right eigenvectors.

2.4.11.1 Theorem. Let $A \in M_n$, unit vectors $x, y \in \mathbf{C}^n$, and $\lambda, \mu \in \mathbf{C}$ be given.

(a) If $Ax = \lambda x$, $y^*A = \mu y^*$, and $\lambda \neq \mu$, then $y^*x = 0$. Let $U = [x \; y \; u_3 \; \ldots \; u_n] \in M_n$ be unitary. Then

$$U^*AU = \begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \mu & 0 \\ 0 & \bigstar & A_{n-2} \end{bmatrix}, \quad A_{n-2} \in M_{n-2} \qquad (2.4.11.2)$$

(b) Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $y^*x = 0$. Let $U = [x\, y\, u_3\, \ldots\, u_n] \in M_n$ be unitary. Then

$$U^*AU = \begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \lambda & 0 \\ 0 & \bigstar & A_{n-2} \end{bmatrix}, \quad A_{n-2} \in M_{n-2} \tag{2.4.11.3}$$

and the algebraic multiplicity of λ is at least two.

(c) Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $y^*x \neq 0$. Let $S = [x\; S_1] \in M_n$, in which the columns of S_1 are any given basis for the orthogonal complement of y. Then S is nonsingular, the first column of S^{-*} is a nonzero scalar multiple of y, and $S^{-1}AS$ has the block form

$$\begin{bmatrix} \lambda & 0 \\ 0 & A_{n-1} \end{bmatrix}, \quad A_{n-1} \in M_{n-1} \tag{2.4.11.4}$$

If the geometric multiplicity of λ is one, then its algebraic multiplicity is also one. Conversely, if A is similar to a block matrix of the form (2.4.11.4), then it has a non-orthogonal pair of left and right eigenvectors associated with the eigenvalue λ.

(d) Suppose that $Ax = \lambda x$, $y^*A = \lambda y^*$, and $x = y$ (such an x is called a *normal eigenvector*). Let $U = [x\; U_1] \in M_n$ be unitary. Then U^*AU has the block form (2.4.11.4).

Proof: (a) Compared with the reduction in (2.3.1), the extra zeroes in the second row of (2.4.11.2) come from the left eigenvector: $y^*Au_i = \mu y^*u_i = 0$ for $i = 3, \ldots, n$.

(b) The zero pattern in (2.4.11.3) is the same as that in (2.4.11.2), and for the same reason. For the assertion about the algebraic multiplicity, see Problem 14 in (1.2).

(c) See (1.4.7) and (1.4.12). If the algebraic multiplicity of λ for a matrix of the form (2.4.11.4) is at least two, then λ is an eigenvalue of A_{n-1}. Then λ has geometric multiplicity at least one as an eigenvalue of A_{n-1}, which means it has geometric multiplicity at least two as an eigenvalue of A.

(d) Compared with the reduction in (2.3.1), the extra zeroes in the first row of (2.4.11.4) appear because x is also a left eigenvector: $x^*AU_1 = \lambda x^*U_1 = 0$. □

Problems

1. Suppose $A = [a_{ij}] \in M_n$ has n distinct eigenvalues. Use (2.4.9.2) to show that there is an $\epsilon > 0$ such that every $B = [b_{ij}] \in M_n$ with $\sum_{i,j=1}^{n} |a_{ij} -$

$b_{ij}|^2 < \epsilon$ has n distinct eigenvalues. Conclude that the set of matrices with distinct eigenvalues is an *open* subset of M_n.

2. Why is the rank of an upper triangular matrix at least as large as the number of its nonzero main diagonal entries? Let $A = [a_{ij}] \in M_n$, and suppose A has exactly $k \geq 1$ nonzero eigenvalues $\lambda_1, \ldots, \lambda_k$. Write $A = UTU^*$, in which U is unitary and $T = [t_{ij}]$ is upper triangular. Show that $\operatorname{rank} A \geq k$, with equality if A is diagonalizable. Explain why

$$|\sum_{i=1}^{k} \lambda_i|^2 \leq k \sum_{i=1}^{k} |\lambda_i|^2 = k \sum_{i=1}^{n} |t_{ii}|^2 \leq k \sum_{i,j=1}^{n} |t_{ij}|^2 = k \sum_{i=1}^{k} |a_{ij}|^2$$

and conclude that $\operatorname{rank} A \geq |\operatorname{tr} A|^2/(\operatorname{tr} A^*A)$.

3. Our *proof* of the Cayley–Hamilton theorem (2.4.3.2) relies on the fact that complex matrices have eigenvalues, but the definition (1.2.3) of the characteristic polynomial does not involve eigenvalues. The Cayley-Hamilton theorem is valid for matrices whose entries come from *any field*. Indeed, it is valid for matrices whose entries come from a *commutative ring with unit*, examples of which are the ring of integers modulo some integer k (which is a field if and only if k is prime) and the ring of polynomials in one or more formal indeterminants with complex coefficients. Provide details for the following proof that any $A \in M_n$ satisfies the identity $p_A(A) = 0$. The proof is valid for A over a commutative ring with unit.
(a) Start with the fundamental identity $(tI-A)[\operatorname{adj}(tI-A)] = \det(tI-A)I = p_A(t)I$ (0.8.2) and write

$$p_A(t)I = It^n + Ia_{n-1}t^{n-1} + Ia_{n-2}t^{n-2} + \cdots + Ia_1t + Ia_0 \qquad (2.4.12)$$

a polynomial in t of degree n with matrix coefficients; each coefficient is the identity matrix.
(b) Explain why $\operatorname{adj}(tI - A)$ is a matrix whose entries are polynomials in t of degree at most $n-1$, and hence it can be written as

$$\operatorname{adj}(tI - A) = A_{n-1}t^{n-1} + A_{n-2}t^{n-2} + \cdots + A_1t + A_0 \qquad (2.4.13)$$

in which each coefficient A_k is an n-by-n matrix whose entries are polynomial functions of the entries of A and $A_0 = (-1)^{n-1} \operatorname{adj} A$.
(c) Use (2.4.13) and compute the product $(tI - A)[\operatorname{adj}(tI - A)]$ as

$$A_{n-1}t^n + (A_{n-2} - AA_{n-1})t^{n-1} + \cdots + (A_0 - AA_1)t - AA_0 \quad (2.4.14)$$

(d) Match the coefficients of (2.4.12) and (2.4.14) to obtain $n+1$ equations

$$\begin{aligned} A_{n-1} &= I \\ A_{n-2} - AA_{n-1} &= a_{n-1}I \\ &\vdots \\ A_0 - AA_1 &= a_1 I \\ -AA_0 &= a_0 I \end{aligned} \tag{2.4.15}$$

(e) For each $k = 1, \ldots, n$, multiply the k^{th} equation in (2.4.15) by A^{n-k+1}, add all $n+1$ equations, and obtain the Cayley-Hamilton theorem $0 = p_A(A)$.
(f) For each $k = 1, \ldots, n-1$, multiply the k^{th} equation in (2.4.15) by A^{n-k}, add only the first n equations, and obtain the identity

$$\operatorname{adj} A = (-1)^{n-1}(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2 A + a_1 I) \tag{2.4.16}$$

which expresses $\operatorname{adj} A$ as an explicit polynomial in A.
(g) Use (2.4.15) to show that the matrix coefficients in the right-hand side of (2.4.13) are $A_{n-1} = I$ and

$$A_{n-k-1} = A^k + a_{n-1}A^{k-1} + \cdots + a_{n-k+1}A + a_{n-k}I \tag{2.4.17}$$

for $k = 1, 2, \ldots, n-1$, so they are actually polynomials in A.

4. Let $A, B \in M_n$ and suppose A commutes with B. Explain why B commutes with $\operatorname{adj} A$, and why $\operatorname{adj} A$ commutes with $\operatorname{adj} B$. If A is nonsingular, deduce that B commutes with A^{-1}.

5. Consider the matrices $\begin{bmatrix} 0 & \epsilon \\ 0 & 0 \end{bmatrix}$ and explain why there can be non-diagonalizable matrices arbitrarily close to a given diagonalizable matrix. Use Problem 1 to explain why this cannot happen if the given matrix has distinct eigenvalues.

6. Show that for

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -2 & 1 & 2 \\ -1 & -2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

$\sigma(aA + bB) = \{a - 2b, 2a - 2b, 3a + b\}$ for all scalars $a, b \in \mathbf{C}$, but A and B are not simultaneously similar to upper triangular matrices. $\sigma(AB) =?$

7. Use the criterion in Problem 6 in (2.3) to show that the two matrices in Example (2.4.8.6) cannot be simultaneously upper triangularized. Apply the same test to the two matrices in Problem 6.

8. An observation in the spirit of McCoy's theorem (2.4.8.7) can sometimes be useful in showing that two matrices are *not* unitarily similar. Let $p(t,s)$ be a polynomial with complex coefficients in two noncommuting variables, and let $A, B \in M_n$ be unitarily similar with $A = UBU^*$ for some unitary $U \in M_n$. Explain why $p(A, A^*) = Up(B, B^*)U^*$. Conclude that if A and B are unitarily similar, then $\operatorname{tr} p(A, A^*) = \operatorname{tr} p(B, B^*)$ for every complex polynomial $p(t,s)$ in two noncommuting variables. How is this related to Specht's theorem (2.2.6)?

9. Let $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ be a given monic polynomial of degree n with zeroes $\lambda_1, \lambda_2, \ldots, \lambda_n$. Let $\mu_k = \lambda_1^k + \lambda_2^k + \cdots + \lambda_n^k$ denote the kth moments of the zeroes, $k = 0, 1, 2, \ldots$ (take $\mu_0 = n$). Provide details for the following proof of *Newton's identities*

$$ka_{n-k} + a_{n-k+1}\mu_1 + a_{n-k+2}\mu_2 + \cdots + a_{n-1}\mu_{k-1} + \mu_k = 0 \qquad (2.4.18)$$

for $k = 1, 2, \ldots, n-1$ and

$$a_0\mu_k + a_1\mu_{k+1} + \cdots + a_{n-1}\mu_{n+k-1} + \mu_{n+k} = 0 \qquad (2.4.19)$$

for $k = 0, 1, 2, \ldots$. First show that if $|t| > R = \max\{|\lambda_i| : i = 1, \ldots, n\}$, then $(t - \lambda_i)^{-1} = t^{-1} + \lambda_i t^{-2} + \lambda_i^2 t^{-3} + \cdots$ and hence

$$f(t) = \sum_{i=1}^{n} (t - \lambda_i)^{-1} = nt^{-1} + \mu_1 t^{-2} + \mu_2 t^{-3} + \cdots \quad \text{for} \quad |t| > R$$

Now show that $p'(t) = p(t)f(t)$ and compare coefficients. Newton's identities show that the first n moments of the zeroes of a monic polynomial of degree n uniquely determine its coefficients. See Problem 18 in (3.3) for a matrix-analytic approach to Newton's identities.

10. Show that $A, B \in M_n$ have the same characteristic polynomials, and hence the same eigenvalues, if and only if $\operatorname{tr} A^k = \operatorname{tr} B^k$ for all $k = 1, 2, \ldots, n$. Deduce that A is nilpotent if and only if $\operatorname{tr} A^k = 0$ for all $k = 1, 2, \ldots, n$.

11. Let $A, B \in M_n$ be given and consider their *commutator* $C = AB - BA$. Show that (a) $\operatorname{tr} C = 0$. (b) Consider $A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and show that a commutator need not be nilpotent; that is, a commutator can have some nonzero eigenvalues, but their sum must be zero. (c) If $\operatorname{rank} C \le 1$, show that C is nilpotent. *Hint:* Problem 2. (d) If $\operatorname{rank} C = 0$, explain why A and B are simultaneously unitarily triangularizable. (e) Suppose $\operatorname{rank} C = 1$ and provide details for the following sketch of a proof that A and B are simultaneously triangularizable by similarity (*Laffey's theorem*): We may assume that A is singular (replace A by $A - \lambda I$ if necessary). If the null space of A is B-invariant, then it is a common nontrivial invariant subspace, so A and B

are simultaneously similar to a block matrix of the form (1.3.17). If the null space of A is *not* B-invariant, let $x \neq 0$ be such that $Ax = 0$ and $ABx \neq 0$. Then $Cx = ABx$ so there is a $z \neq 0$ such that $C = ABxz^T$. For any y, $(z^T y)ABx = Cy = ABy - BAy$, $BAy = AB(y - (z^T y)x)$, and hence $\operatorname{range} BA \subset \operatorname{range} AB \subset \operatorname{range} A$, $\operatorname{range} A$ is B-invariant, and A and B are simultaneously similar to a block matrix of the form (1.3.17). Now assume that $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, $B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix}$, $A_{11}, B_{11} \in M_k, 1 \leq k < n$, and $C = \begin{bmatrix} A_{11}B_{11} - B_{11}A_{11} & X \\ 0 & A_{22}B_{22} - B_{22}A_{22} \end{bmatrix}$ has rank 1. At least one of the diagonal blocks of C is zero, so we may invoke (2.3.3). If one diagonal block has rank 1 and size greater than one, repeat the reduction. A 1-by-1 diagonal block cannot have rank 1.

12. Let $A, B \in M_n$ and let $C = AB - BA$. (a) If C commutes with A, explain why $\operatorname{tr} C^k = \operatorname{tr}(C^{k-1}(AB - BA)) = \operatorname{tr}(AC^{k-1}B - C^{k-1}BA) = 0$ for all $k = 2, \ldots, n$. Deduce *Jacobson's Lemma* from Problem 10: C is nilpotent if it commutes with either A or B. (b) If $n = 2$, show that C commutes with both A and B if and only if A commutes with B. *Hint*: Explain why we may assume that A and B are upper triangular and C is strictly upper triangular. (c) If A is diagonalizable, show that A commutes with C if and only if A commutes with B. *Hint*: If $A = S\Lambda S^{-1}$, $\Lambda = \lambda_1 I_{n_1} \oplus \cdots \oplus \lambda_d I_{n_d}$, and $\lambda_i \neq \lambda_j$ if $i \neq j$, let $\mathcal{C} = S^{-1}CS$ and $\mathcal{B} = S^{-1}BS$. If Λ commutes with $\mathcal{C}$ then $\mathcal{C}$ is block diagonal conformal to Λ. But $\mathcal{C} = \Lambda\mathcal{B} - \mathcal{B}\Lambda$ has zero diagonal blocks, so $\mathcal{C} = 0$. (d) A and B are said to *quasi-commute* if they both commute with their commutator C. If A and B quasi-commute and $p(s,t)$ is any polynomial in two noncommuting variables, show that $p(A, B)$ commutes with C and use (a) to show that $p(A, B)C$ is nilpotent. (e) If A and B quasi-commute, use McCoy's theorem (2.4.8.7) to show that A and B are simultaneously unitarily triangularizable.

13. Provide details for the following alternative proof of Sylvester's theorem (2.4.4.1) on the linear matrix equation $AX - XB = C$: Suppose $A \in M_n$ and $B \in M_m$ have no eigenvalues in common. Consider the linear transformations $T_1, T_2 : M_{n,m} \to M_{n,m}$ defined by $T_1(X) = AX$ and $T_2(X) = XB$. Show that T_1 and T_2 commute, and deduce from (2.4.8.1) that the eigenvalues of $T - T_1 - T_2$ are differences of eigenvalues of T_1 and T_2. Argue that λ is an eigenvalue of T_1 if and only if there is a nonzero $X \in M_{n,m}$ such that $AX - \lambda X = 0$, which can happen if and only if λ is an eigenvalue of A (and every nonzero column of X is a corresponding eigenvector). The spectra of T_1 and A are therefore the same, and similarly for T_2 and B. Thus, T is nonsingular if A and B have no eigenvalues in common. If x is an eigenvector of A

corresponding to the eigenvalue λ and y is a left eigenvector of B corresponding to the eigenvalue μ, consider $X = xy^*$, show that $T(X) = (\lambda - \mu)X$, and conclude that the spectrum of T consists of all possible differences of eigenvalues of A and B.

14. Let $A \in M_n$ and suppose $\operatorname{rank} A = r$. Provide details to show that A is unitarily similar to an upper triangular matrix whose first r rows are linearly independent and whose last $n - r$ rows are zero. Suppose A has k nonzero eigenvalues. Begin by using (2.3.1) to write $A = UTU^*$ in which U is unitary, $T = \begin{bmatrix} B_1 & \bigstar \\ 0 & C_1 \end{bmatrix}$ is upper triangular, the diagonal entries of $B_1 \in M_k$ are the nonzero eigenvalues of A, the diagonal entries of $C_1 \in M_{n-k}$ are zero, $C_1 = \begin{bmatrix} 0 & A_2 \\ 0 & 0 \end{bmatrix}$, and $A_2 \in M_{n-k-1}$. If $A_2 = 0$ or if A_2 is nonsingular, stop. If not, repeat the reduction.

15. Let $A, B \in M_n$ and consider the polynomial in two complex variables defined by $p_{A,B}(s,t) = \det(tB - sA)$. (a) Suppose that A and B are simultaneously triangularizable, with $A = S\mathcal{A}S^{-1}$, $B = S\mathcal{B}S^{-1}$, $\mathcal{A}$ and $\mathcal{B}$ upper triangular, $\operatorname{diag}\mathcal{A} = (\alpha_1, \ldots, \alpha_n)$, and $\operatorname{diag}\mathcal{B} = (\beta_1, \ldots, \beta_n)$. Show that $p_{A,B}(s,t) = \det(t\mathcal{B} - s\mathcal{A}) = \prod_{i=1}^n (t\beta_i - s\alpha_i)$. (b) Now suppose that A and B commute. Deduce that

$$p_{A,B}(B,A) = \textstyle\prod_{i=1}^n (\beta_i A - \alpha_i B) = S(\prod_{i=1}^n (\beta_i \mathcal{A} - \alpha_i \mathcal{B}))S^{-1}$$

Explain why the i,i entry of the upper triangular matrix $\beta_i\mathcal{A} - \alpha_i\mathcal{B}$ is zero. (c) Use Lemma 2.4.1 to show that $p_{A,B}(B,A) = 0$ *if A and B commute*. Explain why this identity is a two-variable generalization of the Cayley-Hamilton theorem. *Hint:* What is $p_{A,I}(1,t)$ and why is $p_{A,I}(I,A) = 0$? (d) Suppose $A, B \in M_n$ commute. For $n = 2$, show that $p_{A,B}(B,A) = (\det B)A^2 - (\operatorname{tr}(A \operatorname{adj} B))AB + (\det A)B^2$. For $n = 3$, show that $p_{A,B}(B,A) = (\det B)A^3 - (\operatorname{tr}(A \operatorname{adj} B))A^2B + (\operatorname{tr}(B \operatorname{adj} A))AB^2 - (\det A)B^3$. What are these identities for $B = I$? (e) Calculate $\det(tB - sA)$ for the matrices in Examples 2.4.8.3 and 2.4.8.4; discuss. (f) Why did we assume commutativity in (b) but not in (a)?

16. Let λ be an eigenvalue of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2$. (a) Explain why $\mu = a + d - \lambda$ is an eigenvalue of A; (b) Explain why $(A - \lambda I)(A - \mu I) = (A - \mu I)(A - \lambda I) = 0$; (c) Deduce that any nonzero column of $\begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix}$ is an eigenvector of A associated with μ, and any nonzero row is the conjugate transpose of a left eigenvector associated with λ; (d) Deduce that any nonzero column of $\begin{bmatrix} \lambda - d & b \\ c & \lambda - a \end{bmatrix}$ is an eigenvector of A associated with μ and any nonzero row is the conjugate transpose of a left eigenvector associated with λ. See Problem 21 in (3.3) for a generalization.

17. Let $A, B \in M_n$ be given and let

$$\mathcal{S}(A,B) = \operatorname{span}\{I, A, B, A^2, AB, BA, B^2, A^3, A^2B, \ldots\}$$

denote the subalgebra of M_n generated by all powers and products of A and B. Then $\mathcal{S}(A,B)$ is a subspace of M_n, so $\dim \mathcal{S}(A,B) \le n^2$. Consider $n = 2$, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, and $B = A^T$; show in this case that $\dim \mathcal{S}(A,B) = n^2$. Use the Cayley-Hamilton theorem to show that $\dim \mathcal{S}(A,I) \le n$. Gerstenhaber's theorem says that if $A, B \in M_n$ commute, then $\dim \mathcal{S}(A,B) \le n$.

18. Suppose that $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \in M_n, A_{11} \in M_k, 1 \le k < n, A_{22} \in M_{n-k}$. Show that A is nilpotent if and only if both A_{11} and A_{22} are nilpotent.

19. Suppose that $A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} \in M_n, B = \begin{bmatrix} B_{11} & B_{12} \\ 0 & B_{22} \end{bmatrix} \in M_n$, $A_{11}, B_{11} \in M_k, 1 \le k < n, A_{22}, B_{22} \in M_{n-k}$. Show that A and B are simultaneously upper triangularizable if and only if both (a) A_{11} and B_{11} are simultaneously upper triangularizable, and (b) A_{22} and B_{22} are simultaneously upper triangularizable. *Hint:* If A and B are simultaneously upper triangularizable and $p(s,t)$ is any polynomial in two noncommuting variables, then $p(A,B)(AB - BA)$ is nilpotent. (2.4.8.7). Explain why $p(A_{ii}, B_{ii})(A_{ii}B_{ii} - B_{ii}A_{ii})$ is nilpotent for $i = 1, 2$.

20. Suppose $A, B \in M_n$ and $AB = 0$, so $C = AB - BA = -BA$. Let $p(s,t)$ be a polynomial in two noncommuting variables. (a) If $p(0,0) = 0$, show that $Ap(A,B)B = 0$ and hence $(p(A,B)C)^2 = 0$. (b) Show that $C^2 = 0$. (c) Use (2.4.8.7) to show that A and B are simultaneously upper triangularizable. (d) Are $\begin{bmatrix} -3 & 3 \\ -4 & 4 \end{bmatrix}$ and $\begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix}$ simultaneously upper triangularizable?

21. Let $A \in M_n$ have eigenvalues $\lambda_1, \ldots, \lambda_n$. The matrix $K = [\operatorname{tr} A^{i+j-2}]_{i,j=1}^n$ is the *moment matrix* associated with A. We always take $A^0 = I$, so $\operatorname{tr} A^0 = n$. Show that $K = V^T V$, in which $V \in M_n$ is the Vandermonde matrix (0.9.11.1) whose i^{th} row is $[\lambda_1^{i-1}\ \lambda_2^{i-1}\ \ldots\ \lambda_n^{i-1}]^T, j = 1, \ldots, n$. Explain why $\det K = (\det V)^2 = \prod_{i<j}(\lambda_j - \lambda_i)^2$; this product is the *discriminant* of A. Conclude that the eigenvalues of A are distinct if and only if its moment matrix is nonsingular. Explain why K (and hence the discriminant of A) is invariant under similarity of A.

22. Suppose that $A \in M_n$ has d distinct eigenvalues $\lambda_1, \ldots, \lambda_d$ with respective multiplicities $\nu_1 \ldots, \nu_d$. The matrix $K_m = [\operatorname{tr} A^{i+j-2}]_{i,j=1}^m$ is the *moment matrix of order m* associated with A, $m = 1, 2, \ldots$; if $m \le n$, it is a leading principal submatrix of the moment matrix K in the preceding problem. Let $v_j^{(m)} = [1\ \lambda_j\ \lambda_j^2\ \ldots\ \lambda_j^{m-1}]^T, j = 1, \ldots, n$ and form the m-by-d matrix

$V_m = [v_1^{(m)}\ v_2^{(m)}\ \ldots\ v_d^{(m)}]$. Let $D = \operatorname{diag}(\nu_1, \ldots, \nu_d) \in M_d$. Show that: (a) V_m has row rank m if $m \leq d$ and has column rank d if $m \geq d$. (b) $K_m = V_m D V_m^T$. *Hint*: $K_m = [\sum_{k=1}^d \nu_k \lambda_k^{i+j-2}]_{i,j=1}^m = \sum_{k=1}^d \nu_k v_k^{(m)} (v_k^{(m)})^T$. (c) If $1 \leq p < q$, K_p is a leading principal submatrix of K_q. (d) K_d is nonsingular. (e) $\operatorname{rank} K_m = d$ if $m \geq d$. *Hint*: $\operatorname{rank} V_m = \operatorname{rank} D = d \Rightarrow \operatorname{rank} K_m \leq d$; K_d nonsingular $\Rightarrow \operatorname{rank} K_m \geq d$. (f) $d = \max\{m \geq 1 : K_m \text{ is nonsingular}\}$ but K_p can be singular for some $p < d$. (g) $K_{n+1}, K_n, \ldots, K_{d+1}$ are all singular but K_d is nonsingular. (h) $K_n = K$, the moment matrix in the preceding problem. Thus, $\operatorname{rank} K$ is exactly the number of distinct eigenvalues of A.

23. Suppose that $T = [t_{ij}] \in M_n$ is upper triangular. Show that $\operatorname{adj} T = [\tau_{ij}]$ is upper triangular and has main diagonal entries $\tau_{ii} = \prod_{j \neq i} t_{jj}$.

24. Let $A \in M_n$ have eigenvalues $\lambda_1, \ldots, \lambda_n$. Show that the eigenvalues of $\operatorname{adj} A$ are $\prod_{j \neq i} \lambda_j, i = 1, \ldots, n$.

25. Let $A, B \in M_2$ and suppose λ_1, λ_2 are the eigenvalues of A. (a) Show that A is unitarily similar to $\begin{bmatrix} \lambda_1 & x \\ 0 & \lambda_2 \end{bmatrix}$ in which $x \geq 0$ and $x^2 = \operatorname{tr} AA^* - |\lambda_1|^2 - |\lambda_2|^2$. (b) Show that A is unitarily similar to B if and only if $\operatorname{tr} A = \operatorname{tr} B$, $\operatorname{tr} A^2 = \operatorname{tr} B^2$, and $\operatorname{tr} AA^* = \operatorname{tr} BB^*$. *Hint:* Problem 10.

26. Let $B \in M_{n,k}$ and $C \in M_{k,m}$ be given. Show that $BCp(BC) = Bp(CB)C$ for any polynomial $p(t)$.

27. Let $A \in M_n$ be given. (a) If $A = BC$ and $B, C^T \in M_{n,k}$, use (2.4.3.2) to show that there is a polynomial $q(t)$ of degree at most $k+1$ such that $q(A) = 0$. *Hint:* Take $p(t) = p_{CB}(t)$ in the preceding problem.

28. Suppose $A \in M_n$ is singular and let $r = \operatorname{rank} A$. Show that there is a polynomial $p(t)$ of degree at most $r + 1$ such that $p(A) = 0$. *Hint:* (0.4.6(e)).

29. Let $A \in M_n$ and suppose $x, y \in \mathbf{C}^n$ are nonzero vectors such that $Ax = \lambda x$ and $y^* A = \lambda y^*$. If λ is a simple eigenvalue of A, show that $A - \lambda I + \kappa x y^*$ is nonsingular for all $\kappa \neq 0$.

30. There is a systematic approach to the calculations illustrated in (2.4.3.3). Let $A \in M_n$ be given and suppose that $p(t)$ is a polynomial of degree greater than n. Use the Euclidean algorithm (polynomial long division) to express $p(t) = h(t)p_A(t) + r(t)$, in which the degree of $r(t)$ is strictly less than n (possibly zero). Explain why $p(A) = r(A)$.

31. Use (2.4.3.2) to prove that if all the eigenvalues of $A \in M_n$ are zero, then $A^n = 0$. *Hint*: What is $p_A(t)$ in this case?

Further Readings and Notes. See [RadRos] for a detailed exposition of simultaneous triangularization. Theorem (2.4.8.7) and its generalizations were proved by N. McCoy, On the Characteristic Roots of Matric Polynomials, *Bull. Amer. Math. Soc*. 42 (1936) 592–600. Our proof for (2.4.8.7) is adapted from M. P. Drazin, J. W. Dungey, and K. W. Gruenberg, Some Theorems on Commutative Matrices, *J. Lond. Math. Soc.* 26 (1951) 221-228, which contains a proof of (2.4.8.10) in the general case $m \geq 2$. The relationship between eigenvalues and linear combinations is discussed in T. Motzkin and O. Taussky, Pairs of Matrices with Property *L*, *Trans. Amer. Math. Soc*. 73 (1952) 108–114. A pair $A, B \in M_n$ such that $\sigma(aA + bB) = \{a\alpha_j + b\beta_{i_j} : j = 1, \ldots, n\}$ for all $a, b \in \mathbf{C}$ is said to have *property L*; the condition (2.4.8.7(c)) is called *property P*. Property *P* implies property *L*, but not conversely. Property *L* is not fully understood, although it is known that a pair of normal matrices with property *L* must commute and must therefore be simultaneously unitarily diagonalizable.

2.5 Normal matrices

The class of normal matrices, which arises naturally in the context of unitary similarity, is important throughout matrix analysis; it includes the unitary, Hermitian, skew-Hermitian, real orthogonal, real symmetric, and real skew-symmetric matrices.

2.5.1 Definition. A matrix $A \in M_n$ is *normal* if $A^*A = AA^*$, that is, if A commutes with its conjugate transpose.

Exercise. If $A \in M_n$ is normal and $\alpha \in \mathbf{C}$, show that αA is normal. *The class of normal matrices of a given size is closed under multiplication by complex scalars.*

Exercise. If $A \in M_n$ is normal, and if B is unitarily similar to A, show that B is normal. *The class of normal matrices of a given size is closed under unitary similarity.*

Exercise. If $A \in M_n$ and $B \in M_m$ are normal, show that $A \oplus B \in M_{n+m}$ is normal. *The class of normal matrices is closed under direct sums.*

Exercise. If $A \in M_n$ and $B \in M_m$, and if $A \oplus B \in M_{n+m}$ is normal, show that A and B are normal.

Exercise. Show that $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ is normal for all $a, b \in \mathbf{C}$.

Exercise. Show that every unitary matrix is normal.

Exercise. Show that every Hermitian or skew-Hermitian matrix is normal.

Exercise. Verify that $A = \begin{bmatrix} 1 & e^{i\pi/4} \\ -e^{i\pi/4} & 1 \end{bmatrix}$ is normal, but no scalar multiple of A is unitary, Hermitian, or skew-Hermitian.

Exercise. Explain why every diagonal matrix is normal. If a diagonal matrix is Hermitian, why must it be real?

Exercise. Show that each of the classes of unitary, Hermitian, and skew-Hermitian matrices is closed under unitary similarity. If A is unitary and $|\alpha| = 1$, show that αA is unitary. If A is Hermitian and α is real, show that αA is Hermitian. If A is skew-Hermitian and α is real, show that αA is skew-Hermitian.

Exercise. Show that a Hermitian matrix must have real main diagonal entries and a skew-Hermitian matrix must have pure imaginary main diagonal entries. What are the main diagonal entries of a real skew-symmetric matrix?

Exercise. Review the proof of (1.3.7) and conclude that $A \in M_n$ is unitarily diagonalizable if and only if there is a set of n orthonormal vectors in $\mathbf{C}^n$, each of which is an eigenvector of A.

There is something special about certain zero blocks in a normal matrix.

2.5.2 Lemma. Suppose $A \in M_n$ is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

in which A_{11} and A_{22} are square. Then A is normal if and only if A_{11} and A_{22} are normal, and $A_{12} = 0$. A block upper triangular matrix is normal if and only if each of its off-diagonal blocks is zero and each of its diagonal blocks is normal; in particular, an upper triangular matrix is normal if and only if it is diagonal.

Proof: If A_{11} and A_{22} are normal and $A_{12} = 0$, then $A = A_{11} \oplus A_{22}$ is a direct sum of normal matrices, so it is normal. Conversely, if A is normal, then

$$AA^* = \begin{bmatrix} A_{11}A_{11}^* + A_{12}A_{12}^* & \bigstar \\ \bigstar & \bigstar \end{bmatrix} = \begin{bmatrix} A_{11}^*A_{11} & \bigstar \\ \bigstar & \bigstar \end{bmatrix} = A^*A$$

so $A_{11}^*A_{11} = A_{11}A_{11}^* + A_{12}A_{12}^*$, which implies that

$$\begin{aligned} \operatorname{tr} A_{11}^*A_{11} &= \operatorname{tr}(A_{11}A_{11}^* + A_{12}A_{12}^*) \\ &= \operatorname{tr}(A_{11}A_{11}^*) + \operatorname{tr}(A_{12}A_{12}^*) = \operatorname{tr}(A_{11}^*A_{11}) + \operatorname{tr}(A_{12}A_{12}^*) \end{aligned}$$

and hence $\operatorname{tr}(A_{12}A_{12}^*) = 0$. Since $\operatorname{tr}(A_{12}A_{12}^*)$ is the sum of squares of the absolute values of the entries of A_{12} (0.2.5.1), it follows that $A_{12} = 0$. Then $A = A_{11} \oplus A_{22}$ is normal, so A_{11} and A_{22} are normal.

Suppose $B = [B_{ij}]_{i,j=1}^k \in M_n$ is normal and block upper triangular, that is, $B_{ii} \in M_{n_i}$ for $i = 1, \dots, k$ and $B_{ij} = 0$ if $i > j$). Partition it as

$$B = \begin{bmatrix} B_{11} & X \\ 0 & \tilde{B} \end{bmatrix}$$

in which $X = [B_{12} \ \dots \ B_{1k}]$ and $\tilde{B} = [B_{ij}]_{i,j=2}^k$ is block upper triangular. Then $X = 0$ and $\tilde{B}$ is normal, so a finite induction permits us to conclude that B is block diagonal. For the converse, we have observed in a preceding exercise that a direct sum of normal matrices is normal. □

Exercise. Let $A \in M_n$ be normal and let $\alpha \in \{1, \dots, n\}$ be a given index set. If $A[\alpha, \alpha^c] = 0$, show that $A[\alpha^c, \alpha] = 0$.

We next catalog the most fundamental facts about normal matrices. The equivalence of (a) and (b) in the following theorem is often called the *spectral theorem for normal matrices*.

2.5.3 Theorem. Let $A = [a_{ij}] \in M_n$ have eigenvalues $\lambda_1, \dots, \lambda_n$. The following statements are equivalent:

(a) A is normal;
(b) A is unitarily diagonalizable;
(c) $\sum_{i,j=1}^n |a_{ij}|^2 = \sum_{i=1}^n |\lambda_i|^2$; and
(d) There is an orthonormal set of n eigenvectors of A.

Proof: Use (2.3.1) to write $A = UTU^*$, in which $U = [u_1 \ \dots \ u_n]$ is unitary and $T = [t_{ij}] \in M_n$ is upper triangular.

If A is normal, then so is T (as is every matrix that is unitarily similar to A). The preceding lemma ensures that T is actually a diagonal matrix, so A is unitarily diagonalizable.

If there is a unitary V such that $A = V\Lambda V^*$ and $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, then $\operatorname{tr} A^*A = \operatorname{tr} \Lambda^*\Lambda$ by (2.2.2), which is the assertion in (c).

The diagonal entries of T are $\lambda_1, \dots, \lambda_n$ in some order, and hence $\operatorname{tr} A^*A = \operatorname{tr} T^*T = \sum_{i=1}^n |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2$. Thus, (c) implies that $\sum_{i<j}^n |t_{ij}|^2 = 0$, so T is diagonal. The factorization $A = UTU^*$ is equivalent to the identity $AU = UT$, which says that $Au_i = t_{ii}u_i$ for each $i = 1, \dots, n$. Thus, each column of U (an orthonormal set) is an eigenvector of A.

Finally, an orthonormal set is linearly independent, so (d) ensures that A is diagonalizable and that a diagonalizing similarity can be chosen with orthonormal columns (1.3.7). This means that A is unitarily similar to a diagonal (and hence normal) matrix, so A is normal. □

A representation of a normal matrix $A \in M_n$ as $A = U\Lambda U^*$, in which U is unitary and Λ is diagonal, is called a *spectral decomposition* of A.

Exercise. Explain why a normal matrix is nondefective, that is, the geometric multiplicity of every eigenvalue is the same as its algebraic multiplicity.

Exercise. If $A \in M_n$ is normal, show that $x \in \mathbf{C}^n$ is a right eigenvector of A corresponding to the eigenvalue λ of A if and only if x is a left eigenvector of A corresponding to λ; that is, $Ax = \lambda x$ is equivalent to $x^*A = \lambda x^*$ if A is normal. *Hint:* Normalize x and write $A = U\Lambda U^*$ with x as the first column of U. Then what is A^*? A^*x? See Problem 20 for another proof.

Exercise. If $A \in M_n$ is normal, and if x and y are eigenvectors of A corresponding to distinct eigenvalues, use the preceding exercise and the principle of biorthogonality to show that x and y are orthogonal.

Once the distinct eigenvalues $\lambda_1, \ldots, \lambda_d$ of a normal matrix $A \in M_n$ are known, it can be unitarily diagonalized via the following conceptual prescription: For each eigenspace $\{x \in \mathbf{C}^n : Ax = \lambda x\}$, determine a basis and orthonormalize it (use the Gram–Schmidt procedure, for example) to obtain an orthonormal basis. The eigenspaces are mutually orthogonal and the dimension of each eigenspace is equal to the multiplicity of the corresponding eigenvalue (normality of A is the reason for both), so the union of these bases is an orthonormal basis for $\mathbf{C}^n$. Arraying these basis vectors as the columns of a matrix U produces a unitary matrix such that U^*AU is diagonal.

However, an eigenspace always has more than one orthonormal basis (Why?), so the diagonalizing unitary matrix constructed in the preceding conceptual prescription is never unique. If $X, Y \in M_{n,k}$ have orthonormal columns ($X^*X = I_k = Y^*Y$), and if $\operatorname{range} X = \operatorname{range} Y$, then each column of X is a linear combination of the columns of Y, that is, $X = YG$ for some $G \in M_k$. Then $I_k = X^*X = (YG)^*(YG) = G^*(Y^*Y)G = G^*G$, so G must be unitary. This observation gives a geometric interpretation for the first part of the following uniqueness theorem.

2.5.4 Theorem. Let $A \in M_n$ be normal and have distinct eigenvalues $\lambda_1, \ldots, \lambda_d$, with respective multiplicities $n_1, \ldots, n_d$. Let $\Lambda = \lambda_1 I_{n_1} \oplus \cdots \oplus \lambda_d I_{n_d}$, and suppose that $U \in M_n$ is unitary and $A = U\Lambda U^*$. (a) $A = V\Lambda V^*$ for some unitary $V \in M_n$ if and only if there are unitary matrices $W_1, \ldots, W_d$ with each $W_i \in M_{n_i}$ such that $U = V(W_1 \oplus \cdots \oplus W_d)$. (b) Two normal matrices are unitarily similar if and only if they have the same eigenvalues.

Proof: (a) If $U\Lambda U^* = V\Lambda V^*$, then $\Lambda U^*V = U^*V\Lambda$, so $W = U^*V$ is unitary and commutes with Λ; (2.4.4.2) ensures that W is block diagonal conformal to Λ. Conversely, if $U = VW$ and $W = W_1 \oplus \cdots \oplus W_d$ with each $W_i \in M_{n_i}$, then W commutes with Λ and $U\Lambda U^* = VW\Lambda W^*V^* = V\Lambda WW^*V^* = V\Lambda V^*$. (b) If $B = V\Lambda V^*$ for some unitary V, then $(UV^*)B(UV^*)^* = (UV^*)V\Lambda V^*(UV^*)^* = U\Lambda U^* = A$. Conversely, if B is similar to A, then they have the same eigenvalues; if B is unitarily similar to a normal matrix, then it is normal. $\square$

We next note that commuting normal matrices may be simultaneously unitarily diagonalized.

2.5.5 Theorem. Let $\mathcal{N} \subseteq M_n$ be a nonempty family of normal matrices. Then $\mathcal{N}$ is a commuting family if and only if it is a simultaneously unitarily diagonalizable family. For any given $A_0 \in \mathcal{N}$ and for any given ordering $\lambda_1, \ldots, \lambda_n$ of the eigenvalues of A_0, there is a unitary $U \in M_n$ such that $U^*A_0U = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and U^*BU is diagonal for every $B \in \mathcal{N}$.

Exercise. Use (2.3.3) and the fact that a triangular normal matrix must be diagonal to prove (2.5.5). The final assertion about A_0 follows as in the proof of (1.3.21), since every permutation matrix is unitary.

Application of (2.5.3) to the special case of Hermitian matrices yields a fundamental result called the *spectral theorem for Hermitian matrices*.

2.5.6 Theorem. If $A \in M_n$ is Hermitian, then (a) the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A are real; (b) A is unitarily diagonalizable; and (c) there is a unitary $U \in M_n$ such that $A = U\Lambda U^*$, in which $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. If $A \in M_n(\mathbf{R})$ is symmetric, then A has real eigenvalues and is real orthogonally diagonalizable.

Proof: A diagonal Hermitian matrix must have real diagonal entries, so (a) follows from (b) and the fact that the set of Hermitian matrices is closed under unitary similarity. Statement (b) follows from (2.5.3) because Hermitian matrices are normal. Statement (c) restates (b) and incorporates the information that the diagonal entries of Λ are necessarily the eigenvalues of A. If $A \in M_n(\mathbf{R})$ is symmetric, then it is Hermitian, but since it is real, has real eigenvalues, and is nondefective, all calculations necessary to diagonalize A (determine each eigenspace and find an orthonormal basis of it) can take place over the real field. $\square$

Exercise. Modify the proof of Theorem 2.5.4 to show that if A is real symmetric and U, V are real orthogonal, then the matrices $W_1, \ldots, W_d$ may be taken to be real orthogonal.

In contrast to the discussion of diagonalizability in Chapter 1, there is no reason to assume distinctness of eigenvalues in (2.5.4) and (2.5.6), and diagonalizability need not be assumed in (2.5.5). A full linearly independent set of eigenvectors (in fact, an orthonormal set) is structurally guaranteed by normality. This is one reason why Hermitian and normal matrices are so important and have such pleasant properties.

We conclude with a discussion of real normal matrices. Such matrices can be diagonalized by a complex unitary similarity, but what special form can they can be put into by a *real orthogonal* similarity? Since a real normal matrix can have non-real eigenvalues, it might not be possible to diagonalize it with a real similarity. On the other hand, any real matrix can be put into a special block upper triangular form by a real orthogonal similarity (2.3.4), and this suggests what to do if the matrix is also normal. Our proof uses (2.3.4) in the same way that (2.3.1) was used in the proof of (2.5.4).

Exercise. Let $x \in \mathbf{C}^n$ be given and write $x = u + iv$, in which $u, v \in \mathbf{R}^n$. Show that x is orthogonal to $\bar{x}$ if and only if u is orthogonal to v and $\|u\|_2 = \|v\|_2$.

2.5.7 Theorem. Suppose that $A \in M_n(\mathbf{R})$ has p complex conjugate pairs of non-real eigenvalues

$$\lambda_1 = a_1 + ib_1, \overline{\lambda_1} = a_1 - ib_1, \ldots, \lambda_p = a_p + ib_p, \overline{\lambda_p} = a_p - ib_p \quad (2.5.8)$$

in which all $a_j, b_j \in \mathbf{R}$ and all $b_j > 0$, and, if $2p < n$, an additional $n - 2p$ real eigenvalues $\mu_1, \ldots, \mu_{n-2p}$. If A is normal, then there is a real orthogonal $Q \in M_n(\mathbf{R})$ such that

$$Q^T A Q = \begin{bmatrix} A_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_{n-p} \end{bmatrix} \quad (2.5.9)$$

is real and block diagonal, and each diagonal block is either 1-by-1 or 2-by-2. There are $n - 2p$ real blocks of the form $[\mu_j]$, one for each of the real eigenvalues $\mu_1, \ldots, \mu_{n-2p}$. There are p real blocks of the form

$$\begin{bmatrix} a_j & b_j \\ -b_j & a_j \end{bmatrix}, \quad a_j, b_j \in \mathbf{R}, b_j > 0 \quad (2.5.10)$$

one for each complex conjugate pair of non-real eigenvalues $\lambda_j, \overline{\lambda_j} = a_j \pm ib_j$. The p 2-by-2 blocks and the $n - 2p$ 1-by-1 blocks may appear in the block diagonal of (2.5.9) in any prescribed order. Any matrix of the form (2.5.9) is normal. Two real normal matrices are real orthogonally similar if and only if they have the same eigenvalues.

Proof: Follow the proof of (2.3.4). Every deflation involving a real eigenvalue is achieved with a real orthogonal similarity. For a conjugate pair of non-real eigenvalues $\lambda_j, \bar{\lambda}_j = a_j \pm ib_j$ with associated eigenvectors x_j and $\bar{x}_j$, the deflation is achieved with a real orthogonal similarity that yields a block of the form (2.5.10) *provided* that x_j and $\bar{x}_j$ are orthogonal and normalized so that $\|x_j\|_2^2 = \|u_j\|_2^2 + \|v_j\|_2^2 = 2$ (u_j and v_j are then real orthogonal unit vectors, per the preceding exercise). Since A is normal, a previous exercise ensures that x_j and $\bar{x}_j$ are indeed orthogonal, and there is no loss of generality to assume that $\|x_j\|_2^2 = 2$. Invoking (2.5.2), we use normality of A once again to conclude that every off-diagonal block in (2.3.5) is zero. For the converse, a computation reveals that any matrix of the form (2.5.10) is normal and has eigenvalues $a_j \pm ib_j$. The final assertion follows from our construction: the diagonal blocks in (2.5.9) are completely determined by the eigenvalues of the real normal matrix A. $\square$

As a consequence of this theorem for real normal matrices, we can deduce real canonical forms for real matrices that are symmetric, skew-symmetric, or real orthogonal.

2.5.11 Corollary. Let $A \in M_n(\mathbf{R})$. Then

(a) $A = A^T$ if and only if there is a real orthogonal $Q \in M_n(\mathbf{R})$ such that

$$Q^T A Q = \begin{bmatrix} \lambda_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_n \end{bmatrix}, \text{ all } \lambda_i \in \mathbf{R} \tag{2.5.12}$$

The parameters $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. Two real symmetric matrices are real orthogonally similar if and only if they have the same eigenvalues.

(b) $A = -A^T$ if and only if there is a real orthogonal $Q \in M_n(\mathbf{R})$ such that

$$Q^T A Q = \begin{bmatrix} 0_{n-2p} & 0 \\ 0 & \tilde{A} \end{bmatrix} \tag{2.5.13}$$

in which

$$\tilde{A} = \begin{bmatrix} 0 & b_1 \\ -b_1 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 0 & b_p \\ -b_p & 0 \end{bmatrix}$$

and all $b_j \in \mathbf{R}$ are nonzero. The nonzero eigenvalues of A are $\pm ib_1, \dots, \pm ib_p$. Two real skew-symmetric matrices are real orthogonally similar if and only if they have the same eigenvalues.

(c) $AA^T = I$ if and only if there is a real orthogonal $Q \in M_n(\mathbf{R})$ such that

$$Q^T A Q = \begin{bmatrix} \Lambda_{n-2p} & 0 \\ 0 & \tilde{A} \end{bmatrix} \tag{2.5.14}$$

in which $\Lambda_{n-2p} = \text{diag}(\pm 1, \ldots, \pm 1) \in M_{n-2p}(\mathbf{R})$ and

$$\tilde{A} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} \cos\theta_p & \sin\theta_p \\ -\sin\theta_p & \cos\theta_p \end{bmatrix}, 0 < \theta_j < \pi$$

The eigenvalues of A are the diagonal entries of Λ_{n-2p} together with $e^{\pm i\theta_1}, \ldots, e^{\pm i\theta_p}$. Two real orthogonal matrices are real orthogonally similar if and only if they have the same eigenvalues.

Proof: In each case, the hypothesis guarantees that A is real and normal, so it can be written in the form (2.5.9). If $A = A^T$, then there can be no 2-by-2 diagonal blocks. If $A = -A^T$, then every 1-by-1 diagonal block is zero and every 2-by-2 diagonal block has a zero main diagonal. If $AA^T = I$, then each 1-by-1 block must be $[\pm 1]$ and each 2-by-2 block has determinant ± 1, so each $a_j^2 + b_j^2 = 1$ and so we may write each $a_j = \cos\theta_j, |b_j| = \sin\theta_j$, that is, $a_j \pm ib_j = e^{\pm i\theta_j}$. □

If one has a family of commuting real normal matrices, they might not be simultaneously real orthogonally diagonalizable, but they can all be brought simultaneously into the block diagonal form (2.5.9).

Exercise. Let $A_1 = \begin{bmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{bmatrix}$ and $A_2 = \begin{bmatrix} a_2 & b_2 \\ -b_2 & a_2 \end{bmatrix}$. Show that A_1 commutes with A_2 for all $a_1, a_2, b_1, b_2 \in \mathbf{C}$.

Exercise. Let $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ and $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ in which $a, b, \lambda_1, \lambda_2 \in \mathbf{C}$ and $b \neq 0$. Show that A commutes with Λ if and only if $\lambda_1 = \lambda_2$.

2.5.15 Theorem. Let $\mathcal{N} \subseteq M_n(\mathbf{R})$ be a commuting family of real normal matrices. Suppose that the non-real eigenvalues of a given $A_0 \in \mathcal{N}$ are $\lambda_1, \overline{\lambda_1}, \ldots, \lambda_p, \overline{\lambda_p}$, in which $\lambda_j = a_j + ib_j$ with all $a_j, b_j \in \mathbf{R}$ and $b_j > 0$. Let $\mu_1, \ldots, \mu_{n-2p}$ be the real eigenvalues of A_0. Then:

(a) There is a single real orthogonal matrix Q such that $Q^T AQ$ has the block diagonal form (2.5.9) for each $A \in \mathcal{N}$ and

$$Q^T A_0 Q = \text{diag}(\mu_1, \ldots, \mu_{n-2p}) \oplus \begin{bmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} a_p & b_p \\ -b_p & a_p \end{bmatrix}$$

(b) Let Q be as in (a) and consider a given index set $\gamma = \{j, j+1\}$ with

$1 \le j < n-2p$, for which $(Q^T A_0 Q)[\gamma] = \begin{bmatrix} \mu_j & 0 \\ 0 & \mu_{j+1} \end{bmatrix}$. If $\mu_j \neq \mu_{j+1}$, then $(Q^T AQ)[\gamma]$ is a diagonal matrix for every $A \in \mathcal{N}$.

(c) Let Q be as in (a) and consider a given index set $\gamma = \{n-2p+2j-1, n-2p+2j\}$ with $1 \le j \le p$. Then for each $A \in \mathcal{N}$, $(Q^T AQ)[\gamma] = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ for some $a, b \in \mathbf{R}$ ($b = 0$ is possible).

Proof: Use (2.3.6) to reduce every member of $\mathcal{N}$ to the form (2.3.5) via one real orthogonal similarity Q. The argument in the proof of (2.5.7) shows that they have the form (2.5.9). If necessary, a final simultaneous permutation similarity achieves the desired presentation of $Q^T A_0 Q$ and preserves the block diagonal structure of $Q^T \mathcal{N} Q$. The assertions in (b) and (c) about the common block form of all the matrices in $Q^T \mathcal{N} Q$ follow from commutativity and the two preceding exercises. $\square$

A matrix identity of the form $AX = XB$ is known as an *intertwining relation*. A familiar intertwining relation is the commutativity equation $AB = BA$; other examples are $AB = BA^T$, $AB = B\bar{A}$, and $AB = BA^*$. In (2.4.4) we made use of the important fact that if $AX = XB$ and $A, B \in M_n$, then $p(A)X = Xp(B)$ for any polynomial $p(t)$.

A fundamental principle worth keeping in mind is that if $AX = XB$ and if there is something special about the structure of A and B, then there is likely to be something special about the structure of X. One may be able to discover what that special structure is by replacing A and B by canonical forms and studying the resulting intertwining relation involving the canonical forms and a transformed X.

If A and B are normal (either complex or real) and satisfy an intertwining relation, the *Fuglede-Putnam theorem* says that A^* and B^* satisfy the same intertwining relation. The key to understanding this result is the scalar case: if $a, b \in \mathbf{C}$, then $ab = 0$ if and only if $a\bar{b} = 0$.

2.5.16 Theorem (Fuglede-Putnam). Let $A \in M_n$ and $B \in M_m$ be normal and let $X \in M_{n,m}$ be given. Then $AX = XB$ if and only if $A^*X = XB^*$.

Proof: Let $A = U\Lambda U^*$ and $B = VMV^*$ be spectral decompositions in which $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $M = \operatorname{diag}(\mu_1, \ldots, \mu_m)$. Let $U^*XV = [\xi_{ij}]$. Then $AX = XB \iff U\Lambda U^*X = XVMV^* \iff \Lambda(U^*XV) = (U^*XV)M \iff \lambda_i\xi_{ij} = \xi_{ij}\mu_j$ for all $i, j \iff \xi_{ij}(\lambda_i - \mu_j) = 0$ for all $i, j \iff \xi_{ij}\overline{(\lambda_i - \mu_j)} = 0$ for all $i, j \iff \overline{\lambda_i}\xi_{ij} = \xi_{ij}\overline{\mu_j}$ for all $i, j \iff \bar{\Lambda}(U^*XV) = (U^*XV)\bar{M} \iff U\bar{\Lambda}U^*X = XV\bar{M}V^* \iff A^*X = XB^*$. $\square$

The preceding two theorems lead to a useful representation for normal matrices that commute with their transpose or, equivalently, with their complex conjugate.

Exercise. If $A \in M_n$, explain why $\bar{A}A = A\bar{A}$ if and only if $A\bar{A}$ is real.

Exercise. Let $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \in M_2$ be given. Explain why: (a) A is nonsingular if and only if $A = c\begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}, c \neq 0$, and $\alpha^2 + \beta^2 = 1$. (b) A is singular and nonzero if and only A is a nonzero scalar multiple of $\begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$ or its complex conjugate.

2.5.17 Theorem. Let $A \in M_n$ be normal. The following three statements are equivalent:
(a) $\bar{A}A = A\bar{A}$;
(b) $A^T A = AA^T$; and
(c) There is a real orthogonal Q such that $Q^T AQ$ is a direct sum of blocks, each of which is either a zero block or a nonzero scalar multiple of

$$[1], \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \begin{bmatrix} a & b \\ -b & a \end{bmatrix}, \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}, \text{ or } \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}, \quad a, b \in \mathbb{C}, \tag{2.5.18}$$

in which $a \neq 0 \neq b$ and $a^2 + b^2 = 1$.
Conversely, if A is real orthogonally similar to a direct sum of complex scalar multiples of blocks of the form (2.5.18), then A is normal and $A\bar{A} = \bar{A}A$.

Proof: Equivalence of (a) and (b) follows from the preceding theorem: $\bar{A}A = A\bar{A}$ if and only if $A^T A = (\bar{A})^* A = A(\bar{A})^* = AA^T$.

Suppose that $\bar{A}A = A\bar{A}$ and write $A = B + iC$, in which $B, C \in M_n(\mathbb{R})$. Then $B = (A + \bar{A})/2$ and $C = (A - \bar{A})/2i$ are real, normal, and commute because the normal matrices A and $\bar{A}$ commute. Suppose that B has β complex conjugate pairs of non-real nonzero eigenvalues and that C has γ complex conjugate pairs of non-real nonzero eigenvalues. If $\beta < \gamma$, (2.5.15) ensures that there is a real orthogonal Q such that

$$\begin{aligned} Q^T BQ &= \Lambda_B \oplus B_1 \oplus \cdots \oplus B_\beta \oplus \hat{B}_1 \oplus \cdots \oplus \hat{B}_{\gamma-\beta} \\ Q^T CQ &= \Lambda_C \oplus C_1 \oplus \cdots \oplus C_\beta \oplus C_{\beta+1} \oplus \cdots \oplus C_\gamma \end{aligned}$$

in which $\Lambda_B, \Lambda_C \in M_{n-2\gamma}$ are real diagonal, each of the 2-by-2 real blocks $B_1, \ldots, B_\beta, C_1, \ldots, C_\gamma$ has the form (2.5.10), and $\hat{B}_j = \mu_j I_2$ for some real μ_j for each $j = 1, \ldots, \gamma - \beta$. Then $Q^T AQ = Q^T(B + iC)Q$ is the direct sum of the complex diagonal matrix $\Lambda_B + i\Lambda_C$ and blocks of the form $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$

for some complex a, b ($a = 0$ and/or $b = 0$ are possible). A similar argument reaches the same conclusion if $\beta = \gamma$ or $\beta > \gamma$. Any nonzero 1-by-1 direct summand of $Q^T AQ$ is a nonzero scalar multiple of the first block in (2.5.18); any 2-by-2 direct summand of $Q^T AQ$ in which $a = 0$ and $b \neq 0$ is a nonzero scalar multiple of the second block; any nonsingular 2-by-2 direct summand of $Q^T AQ$ in which $a \neq 0 \neq b$ is a nonzero scalar multiple of the third block; and any singular nonzero 2-by-2 direct summand of $Q^T AQ$ is a nonzero scalar multiple of the fourth block or its complex conjugate. $\square$

Two special cases of the preceding theorem play an important role in the next section: unitary matrices that are either symmetric or skew-symmetric.

Exercise. Show that: the first two blocks in (2.5.18) are unitary; the third block is complex orthogonal but not unitary; the fourth and fifths blocks are neither unitary nor complex orthogonal.

2.5.18 Corollary. Let $U \in M_n$ be unitary. If $U = U^T$, then there are real scalars $\theta_1, \ldots, \theta_n \in \mathbf{R}$ and a real orthogonal $Q \in M_n(\mathbf{R})$ such that

$$U = Q \begin{bmatrix} e^{i\theta_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & e^{i\theta_n} \end{bmatrix} Q^T \tag{2.5.19.1}$$

If $U = -U^T$, then n is even and there are real scalars $\theta_1, \ldots, \theta_{n/2} \in \mathbf{R}$ and a real orthogonal $Q \in M_n(\mathbf{R})$ such that

$$U = Q \left(e^{i\theta_1} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \oplus \cdots \oplus e^{i\theta_{n/2}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right) Q^T \tag{2.5.19.2}$$

Conversely, any matrix of the form (2.5.19.1) is unitary and symmetric; any matrix of the form (2.5.19.2) is unitary and skew symmetric.

Proof: Suppose that U is symmetric. In the representation described in the preceding theorem, only symmetric unitary blocks can appear, so $Q^T UQ$ is a direct sum of blocks of the form $c[1]$, in which $|c| = 1$.

Now suppose that U is skew symmetric. In the representation described in the preceding theorem, only unitary skew-symmetric blocks can appear, so the only blocks that can occur are of the form $c\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ in which $|c| = 1$. In particular, n must be even. $\square$

Problems

1. Show that $A \in M_n$ is normal if and only if $(Ax)^*(Ax) = (A^*x)^*(A^*x)$ for all $x \in \mathbf{C}^n$, that is, $\|Ax\|_2 = \|A^*x\|_2$ for all $x \in \mathbf{C}^n$.

2. Show that a normal matrix is unitary if and only if all its eigenvalues have absolute value 1.

3. Show that a normal matrix is Hermitian if and only if all its eigenvalues are real.

4. Show that a normal matrix is skew-Hermitian if and only if all its eigenvalues are pure imaginary (have real part equal to 0).

5. If $A \in M_n$ is skew Hermitian (respectively Hermitian), show that *iA* is Hermitian (respectively skew Hermitian).

6. Show that $A \in M_n$ is normal if and only if it commutes with some normal matrix with distinct eigenvalues.

7. Consider matrices $A \in M_n$ of the form $A = B^{-1}B^*$ for a nonsingular $B \in M_n$, as in (2.1.9). (a) Show that A is unitary if and only if B is normal. (b) If B has the form $B = HNH$, in which N is normal and H is Hermitian (and both are nonsingular), show that A is similar to a unitary matrix.

8. Write $A \in M_n$ as $A = H(A) + iK(A)$ in which $H(A)$ and $K(A)$ are Hermitian. Show that A is normal if and only if $H(A)$ and $K(A)$ commute.

9. Write $A \in M_n$ as $A = H(A) + iK(A)$ in which $H(A)$ and $K(A)$ are Hermitian. (0.2.5) If every eigenvector of $H(A)$ is an eigenvector of $K(A)$, show that A is normal. What about the converse? Consider $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$.

10. Suppose $A, B \in M_n$ are both normal. If A and B commute, show that AB and $A \pm B$ are all normal. What about the converse? Verify that $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$, AB, and BA are all normal, but A and B do not commute.

11. For any complex number $z \in \mathbf{C}$, show that there are $\theta, \tau \in \mathbf{R}$ such that $\bar{z} = e^{i\theta}z$ and $|z| = e^{i\tau}z$. Notice that $[e^{i\theta}] \in M_1$ is a unitary matrix. What do diagonal unitary matrices $U \in M_n$ look like?

12. Generalize Problem 11 to show that if $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in M_n$, then there are diagonal unitary matrices U and V such that $\bar{\Lambda} = U\Lambda = \Lambda U$ and $|\Lambda| = \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_n|) = V\Lambda = \Lambda V$.

13. Use Problem 12 to show that $A \in M_n$ is normal if and only if there is a unitary $V \in M_n$ such that $A^* = AV$.

14. Let $A \in M_n(\mathbf{R})$ be given. Explain why A is normal and all its eigenvalues are real if and only if A is symmetric.

15. Show that two normal matrices are similar if and only if they have the same characteristic polynomial. Is this true if we omit the assumption that both matrices are normal? Consider $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

16. If $U, V, \Lambda \in M_n$ and U, V are unitary, show that $U\Lambda U^*$ and $V\Lambda V^*$ are unitarily similar. Deduce that two normal matrices are similar if and only if they are unitarily similar. Give an example of two diagonalizable matrices that are similar but not unitarily similar.

17. If $A \in M_n$ is normal and $p(t)$ is a given polynomial, use (2.5.1) to show that $p(A)$ is normal. Give another proof of this fact using (2.5.4).

18. If $A \in M_n$ and there is nonzero polynomial $p(t)$ such that $p(A)$ is normal, does it follow that A is normal? *Hint:* Consider $A = \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$ and A^2.

19. Let $A \in M_n$ and $a \in \mathbf{C}$ be given. Show that A is normal if and only if $A + aI$ is normal.

20. Let $A \in M_n$ be normal and suppose $x \in \mathbf{C}^n$ is a right eigenvector of A corresponding to the eigenvalue λ. Use Problems 1 and 19 to show that x is a left eigenvector of A corresponding to the same eigenvalue λ. *Hint:* $\|(A - \lambda I)x\|_2 = \|(A - \lambda I)^* x\|_2$.

21. Suppose $A \in M_n$ is normal. Use the preceding problem to show that $Ax = 0$ if and only if $A^*x = 0$, that is, the null space of A is the same as that of A^*. Consider $B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$ to show that the null space of a non-normal matrix B need not be the same as the null space B^*, even if B is diagonalizable.

22. Use (2.5.6) to show that the characteristic polynomial of a complex Hermitian matrix has real coefficients.

23. Show that $\begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}$ and $\begin{bmatrix} i & i \\ i & -1 \end{bmatrix}$ are both symmetric, but one is normal and the other is not. This is an important difference between real symmetric matrices and complex symmetric matrices.

24. If $A \in M_n$ is both normal and nilpotent, show that $A = 0$.

25. Suppose $A \in M_n$ and $B \in M_m$ are normal and let $X \in M_{n,m}$ be given. Explain why $\bar{B}$ is normal and deduce that $AX = X\bar{B}$ if and only if $A^*X = XB^T$. *Hint:* (2.5.16)

26. Let $A \in M_n$ be given. (a) If there is a polynomial $p(t)$ such that $A^* = p(A)$, show that $A \in M_n$ is normal. (b) If A is normal, show that there is a polynomial $p(t)$ of degree at most $n - 1$ such that $A^* = p(A)$. (c) If A

is real and normal, show that there is a polynomial $p(t)$ with real coefficients and degree at most $n-1$ such that $A^T = p(A)$. (d) If A is normal, show that there is a polynomial $p(t)$ with real coefficients and degree at most $2n-1$ such that $A^* = p(A)$. (e) If A is normal and $B \in M_m$ is normal, show that there is a polynomial $p(t)$ of degree at most $n+m-1$ such that $A^* = p(A)$ and $B^* = p(B)$. (f) If A is normal and $B \in M_m$ is normal, show that there is a polynomial $p(t)$ with real coefficients and degree at most $2n+2m-1$ such that $A^* = p(A)$ and $B^* = p(B)$. (g) Use (e) to prove (2.5.16). (h) Use (f) to prove the assertion in Problem 25. *Hint:* These are all classical polynomial interpolation problems. Look carefully at (0.9.11.4).

27. Let $A, B \in M_{n,m}$ be given. (a) If AB^* and B^*A are both normal, show that $BA^*A = AA^*B$. *Hint:* (2.5.16). $(AB^*)A = A(B^*A)$. (b) Suppose that $n = m$. Prove that $A\bar{A}$ is normal (such a matrix is called *congruence normal*) if and only if $AA^*A^T = A^TA^*A$. *Hint:* $AA^*A^T = A(A\bar{A})^T$ and $A^TA^*A = (A\bar{A})^*A$.

28. Let Hermitian matrices $A, B \in M_n$ be given and assume that AB is normal. (a) Why is BA normal? (b) Show that A commutes with B^2 and B commutes with A^2. *Hint:* $(AB)A = A(BA)$. (2.5.16). (c) If there is a polynomial $p(t)$ such that *either* $A = p(A^2)$ or $B = p(B^2)$, show that A commutes with B and AB is actually Hermitian. (d) Explain why the condition in (c) is met if *either* A or B has the property that whenever λ is a nonzero eigenvalue, then $-\lambda$ is not also an eigenvalue. For example, if either A or B has all nonnegative eigenvalues, this condition is met. (d) Discuss the example $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & i \\ -i & 0 \end{bmatrix}$.

29. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2$ and assume that $bc \neq 0$. (a) Show that A is normal if and only if there is some $\theta \in \mathbf{R}$ such that $c = e^{i\theta}b$ and $a - d = e^{i\theta}b(\bar{a} - \bar{d})/\bar{b}$. In particular, if A is normal it is necessary that $|c| = |b|$. (b) If A is real, deduce from part (a) that it is normal if and only if either $c = b$ ($A = A^T$) or $c = -b$ and $a = d$ ($AA^T = (a^2 + b^2)I$ and $A = -A^T$ if $a = 0$).

30. Show that a given $A \in M_n$ is normal if and only if $(Ax)^*(Ay) = (A^*x)^*(A^*y)$ for all $x, y \in \mathbf{C}^n$, that is, for all x and y, the angle between Ax and Ay is the same as the angle between A^*x and A^*y. Compare with Problem 1. What does this condition say if we take $x = e_i$ and $y = e_j$ (the standard Euclidean basis vectors)? If $(Ae_i)^*(Ae_j) = (A^*e_i)^*(A^*e_j)$ for all $i, j = 1, \ldots, n$, show that A is normal.

31. Let $A \in M_n(\mathbf{R})$ be a real normal matrix, that is, $AA^T = A^TA$. If AA^T has n distinct eigenvalues, show that A is symmetric. *Hint:* Use (2.5.7).

32. If $A \in M_3(\mathbf{R})$ is real orthogonal, observe that A has either one or three real eigenvalues. If it has a positive determinant, use (2.5.11) to show that it is orthogonally similar to the direct sum of $[1] \in M_1$ and a plane rotation. Discuss the geometrical interpretation of this as a rotation by an angle θ around some fixed axis passing through the origin in $\mathbf{R}^3$. This is part of Euler's theorem in mechanics: Every motion of a rigid body is the composition of a translation and a rotation about some axis.

33. If $\mathcal{F} \subseteq M_n$ is a commuting family of normal matrices, show that there exists a single Hermitian matrix B such that for each $A_\alpha \in \mathcal{F}$ there is a polynomial $p_\alpha(t)$ of degree at most $n-1$ such that $A_\alpha = p_\alpha(B)$. Notice that B is fixed for all of $\mathcal{F}$ but the polynomial may depend on the element of $\mathcal{F}$. *Hint:* Let $U \in M_n$ be a unitary matrix that simultaneously diagonalizes every member of $\mathcal{F}$, let $B = U\mathrm{diag}(1,2,\ldots,n)U^*$, let $A_\alpha = U\Lambda_\alpha U^*$ with $\Lambda_\alpha = \mathrm{diag}(\lambda_1^{(\alpha)},\ldots,\lambda_n^{(\alpha)})$, and take $p_\alpha(t)$ to be the Lagrange interpolation polynomial such that $p_\alpha(k) = \lambda_k^{(\alpha)}, k = 1,2,\ldots,n$.

34. Let $A \in M_n$. We say that x is a *normal eigenvector* of A if it is both a right and left eigenvector of A (necessarily associated with the same eigenvalue). (a) If x is a normal eigenvector of A associated with an eigenvalue λ, show that A is unitarily similar to $[\lambda] \oplus A_1$, in which $A_1 \in M_{n-1}$ is upper triangular. (b) Show that A is normal if and only if every eigenvector of A is a normal eigenvector. *Hint:* Let $U_1 \in M_n$ be a unitary matrix whose first column is an eigenvector of A and of A^*. Inspect the first row of $U_1^*AU_1$ in the proof of (2.3.1) and continue.

35. Let $x,y \in \mathbf{C}^n$ be given. Show that $xx^* = yy^*$ if and only if there is some real θ such that $x = e^{i\theta}y$. *Hint*: If $xx^* = yy^*$ and $x_k \neq 0$ then $x_j = (\overline{y_k/x_k})y_j$ for all $j = 1,\ldots,n$.

36. For any $A \in M_n$, show that $\begin{bmatrix} A & A^* \\ A^* & A \end{bmatrix} \in M_{2n}$ is normal. Thus, any square matrix can be a principal submatrix of a normal matrix. Can any square matrix be a principal submatrix of a Hermitian matrix? of a unitary matrix?

37. Let $n \geq 2$ and suppose $A = \begin{bmatrix} a & x^* \\ y & B \end{bmatrix} \in M_n$ is normal with $B \in M_{n-1}$ and $x,y \in \mathbf{C}^{n-1}$. (a) Show that $\|x\|_2 = \|y\|_2$ and $xx^* - yy^* = BB^* - B^*B$. (b) Explain why $\mathrm{rank}(FF^* - F^*F) \neq 1$ for every square complex matrix F. *Hint*: Problem 14 in (1.3). (c) Explain why there are two mutually exclusive possibilities: Either (i) the principal submatrix B is normal or (ii) $\mathrm{rank}(BB^* - B^*B) = 2$. (d) Explain why B is normal if and only if $x = e^{i\theta}y$ for some real θ. *Hint*: Problem 35. (e) Discuss the example $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, x = [-\sqrt{2}\ \ 1]^T, y = [1\ \ -\sqrt{2}]^T, a = 1-\sqrt{2}$.

38. Let $A \in M_n$ and let $C = AA^* - A^*A$. Explain why C is Hermitian. Conclude that C is nilpotent if and only if $C = 0$. Use Problem 12 in (2.4) to prove that A is normal if and only if it commutes with C.

39. Suppose that $U \in M_n$ is unitary, so all its eigenvalues have modulus one. (a) If U is symmetric, show that its eigenvalues uniquely determine its representation (2.5.19.1), up to permutation of the diagonal entries. (b) If U is skew symmetric, explain how the scalars $e^{i\theta_j}$ in (2.5.19.2) are related to its eigenvalues. Why must the eigenvalues of U occur in $\pm$ pairs? Show that the eigenvalues of U uniquely determine its representation (2.5.19.2), up to permutation of direct summands.

40. Let $A = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix} \in M_4$, in which $B = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$. Verify that A commutes with A^T, A commutes with $\bar{A}$, but A does not commute with A^*, that is, A is not normal.

41. Let $z \in \mathbf{C}^n$ be nonzero and write $z = x + iy$ with $x, y \in \mathbf{R}^n$. (a) Show that the following three statements are equivalent: (1) $\{z, \bar{z}\}$ is linearly dependent; (2) $\{x, y\}$ is linearly dependent; (3) There is a unit vector $u \in \mathbf{R}^n$ and a nonzero $c \in \mathbf{C}$ such that $z = cu$. (b) Show that the following are equivalent: (1) $\{z, \bar{z}\}$ is linearly independent; (2) $\{x, y\}$ is linearly independent; (3) There are real orthonormal vectors $v, w \in \mathbf{R}^n$ such that $\operatorname{span}\{z, \bar{z}\} = \operatorname{span}\{v, w\}$ (over $\mathbf{C}$).

42. For $A \in M_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$, the function $\Delta(A) = \operatorname{tr} A^*A - \sum_{i=1}^n |\lambda_i|^2$ is called the *defect from normality*. The identity (2.3.2a) ensures that $\Delta(A) \geq 0$ and 2.5.3(c) tells us that A is normal if and only if $\Delta(A) = 0$. Let $A, B \in M_n$, suppose A, B, and AB are all normal, and let $\lambda_i(AB), \lambda_i(BA), i = 1, \ldots, n$ be the eigenvalues of AB and BA, respectively. Provide details for the following computation:

$$\begin{aligned} \sum_{i=1}^n |\lambda_i(BA)|^2 &= \sum_{i=1}^n |\lambda_i(AB)|^2 = \operatorname{tr}((AB)^*(AB)) \\ &= \operatorname{tr}(B^*A^*AB) = \operatorname{tr}(A^*B^*BA) = \operatorname{tr}((BA)^*(BA)) \end{aligned}$$

Now explain why BA is normal.

43. Suppose $A \in M_n$ is normal. Show that the main diagonal entries of A are its eigenvalues if and only if A is a diagonal matrix. If $n = 2$ and one of the main diagonal entries of A is an eigenvalue of A, explain why A must be a diagonal matrix.

44. (a) Show that $A \in M_n$ is Hermitian if and only if $\operatorname{tr} A^2 = \operatorname{tr} A^*A$. (b) Show that Hermitian matrices $A, B \in M_n$ commute if and only if $\operatorname{tr}(AB)^2 = \operatorname{tr}(A^2B^2)$. *Hint:* $\operatorname{tr}(A^2B^2) = \operatorname{tr}((AB)(AB)^*)$.

45. Let $\mathcal{N} \subseteq M_n(\mathbf{R})$ be a commuting family of real symmetric matrices. Show that there is a single real orthogonal matrix Q such that Q^TAQ is diagonal for every $A \in \mathcal{N}$.

46. Use (2.3.1) to show that any non-real eigenvalues of a real matrix must occur in complex conjugate pairs. *Hint:* If $A \in M_n(\mathbf{R})$ and $T = U^*AU$ is upper triangular, then $\bar{T} = U^TA\bar{U}$ is unitarily similar to A and hence to T, so the sets of main diagonal entries of T and $\bar{T}$ are identical.

47. Suppose $A \in M_n$ is normal and has eigenvalues $\lambda_1, \dots, \lambda_n$. Show that: (a) $\operatorname{adj} A$ is normal and has eigenvalues $\prod_{j \neq i} \lambda_j, i = 1, \dots, n$. (b) $\operatorname{adj} A$ is Hermitian if A is Hermitian. (c) $\operatorname{adj} A$ has positive (respectively, nonnegative) eigenvalues if A has positive (respectively, nonnegative) eigenvalues. (c) $\operatorname{adj} A$ is unitary if A is unitary.

48. Let $A \in M_n$ be normal, suppose $\operatorname{rank} A = r$, and suppose the first r rows of A are linearly independent. Let $A = U\Lambda U^*$, in which $\Lambda = \Lambda_1 \oplus 0_{n-r}$ and $\Lambda_1 \in M_r$ is a nonsingular diagonal matrix. Partition $A = [A_{ij}]_{i,j=1}^2$ and $U = [U_{ij}]_{i,j=1}^2$ conformal to Λ. (a) Show that $[A_{11}\ A_{12}] = U_{11}\Lambda_1[U_{11}^*\ U_{21}^*]$ and conclude that U_{11} is nonsingular. (b) Explain why every normal matrix is rank principal. (0.7.6) (c) Consider $\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$ and explain why the hypothesis of normality in (b) may not be weakened to diagonalizability.

49. Suppose $A \in M_n$ is upper triangular and diagonalizable. Show that it can be diagonalized via an upper triangular similarity. *Hint:* If $A = S\Lambda S^{-1}$, let $S = RQ$ be an RQ factorization. Then $R^{-1}AR$ is normal.

50. The reversal matrix K_n (0.9.5.1) is real symmetric. Check that it is also real orthogonal, and explain why its eigenvalues can only be ± 1. Check that $\operatorname{tr} K_n = 0$ if n is even and $\operatorname{tr} K_n = 1$ if n is odd. Explain why: if n is even, the eigenvalues of K_n are ± 1, each with multiplicity $n/2$; if n is odd, the eigenvalues of K_n are $+1$ with multiplicity $(n+1)/2$ and -1 with multiplicity $(n-1)/2$.

51. Let $A \in M_n$ be normal, let $A = U\Lambda U^*$ be a spectral decomposition in which $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, let $x \in \mathbf{C}^n$ be any given unit vector, and let $\xi = [\xi_i] = Ux$. Explain why $x^*Ax = \sum_{i=1}^n |\xi_i|^2\lambda_i$, why the x^*Ax lies in the convex hull of the eigenvalues of A, and why each complex number in the convex hull of the eigenvalues of A equals x^*Ax for some unit vector x. Thus, if A is normal, $x^*Ax \neq 0$ for every unit vector x if and only if 0 is not in the convex hull of the eigenvalues of A.

52. Let $A, B \in M_n$ be nonsingular. The matrix $C = ABA^{-1}B^{-1}$ is the *multiplicative commutator* of A and B. Explain why $C = I$ if and only if A commutes with B. Suppose that A and C are normal and 0 is not in the convex hull of the eigenvalues of B. Provide details for the following sketch of a proof that A commutes with C if and only if A commutes with B (this is the *Marcus-Thompson theorem*): Let $A = U\Lambda U^*$ and $C = UMU^*$ be spectral decompositions in which $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $M = \operatorname{diag}(\mu_1, \ldots, \mu_n)$. Let $\mathcal{B} = U^*BU = [\beta_{ij}]$. Then all $\beta_{ii} \neq 0$ and $M = U^*CU = \Lambda\mathcal{B}\Lambda^{-1}\mathcal{B}^{-1} \Rightarrow M\mathcal{B} = \Lambda\mathcal{B}\Lambda^{-1} \Rightarrow \mu_i\beta_{ii} = \beta_{ii} \Rightarrow M = I \Rightarrow C = I$. Compare with Problem 12(c) in (2.4).

53. Let $U, V \in M_n$ be unitary and suppose that all of the eigenvalues of V lie on an open arc of the unit circle of length π; such a matrix is called a *cramped unitary matrix*. Let $C = UVU^*V^*$ be the multiplicative commutator of U and V. Use the preceding problem to prove *Frobenius's theorem*: U commutes with C if and only if U commutes with V.

54. If $A, B \in M_n$ are normal, show that: (a) The null space of A is orthogonal to the range of A. (b) The null space of A is contained in the null space of B if and only if the range of A contains the range of B.

55. Verify the following improvement of (2.2.8) for normal matrices: If $A, B \in M_n$ are normal, then A is unitarily similar to B if and only if $\operatorname{tr} A^k = \operatorname{tr} B^k$, $k = 1, 2, \ldots, n$. *Hint:* Problem 15, and Problem 10 in (2.4).

56. Let $A \in M_n$ and an integer $k \geq 2$ be given, and let $\omega = e^{2\pi i/(k+1)}$. Show that $A^k = A^*$ if and only if A is normal and its spectrum is contained in the set $\{0, 1, \omega, \omega^2, \ldots, \omega^k\}$. If $A^k = A^*$ and A is nonsingular, explain why it is unitary. *Hint*: Problem 26.

57. Let $A \in M_n$ be given. Use (2.5.17) to show that A is normal and symmetric if and only if there is a real orthogonal $Q \in M_n$ and a diagonal $\Lambda \in M_n$ such that $A = Q\Lambda Q^T$. *Hint*: Which of the blocks in (2.5.18) are symmetric?

58. Let $A \in M_n$ be normal. Then $A\bar{A} = 0$ if and only if $AA^T = A^TA = 0$. (a) Use (2.5.17) to prove this. (b) Provide details for an alternative proof: $A\bar{A} = 0 \Rightarrow 0 = A^*A\bar{A} = AA^*\bar{A} \Rightarrow \bar{A}A^TA = 0 \Rightarrow (A^TA)^*(A^TA) = 0 \Rightarrow A^TA = 0$ (0.2.5.1).

Further readings. For a discussion of 89 characterizations of normality, see R. Grone, C. Johnson, E. Sa, and H. Wolkowicz, Normal Matrices, *Linear Al-*

gebra Appl. 87 (1987) 213-225 as well as L. Elsner and Kh. Ikramov, Normal Matrices: An Update, *Linear Algebra Appl.* 285 (1998) 291-303.

2.6 Unitary equivalence and the singular value decomposition

Suppose that a given matrix A is the basis representation of a linear transformation $T : V \to V$ on an n-dimensional complex vector space, with respect to a given orthonormal basis. A unitary similarity $A \to UAU^*$ corresponds to changing the basis from the given one to another orthonormal basis; the unitary matrix U is the change of basis matrix.

If $T : V_1 \to V_2$ is a linear transformation from an n-dimensional complex vector space into an m-dimensional one, and if $A \in M_{m,n}$ is its basis representation with respect to given orthonormal bases of V_1 and V_2, then the unitary equivalence $A \to UAW^*$ corresponds to changing the bases of V and W from the given ones to other orthonormal bases.

A unitary equivalence $A \to UAV$ involves *two* unitary matrices that can be selected independently. This additional flexibility permits us to achieve some reductions to special forms that are unattainable with unitary similarity.

In order to ensure that we can reduce $A, B \in M_n$ to upper triangular form by the same unitary similarity, some condition (commutativity, for example) must be imposed on them. However, we can reduce any two given matrices to upper triangular form by the same unitary equivalence.

2.6.1 Theorem. Let $A, B \in M_n$. There are unitary $V, W \in M_n$ such that $A = VT_AW^*$, $B = VT_BW^*$, and T_A, T_B are upper triangular. If B is nonsingular, the main diagonal entries of $T_B^{-1}T_A$ are the eigenvalues of $B^{-1}A$.

Proof: Suppose that B is nonsingular, and use (2.3.1) to write $B^{-1}A = UTU^*$, in which U is unitary and T is upper triangular. Then use the QR factorization (2.1.14) to write $BU = QR$, in which Q is unitary and R is upper triangular. Then $A = BUTU^* = Q(RT)U^*$, RT is upper triangular, and $B = QRU^*$. Moreover, the eigenvalues of $B^{-1}A = UR^{-1}Q^*QRTU^* = UTU^*$ are the main diagonal entries of T.

If both A and B are singular, there is a $\delta > 0$ such that $B_\varepsilon = B + \varepsilon I$ is nonsingular whenever $0 < \varepsilon < \delta$. (1.2.17) For any ε satisfying this constraint, we have shown that there are unitary $V_\varepsilon, W_\varepsilon \in M_n$ such that $V_\varepsilon^* AW_\varepsilon$ and $V_\varepsilon^* BW_\varepsilon$ are both upper triangular. Choose a sequence of nonzero scalars ε_k such that $\varepsilon_k \to 0$ and both $\lim_{k\to\infty} V_\varepsilon = V$ and $\lim_{k\to\infty} W_\varepsilon = W$ exist; each of the limits V and W is unitary. (2.1.8) Then each of $\lim_{k\to\infty} V_\varepsilon^* AW_\varepsilon =$

$V^*AW \equiv T_A$ and $\lim_{k\to\infty} V_\varepsilon^* BW_\varepsilon = V^*BW \equiv T_B$ is upper triangular. We conclude that $A = VT_AW^*$ and $B = VT_BW^*$, as asserted. $\square$

There is also a real version of this theorem, which uses the following fact.

Exercise. Suppose that $A, B \in M_n$, A is upper triangular, and B is upper quasi-triangular. Show that AB is upper quasi-triangular conformal to B.

2.6.2 Theorem. Let $A, B \in M_n(\mathbf{R})$. There are real orthogonal $V, W \in M_n$ such that $A = VT_AW^T$, $B = VT_BW^T$, T_A is real and upper quasi-triangular, and T_B is real and upper triangular.

Proof: If B is nonsingular, one uses (2.3.4) to write $B^{-1}A = UTU^T$, in which U is real orthogonal and T is real and upper quasi-triangular. Use (2.1.14(d)) to write $BU = QR$, in which Q is real orthogonal and R is real and upper triangular. Then RU is upper quasi-triangular, $A = Q(RT)U^T$, and $B = QRU^T$. If both A and B are singular, one can use a real version of the limit argument in the preceding proof. $\square$

Although only square matrices that are normal can be diagonalized by unitary similarity, any complex matrix can be diagonalized by unitary equivalence.

2.6.3 Theorem (Singular Value Decomposition). Let $A \in M_{n,m}$ be given, let $q = \min\{m, n\}$, and suppose $\operatorname{rank} A = r$.
(a) There are unitary matrices $V \in M_n$ and $W \in M_m$, and a square diagonal matrix

$$\Sigma_q = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_q \end{bmatrix}, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_q \tag{2.6.3.1}$$

such that $A = V\Sigma W^*$, in which $\Sigma = \Sigma_q$ if $m = n$,

$$\Sigma = \begin{bmatrix} \Sigma_q & 0 \end{bmatrix} \in M_{n,m} \text{ if } m > n, \text{ and } \Sigma = \begin{bmatrix} \Sigma_q \\ 0 \end{bmatrix} \in M_{n,m} \text{ if } n > m \tag{2.6.3.2}$$

(b) The parameters $\sigma_1, \ldots, \sigma_r$ are the positive square roots of the decreasingly ordered nonzero eigenvalues of AA^*, which are the same as the decreasingly ordered nonzero eigenvalues of A^*A.

Proof: First suppose that $m = n$. The Hermitian matrices $AA^* \in M_n$ and

$A^*A \in M_n$ have the same eigenvalues (1.3.22), so they are unitarily similar (2.5.4(d)) and hence there is a unitary U such that $A^*A = U(AA^*)U^*$. Then

$$(UA)^*(UA) = A^*U^*UA = A^*A = UAA^*U^* = (UA)(UA)^*$$

so UA is normal. Let $\lambda_1 = |\lambda_1|e^{i\theta_1}, \ldots, \lambda_n = |\lambda_n|e^{i\theta_n}$ be the eigenvalues of UA, ordered so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Then $r = \operatorname{rank} A = \operatorname{rank} UA$ is the number of nonzero eigenvalues of the normal matrix UA, so $|\lambda_r| > 0$ and $\lambda_{r+1} = \cdots = \lambda_n$. Let $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, let $D = \operatorname{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$, let $\Sigma_q = \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_n|)$, and let X be a unitary matrix such that $UA = X\Lambda X^*$. Then D is unitary and $A = U^*X\Lambda X^* = U^*X\Sigma_q DX^* = (U^*X)\Sigma_q(DX^*)$ exhibits the desired factorization, in which $V = U^*X$ and $W = XD^*$ are unitary, and $\sigma_j = |\lambda_j|, j = 1, \ldots, n$.

Now suppose that $m > n$. Then $r \leq n$, so the null space of A has dimension $m - r \geq m - n$. Let $\{x_1, \ldots, x_{m-n}\}$ be any set of orthonormal vectors in the null space of A, let $X_2 = [x_1 \ , \ldots \ x_{m-n}] \in M_{n,m-n}$, and let $X = [X_1 \ X_2] \in M_m$ be unitary, that is, extend the given orthonormal set to a basis of $\mathbf{C}^m$. Then $AX = [AX_1 \ AX_2] = [AX_1 \ 0]$ and $AX_1 \in M_n$. Using the preceding case, write $AX_1 = V\Sigma_q W^*$, in which $V, W \in M_n$ are unitary and Σ_q is as in (2.6.3.1). This gives

$$A = [AX_1 \ 0]X^* = [V\Sigma_q W^* \ 0]X^* = V[\Sigma_q \ 0]\left(\begin{bmatrix} W^* & 0 \\ 0 & I_{m-n} \end{bmatrix} X^*\right)$$

which is a factorization of the asserted form.

If $n > m$, apply the preceding case to A^*.

Using the factorization $A = V\Sigma W^*$, notice that $\operatorname{rank} A = \operatorname{rank} \Sigma$ since V and W are nonsingular. But $\operatorname{rank} \Sigma$ equals the number of nonzero (and hence positive) diagonal entries of Σ, as asserted. Now compute $AA^* = V\Sigma W^*W\Sigma^T V^* = V\Sigma\Sigma^T V^*$, which is unitarily similar to $\Sigma\Sigma^T$. If $n = m$, then $\Sigma\Sigma^T = \Sigma_q^2 = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. If $m > n$, then $\Sigma\Sigma^T = [\Sigma_q \ 0][\Sigma_q \ 0]^T = \Sigma_q^2 + 0_n = \Sigma_q^2$. Finally, if $n > m$, then

$$\Sigma\Sigma^T = \begin{bmatrix} \Sigma_q \\ 0 \end{bmatrix} \begin{bmatrix} \Sigma_q & 0 \end{bmatrix} = \begin{bmatrix} \Sigma_q^2 & 0 \\ 0 & 0_{n-m} \end{bmatrix}$$

In each case, the nonzero eigenvalues of AA^* are $\sigma_1^2, \ldots, \sigma_r^2$, as asserted. □

The diagonal entries of the matrix Σ in (2.6.3) (that is, the scalars $\sigma_1, \ldots, \sigma_q$) are the *singular values* of A. The *multiplicity* of a singular value σ of A is the multiplicity of σ^2 as an eigenvalue of AA^* or, equivalently, of A^*A. The rank of A is *equal* to the number of its nonzero singular values, while it is *not less than* (and can be greater than) the number of its nonzero eigenvalues.

The singular values of A are uniquely determined by the eigenvalues of A^*A (equivalently, by the eigenvalues of AA^*), so the diagonal factor Σ in the singular value decomposition of A is determined up to permutation of its diagonal entries; a conventional choice that makes Σ unique is to require that the singular values be arranged in non-increasing order, but other choices are possible. The following theorem gives a precise formulation of the assertion that the singular values of a matrix depend continuously on its entries.

2.6.4 Theorem. Let an infinite sequence $A_1, A_2, \ldots \in M_{n,m}$ be given, suppose that $\lim_{k\to\infty} A_k = A$ (entry-wise convergence), and let $q = \min\{m, n\}$. Let $\sigma_1(A) \geq \cdots \geq \sigma_q(A)$ and $\sigma_1(A_k) \geq \cdots \geq \sigma_q(A_k)$ be the non-increasingly ordered singular values of A and A_k, respectively, for $k = 1, 2, \ldots$. Then $\lim_{i\to\infty} \sigma_i(A_k) = \sigma_i(A)$ for each $i = 1, \ldots, q$.

Proof: If the assertion of the theorem is false, then there is some $\varepsilon_0 > 0$ and an infinite sequence of positive integers $k_1 < k_2 < \cdots$ such that for every $j = 1, 2, \ldots$ we have

$$\max_{i=1,\ldots,q} \left|\sigma_i(A_{k_j}) - \sigma_i(A)\right| > \varepsilon_0 \tag{2.6.4.1}$$

For each $j = 1, 2, \ldots$ let $A_{k_j} = V_{k_j} \Sigma_{k_j} W_{k_j}^*$, in which $V_{k_j} \in M_n$ and $W_{k_j} \in M_m$ are unitary and $\Sigma_{k_j} \in M_{n,m}$ is the nonnegative diagonal matrix such that $\operatorname{diag} \Sigma_{k_j} = [\sigma_1(A_{k_j}) \ \ldots \ \sigma_q(A_{k_j})]^T$. Lemma (2.1.8) ensures that there is an infinite sub-subsequence $k_{j_1} < k_{j_2} < \cdots$ and unitary matrices V and W such that $\lim_{\ell\to\infty} V_{k_{j_\ell}} = V$ and $\lim_{\ell\to\infty} W_{k_{j_\ell}} = W$. Then

$$\lim_{\ell\to\infty} \Sigma_{k_{j_\ell}} = \lim_{\ell\to\infty} V_{k_{j_\ell}}^* A_{k_{j_\ell}} W_{k_{j_\ell}} = \left(\lim_{\ell\to\infty} V^*\right)\left(\lim_{\ell\to\infty} A\right)\left(\lim_{\ell\to\infty} W\right) = V^*AW$$

exists and is a nonnegative diagonal matrix with non-increasingly ordered diagonal entries; we denote it by Σ and observe that $A = V\Sigma W^*$. Uniqueness of the singular values of A ensures that $\operatorname{diag}\Sigma = [\sigma_1(A) \ \ldots \ \sigma_q(A)]^T$, contradicts (2.6.4.1), and proves the theorem. $\square$

The unitary factors in a singular value decomposition are never unique. For example, if $A = V\Sigma W^*$, we may replace V by $-V$ and W by $-W$. The following theorem describes in an explicit and very useful fashion how, given one pair of unitary factors, all possible pairs of unitary factors can be obtained.

2.6.5 Theorem. Let $A \in M_{n,m}$ be given with $\operatorname{rank} A = r$. Let $s_1 > \cdots > s_d > 0$ be the decreasingly ordered distinct positive singular values of A, with respective multiplicities $n_1, \ldots, n_d$. Let $A = V\Sigma W^*$ be a given singular value decomposition with Σ as in (2.6.3.1) or (2.6.3.2), so that $\Sigma^T\Sigma = s_1^2 I_{n_1} \oplus \cdots \oplus s_d^2 I_{n_d} \oplus 0_{n-r}$ and $\Sigma\Sigma^T = s_1^2 I_{n_1} \oplus \cdots \oplus s_d^2 I_{n_d} \oplus 0_{m-r}$ (one or both of the zero direct summands are absent if A has full rank). Let $\hat{V} \in M_n$ and $\hat{W} \in M_m$ be given unitary matrices. Then $A = \hat{V}\Sigma\hat{W}^*$ if and only if there are unitary matrices $U_1 \in M_{n_1}, \ldots, U_d \in M_{n_d}, \tilde{V} \in M_{n-r}$, and $\tilde{W} \in M_{m-r}$ such that

$$\hat{V} = V(U_1 \oplus \cdots \oplus U_d \oplus \tilde{V}) \text{ and } \hat{W} = W(U_1 \oplus \cdots \oplus U_d \oplus \tilde{W}) \quad (2.6.5.1)$$

If A is real and the factors $V, W, \hat{V}, \hat{W}$ are real orthogonal, then the matrices $U_1, \ldots, U_d, \tilde{V}, \tilde{W}$ may be taken to be real orthogonal.

Proof: The Hermitian matrix A^*A is represented as $A^*A = (V\Sigma W^*)^*(V\Sigma W^*) = W\Sigma^T\Sigma W^*$ and also as $A^*A = \hat{W}\Sigma^T\Sigma\hat{W}^*$. Theorem 2.5.4 ensures that there are unitary matrices $W_1, \ldots, W_d, W_{d+1}$ with $W_i \in M_{n_i}$ for $i = 1, \ldots, d$ such that $\hat{W} = W(W_1 \oplus \cdots \oplus W_d \oplus W_{d+1})$. We also have $AA^* = V\Sigma\Sigma^T V^* = \hat{V}\Sigma\Sigma^T\hat{V}^*$, so (2.5.4) again tells us that there are unitary matrices $V_1, \ldots, V_d, V_{d+1}$ with $V_i \in M_{n_i}$ for $i = 1, \ldots, d$ such that $\hat{V} = V(V_1 \oplus \cdots \oplus V_d \oplus V_{d+1})$. Since $A = V\Sigma W^* = \hat{V}\Sigma\hat{W}^*$, we have $\Sigma = (V^*\hat{V})\Sigma(\hat{W}^*W)$, that is, $s_i I_{n_i} = V_i(s_i I_{n_i})W_i^*$ for $i = 1, \ldots, d+1$, or $V_i W_i^* = I_{n_i}$ for each $i = 1, \ldots, d$, which means that $V_i = W_i$ for each $i = 1, \ldots, d$. The final assertion about a real A follows from the preceding argument and the exercise following Theorem 2.5.6. $\square$

The singular value decomposition is a very important tool in matrix analysis, with myriad applications in engineering, numerical computation, statistics, image compression, and many other areas; for more details see Chapter 7.

We close this chapter with two application of the preceding uniqueness theorem: singular value decompositions of symmetric or skew-symmetric matrices can be chosen to be unitary congruences, and a real matrix has a singular value decomposition in which all three factors are real.

2.6.6 Corollary. Let $A \in M_n$ and let $r = \operatorname{rank} A$.
(a) (L. Autonne) $A = A^T$ if and only if there is a unitary $U \in M_n$ and a nonnegative diagonal matrix Σ such that $A = U\Sigma U^T$. The diagonal entries of Σ are the singular values of A.
(b) If $A = -A^T$, then r is even and there is a unitary $U \in M_n$ and positive real scalars $s_1, \ldots, s_{r/2}$ such that

$$A = U\left(\begin{bmatrix} 0 & s_1 \\ -s_1 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 0 & s_{r/2} \\ -s_{r/2} & 0 \end{bmatrix} \oplus 0_{n-r}\right)U^T \quad (2.6.6.1)$$

The nonzero singular values of A are $s_1, s_1, \ldots, s_{r/2}, s_{r/2}$. Conversely, any matrix of the form (2.6.6.1) is skew symmetric.

Proof: Let $s_1, \dots, s_d$ be the distinct positive singular values of A, with respective multiplicities $n_1, \dots, n_d$, and let $A = V\Sigma W^*$ be a given singular value decomposition in which $\Sigma = s_1 I_{n_1} \oplus \cdots \oplus s_d I_{n_d} \oplus 0_{n-r}$ and $V, W \in M_n$ are unitary; the zero block is missing if A is nonsingular.
(a) We have $V\Sigma W^* = A = A^T = \bar{W}\Sigma V^T = \bar{W}\Sigma \bar{V}^*$, so the preceding theorem ensures that there are unitary matrices $U_V = U_1 \oplus \cdots \oplus U_d \oplus \tilde{V}$ and $U_W = U_1 \oplus \cdots \oplus U_d \oplus \tilde{W}$, in which $U_i \in M_{n_i}$, $i = 1, \dots, d$, such that $\bar{V} = WU_W$ and $\bar{W} = VU_V$, that is, $U_W = W^*\bar{V}$ and $U_V = V^*\bar{W} = U_W^T$. In particular, $U_j = U_j^T$ for $j = 1, \dots, d$, that is, each U_j is unitary and symmetric. Corollary (2.5.19) tells us that there are real orthogonal matrices Q_j and real parameters $\theta_1^{(j)}, \dots, \theta_{n_j}^{(j)}$ such that $U_j = Q_j \operatorname{diag}(e^{i\theta_1^{(j)}}, \dots, e^{i\theta_{n_j}^{(j)}})Q_j^T$, $j = 1, \dots, d$. For each $j = 1, \dots, d$, let $R_j = Q_j \operatorname{diag}(e^{i\theta_1^{(j)}/2}, \dots, e^{i\theta_{n_j}^{(j)}/2})Q_j^T$ and let $R = R_1 \oplus \cdots \oplus R_d \oplus I_{n-r}$. Then R is symmetric and unitary and $U_V\Sigma = R\Sigma R$, so $A = \bar{W}\Sigma V^T = VU_V\Sigma V^T = VR\Sigma RV^T = (VR)\Sigma(VR)^T$ is a factorization of the asserted form.
(b) Starting with the identity $V\Sigma W^* = -\bar{W}\Sigma V^T = -\bar{W}\Sigma\bar{V}^*$ and proceeding exactly as in (a), we have $\bar{V} = WU_W$, $\bar{W} = -VU_V$, that is, $U_W = W^*\bar{V}$ and $U_V = -V^*\bar{W} = -U_W^T$. In particular, $U_j = -U_j^T$ for $j = 1, \dots, d$, that is, each U_j is unitary and skew symmetric. Corollary (2.5.19) ensures that, for each $j = 1, \dots, d$, n_j is even and there are real orthogonal matrices Q_j and real parameters $\theta_1^{(j)}, \dots, \theta_{n_j/2}^{(j)}$ such that

$$U_j = Q_j \left(e^{i\theta_1^{(j)}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \oplus \cdots \oplus e^{i\theta_{n_j/2}^{(j)}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right) Q_j^T$$

Define the real orthogonal matrix $Q = Q_1 \oplus \cdots \oplus Q_d \oplus I_{n-r}$ and the skew-symmetric matrices

$$S_j = e^{i\theta_1^{(j)}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \oplus \cdots \oplus e^{i\theta_{n_j/2}^{(j)}} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, j = 1, \dots, d$$

Let $S = S_1 \oplus \cdots \oplus S_d \oplus 0_{n-r}$. Then $U_V\Sigma = QSQ^T\Sigma = QS\Sigma Q^T$, so $A = -\bar{W}\Sigma V^T = VU_V\Sigma V^T = VQS\Sigma Q^T V^T = (VQ)S\Sigma(VQ)^T$ is a factorization of the asserted form and $\operatorname{rank} A = n_1 + \cdots + n_d$ is even. $\square$

2.6.7 Corollary. Let $A \in M_{n,m}(\mathbf{R})$ and suppose that $\operatorname{rank} A = r$. Then $A = P\Sigma Q^T$, in which $P \in M_n(\mathbf{R})$ and $Q \in M_m(\mathbf{R})$ are real orthogonal and $\Sigma \in M_{n,m}(\mathbf{R})$ is nonnegative diagonal and has the form (2.6.3.1) or (2.6.3.2).

Proof: Using the notation of (2.6.4), let $A = V\Sigma W^*$ be a given singular value decomposition. We have $V\Sigma W^* = A = \bar{A} = \bar{V}\Sigma\bar{W}^*$ and (2.6.4) ensures that there are unitary matrices $U_V = U_1 \oplus \cdots \oplus U_d \oplus \tilde{V} \in M_n$ and

$U_W = U_1 \oplus \cdots \oplus U_d \oplus \tilde{W} \in M_m$ such that $\bar{V} = VU_V$ and $\bar{W} = WU_W$. Then $U_V = V^*\bar{V} = \bar{V}^T\bar{V}$ and $U_W = \bar{W}^T W$ are unitary and symmetric, so $\tilde{V}$, $\tilde{W}$, and each U_i is unitary and symmetric. Corollary (2.5.18) tells us that there are real orthogonal matrices $Q_1, \ldots, Q_d, Q_{\tilde{V}}, Q_{\tilde{W}}$ and real parameters $\theta_k^{(j)}$ such that $U_j = Q_j \operatorname{diag}(e^{i\theta_1^{(j)}}, \ldots, e^{i\theta_{n-r}^{(j)}})Q_j^T, j = 1, \ldots, d$; $U_{\tilde{V}} = Q_{\tilde{V}} \operatorname{diag}(e^{i\theta_1^{(d+1)}}, \ldots, e^{i\theta_{n_j}^{(d+1)}})Q_{\tilde{V}}^T$; and $U_{\tilde{W}} = Q_{\tilde{W}} \operatorname{diag}(e^{i\theta_1^{(d+2)}}, \ldots, e^{i\theta_{m-r}^{(d+2)}})Q_{\tilde{W}}^T$. For each $j = 1, \ldots, d$, let $R_j = Q_j \operatorname{diag}(e^{i\theta_1^{(j)/}/2}, \ldots, e^{i\theta_{n_j}^{(j)}/2})Q_j^T$; let $R_{\tilde{V}} = Q_{\tilde{V}} \operatorname{diag}(e^{i\theta_1^{(d+1)}/2}, \ldots, e^{i\theta_{n-r}^{(d+1)}/2})Q_{\tilde{V}}^T$; and let $R_{\tilde{W}} = Q_{\tilde{W}} \operatorname{diag}(e^{i\theta_1^{(d+2)}/2}, \ldots, e^{i\theta_{m-r}^{(d+2)}/2})Q_{\tilde{W}}^T$. Finally, let $R_V = R_1 \oplus \cdots \oplus R_d \oplus R_{\tilde{V}}$ and $R_W = R_1 \oplus \cdots \oplus R_d \oplus R_{\tilde{W}}$. Then R_V and R_W are symmetric and unitary, $R_V^{-1} = R_V^* = \overline{R_V}$, $R_W^{-1} = R_W^* = \overline{R_W}$, $R_V^2 = U_V$, $R_W^2 = U_W$, and $R_V \Sigma \overline{R_W} = \Sigma$, so

$$\begin{aligned} A &= \bar{V}\Sigma\bar{W}^* = VU_V\Sigma(WU_W)^* = VR_V^2\Sigma(WR_W^2)^* \\ &= (VR_V)(R_V\Sigma\overline{R_W})(WR_W)^* = (VR_V)\Sigma(WR_W)^* \end{aligned}$$

We conclude the argument by observing that $\bar{V} = VU_V = VR_V^2$ and $\bar{W} = WU_W = WR_W^2$, so $VR_V = \bar{V}R_V^* = \overline{VR_V}$ and $WR_W = \bar{W}R_W^* = \overline{WR_W}$. That is, both VR_V and WR_W are unitary and real, so they are both real orthogonal. $\square$

Problems

1. Let $A \in M_{n,m}$ with $n \geq m$. Show that A has full column rank if and only if all of its singular values are positive.

2. Suppose that $A, B \in M_{n,m}$ can be simultaneously diagonalized by unitary equivalence, that is, suppose that there are unitary matrices $V \in M_n$ and $W \in M_m$ such that each of $V^*AW = \Lambda$ and $V^*BW = M$ is a diagonal matrix (0.9.1). Show that both AB^* and B^*A are normal.

3. Let $A, B \in M_{n,m}$ be given. Show that AB^* and B^*A are both normal if and only if there are unitary matrices $X \in M_n$ and $Y \in M_m$ such that $A = X\Sigma Y^*$, $B = X\Lambda Y^*$, $\Sigma, \Lambda \in M_{n,m}$ are diagonal, and $\Sigma \in M_{n,m}$ has the form (2.6.3.1,2). *Hint*: We may assume that $A = \Sigma$ (write $A = V\Sigma W^*$, so $\Sigma\tilde{B}^*$ and $\tilde{B}^*\Sigma$ are normal, $\tilde{B} = V^*BW$) and $n \leq m$ (if $n > m$, consider B^* and A^*). If ΣB^* and $B^*\Sigma$ are normal, then $\Sigma\Sigma^T B = B\Sigma^T\Sigma$ (Problem 27 in (2.5)). Let $s_1, \ldots, s_d$ be the distinct and decreasingly ordered singular values of A, write $\Sigma_q = s_1 I_{n_1} \oplus \cdots \oplus s_d I_{n_d}$, partition $B = [B^{(1)}\ B^{(2)}]$ with $B^{(1)} \in M_n$, and partition $B^{(1)} = [B_{ij}]_{i,j=1}^d$ conformally to Σ_q. Then $\Sigma\Sigma^T B = B\Sigma^T\Sigma \Rightarrow \Sigma_q^2 B^{(1)} = B_q^{(1)}\Sigma^2$ and $\Sigma_q^2 B^{(2)} = 0$. If $s_d > 0$, then $B^{(2)} = 0$, $B^{(1)} = B_{11} \oplus \cdots \oplus B_{dd}$, and every B_{ii} is normal. If $s_d = 0$, then

$B^{(2)} = \begin{bmatrix} 0 \\ C \end{bmatrix}$, in which C is n_d-by-$(m-n)$, $B^{(1)} = B_{11} \oplus \cdots \oplus B_{dd}$, and each $B_{11}, \ldots, B_{d-1,d-1}$ is normal; replace each normal B_{ii} with its spectral decomposition and replace $[B_{dd}\ C]$ with its singular value decomposition.

4. Let $A, B \in M_{n,m}$ be given. (a) Show that AB^* and B^*A are both Hermitian if and only if there are unitary matrices $X \in M_n$ and $Y \in M_m$ such that $A = X\Sigma Y^*$, $B = X\Lambda Y^*$, $\Sigma, \Lambda \in M_{n,m}(\mathbf{R})$ are diagonal, and Σ has the form (2.6.3.1,2). (b) If A and B are real, show that AB^T and B^TA are both real symmetric if and only if there are real orthogonal matrices $X \in M_n(\mathbf{R})$ and $Y \in M_m(\mathbf{R})$ such that $A = X\Sigma Y^T$, $B = X\Lambda Y^T$, $\Sigma, \Lambda \in M_{n,m}(\mathbf{R})$ are diagonal, and Σ has the form (2.6.3.1,2). (c) In both (a) and (b), show that Λ can be chosen to have nonnegative diagonal entries if and only if all the eigenvalues of AB^* and B^*A are nonnegative.

5. Let $A \in M_{n,m}$ be given and write $A = B + iC$, in which $B, C \in M_{n,m}(\mathbf{R})$. Show that there are real orthogonal matrices $X \in M_n(\mathbf{R})$ and $Y \in M_m(\mathbf{R})$ such that $A = X\Lambda Y^T$ and $\Lambda \in M_{n,m}(\mathbf{C})$ is diagonal if and only if both BC^T and C^TB are real symmetric.

6. Let $A \in M_n$ be given and let $A = QR$ be a QR factorization (2.1.14). (a) Explain why QR is normal if and only if RQ is normal. (b) Show that A is normal if and only if Q and R^* can be simultaneously diagonalized by unitary equivalence.

7. Show that two complex matrices of the same size are unitarily equivalent if and only if they have the same singular values.

8. Let $A \in M_{n,k}$ and $B \in M_{k,m}$ be given. Use the singular value decomposition to show that $\operatorname{rank} AB \le \min\{\operatorname{rank} A, \operatorname{rank} B\}$. *Hint:* Let $A = V\Sigma W^*$. Then $\operatorname{rank} AB = \operatorname{rank} \Sigma W^*B$, and ΣW^*B has at most $\operatorname{rank} A$ nonzero rows.

9. Let $A \in M_n$ be given. Suppose $\operatorname{rank} A = r$, form $\Sigma_1 = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$ from its decreasingly ordered positive singular values, and let $\Sigma = \Sigma_1 \oplus 0_{n-r}$.. Suppose $W \in M_n$ is unitary and $A^*A = W\Sigma^2W^*$. Show that there is a unitary $V \in M_n$ such that $A = V\Sigma W^*$. *Hint:* Let $D = \Sigma_1 \oplus I_{n-r}$, show that $(AWD^{-1})^*(AWD^{-1}) = I_r \oplus 0_{n-r}$, and conclude that $AWD^{-1} = [V_1\ 0_{n,n-r}]$, in which V_1 has orthonormal columns. Let $V = [V_1\ V_2]$ be unitary.

10. Let $A, B \in M_n$ be given, let $\sigma_1 \ge \cdots \ge \sigma_n \ge 0$ be the singular values of A, and let $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$. Show that the following three statements are equivalent: (a) $A^*A = B^*B$. (b) There are unitary matrices $W, X, Y \in M_n$ such that $A = X\Sigma W^*$ and $B = Y\Sigma W^*$. (c) There is a unitary $U \in M_n$ such that $A = UB$.

11. Let $A \in M_{n,m}$ and a normal $B \in M_m$ be given. Show that A^*A commutes with B if and only if there are unitary matrices $V \in M_n$ and $W \in M_m$, and diagonal matrices $\Sigma \in M_{n,m}$ and $\Lambda \in M_m$, such that $A = V\Sigma W^*$ and $B = W\Lambda W^*$. *Hint:* (2.5.5) and Problem 9.

12. Let $A \in M_n$ have a singular value decomposition $A = V\Sigma W^*$, in which $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \cdots \geq \sigma_n$. (a) Show that $\operatorname{adj} A$ has a singular value decomposition $\operatorname{adj} A = X^*SY$ in which $X = (\det W)(\operatorname{adj} W)$, $Y = (\det V)(\operatorname{adj} V)$, and $S = \operatorname{diag}(s_1, \ldots, s_n)$, in which each $s_i = \prod_{j \neq i} \sigma_j$. (b) Use (a) to explain why $\operatorname{adj} A = 0$ if $\operatorname{rank} A \leq n - 2$. (c) If $\operatorname{rank} A = n - 1$ and $v_n, w_n \in \mathbf{C}^n$ are the last columns of V and W, respectively, show that $\operatorname{adj} A = \sigma_1 \cdots \sigma_{n-1} e^{i\theta} w_n v_n^*$, in which $\det(VW^*) = e^{i\theta}$, $\theta \in \mathbf{R}$.

13. Let $A \in M_n$ and let $A = V\Sigma W^*$ be a singular value decomposition. (a) Show that A is unitary if and only if $\Sigma = I$. (b) Show that A is a scalar multiple of a unitary matrix if and only if Ax is orthogonal to Ay whenever $x, y \in \mathbf{C}^n$ are orthogonal. *Hint:* It suffices to consider only the case in which $A = \Sigma$.

14. Suppose $A \in M_n$ is normal and has spectral decomposition $A = U\Lambda U^*$, in which U is unitary, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \operatorname{diag}(e^{i\theta_1}|\lambda_1|, \ldots, e^{i\theta_n}|\lambda_n|)$, and the eigenvalues are ordered so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Let $D = \operatorname{diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})$ and $\Sigma = \operatorname{diag}(|\lambda_1|, \ldots, |\lambda_n|)$. Explain why $A = (UD)\Sigma U^*$ is a singular value decomposition of A and why the singular values of A are exactly the absolute values of its eigenvalues.

15. Let $A = [a_{ij}] \in M_n$ have eigenvalues $\lambda_1, \ldots, \lambda_n$ ordered so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$ and singular values $\sigma_1, \ldots, \sigma_n$ ordered so that $\sigma_1 \geq \cdots \geq \sigma_n$. Show that: (a) $\sum_{i,j=1}^n |a_{ij}|^2 = \operatorname{tr} A^*A = \sum_{i=1}^n \sigma_i^2$. (b) $\sum_{i=1}^n \sigma_i^2 \geq \sum_{i=1}^n |\lambda_i|^2$ with equality if and only if A is normal. (c) $\sigma_i = |\lambda_i|$ for all $i = 1, \ldots, n$ if and only if A is normal. (d) If $|a_{ii}| = \sigma_i$ for all $i = 1, \ldots, n$, then A is diagonal. (e) If A is normal and $|a_{ii}| = |\lambda_i|$ for all $i = 1, \ldots, n$, then A is diagonal. *Hint:* Problem 42 in (4.5).

16. Let $U, V \in M_n$ be unitary. (a) Show that there are always unitary $X, Y \in M_n$ and a diagonal unitary $D \in M_n$ such that $U = XDY$ and $V = Y^*DX^*$. (b) Explain why the unitary equivalence map $A \to UAV = XDYAY^*DX^*$ on M_n is the composition of a unitary similarity, a diagonal unitary congruence, and a unitary similarity. *Hint:* Problem 3.

17. Let $A \in M_{n,m}$. Use the singular value decomposition to explain why $\operatorname{rank} A = \operatorname{rank} AA^* = \operatorname{rank} A^*A$.

18. Let $A \in M_n$ be idempotent and suppose that $\operatorname{rank} A = r$. (a) Show that A is unitarily similar to $\left[\begin{smallmatrix} I_r & X \\ 0 & 0_{n-r} \end{smallmatrix}\right]$. (Problem 5 in (1.1)) (b) Let $X = V\Sigma W^*$ be a singular value decomposition. Show that A is unitarily similar to $\left[\begin{smallmatrix} I_r & \Sigma \\ 0 & 0_{n-r} \end{smallmatrix}\right]$ via $V \oplus W$, and hence the singular values of A are the diagonal entries of $(I_r + \Sigma\Sigma^T) \oplus 0_{n-r}$; let $\sigma_1, \ldots, \sigma_g$ be the singular values of A that are greater than 1. (c) Show that A is unitarily similar to $0_{n-r-g} \oplus I_{r-g} \oplus \left[\begin{smallmatrix} 1 & (\sigma_1^2-1)^{1/2} \\ 0 & 0 \end{smallmatrix}\right] \oplus \cdots \oplus \left[\begin{smallmatrix} 1 & (\sigma_g^2-1)^{1/2} \\ 0 & 0 \end{smallmatrix}\right]$.

19. Let $U = \left[\begin{smallmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{smallmatrix}\right] \in M_{k+\ell}$ be unitary with $U_{11} \in M_k$, $U_{22} \in M_\ell$, and $k \le \ell$. Let $\sigma_1(X) \ge \sigma_2(X) \ge \cdots$ denote the non-increasingly ordered singular values of a matrix X. Show that

$$\sigma_i(U_{11}) = \sigma_i(U_{22}) \text{ and } \sigma_i(U_{12}) = \sigma_i(U_{21}) = (1-\sigma_i^2(U_{11}))^{1/2}, i = 1, \ldots, k$$

and $\sigma_i(U_{22}) = 1, \quad i = k+1, \ldots, \ell$. In particular, $|\det U_{11}| = |\det U_{22}|$ and $\det U_{12}U_{12}^* = \det U_{21}^* U_{21}$. Explain why these results imply Lemma 2.1.10. *Hint:* Write out the identities $U^*U = I$ and $UU^* = I$ as block matrices and use (1.3.22). $U_{11}U_{11}^* + U_{12}U_{12}^* = I \Rightarrow U_{11}U_{11}^* = I - U_{12}U_{12}^* \Rightarrow \sigma_i^2(U_{11}) = 1 - \sigma_i^2(U_{12})$.

20. Let $A \in M_n$ be symmetric. Suppose that the special singular value decomposition in Corollary (2.6.6a) is known if A is *nonsingular*. Provide complete details for the following two approaches to showing that it is valid even if A is singular. (a) Consider $A_\varepsilon = A + \varepsilon I$; use (2.1.8) and (2.6.4). (b) Let the columns of $U_1 \in M_{n,\nu}$ be an orthonormal basis for the null space of A and let $U = [U_1\ U_2] \in M_n$ be unitary. Let $U^T A U = [A_{ij}]_{i,j=1}^2$ (partitioned conformally to U). Explain why A_{11}, A_{12}, and A_{21} are zero matrices, while A_{22} is nonsingular and symmetric.

21. Let $A, B \in M_n$ be symmetric. Show that $A\bar{B}$ is normal if and only if there is a unitary $U \in M_n$ such that $A = U\Sigma U^T$, $B = U\Lambda U^T$, $\Sigma, \Lambda \in M_n$ are diagonal, and the diagonal entries of Σ are nonnegative. *Hint*: (cf. Problem 3) If $A\bar{B} = AB^*$ is normal, then $(A\bar{B})^T = \bar{B}A = B^*A$ is normal. We may take $A = \Sigma$ (use (2.6.6a) to write $A = U\Sigma U^T$, so $\Sigma\bar{\tilde{B}}$ is normal and $\tilde{B} = U^*B\bar{U}$ is symmetric). If $\Sigma\bar{B}$ and $\bar{B}\Sigma$ are normal and B is symmetric, then $\Sigma^2 B = B\Sigma^2$ (Problem 27 in (2.5)). Write $\Sigma = s_1 I_{n_1} \oplus \cdots \oplus s_d I_{n_d}$, in which $s_1 > \cdots > s_d \ge 0$ and partition $B = [B_{ij}]_{i,j=1}^d$ conformally to Σ. Then $\Sigma^2 B = B\Sigma^2 \Rightarrow B = B_{11} \oplus \cdots \oplus B_{dd}$ and each B_{ii} is symmetric; B_{ii} is also normal if $s_i > 0$. If $s_i > 0$, replace B_{ii} with $Q_i\Lambda_i Q_i^T$, in which Q_i is real orthogonal and Λ_i is diagonal (Problem 56 in (2.5)); if $s_d = 0$, replace B_{dd} with the special singular value decomposition in (2.6.6a).

22. Let $A, B \in M_n$ be symmetric. (a) Show that $A\bar{B}$ is Hermitian if and only if there is a unitary $U \in M_n$ such that $A = U\Sigma U^T$, $B = U\Lambda U^T$, $\Sigma, \Lambda \in M_n(\mathbf{R})$ are diagonal, and the diagonal entries of Σ are nonnegative. (b) Show that $A\bar{B}$ is Hermitian and has nonnegative eigenvalues if and only if there is a unitary $U \in M_n$ such that $A = U\Sigma U^T$, $B = U\Lambda U^T$, $\Sigma, \Lambda \in M_n(\mathbf{R})$ are diagonal, and the diagonal entries of Σ and Λ are nonnegative.

23. Let $A \in M_n$ be given. Suppose that $\operatorname{rank} A = r \geq 1$ and $A^2 = 0$. Provide details for the following outline of a proof that A is unitarily *similar* to

$$\sigma_1 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus \cdots \oplus \sigma_r \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \oplus 0_{n-2r} \tag{2.6.8}$$

in which $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the positive singular values of A. (a) $\operatorname{range} A \subseteq \operatorname{nullspace} A$ and hence $2r \leq n$. (b) Let the columns of $U_2 \in M_{n,n-r}$ be an orthonormal basis for the null space of A^*, so $U_2^* A = 0$. Let $U = [U_1\ U_2] \in M_n$ be unitary. Explain why the columns of $U_1 \in M_{n,r}$ are an orthonormal basis for the range of A and $AU_1 = 0$. (c) $U^*AU = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix}$, in which $B \in M_{r,n-r}$ and $\operatorname{rank} B = r$. (d) $B = V[\Sigma_r\ \ 0_{r,n-2r}]W^*$, in which $V \in M_r$ and $W \in M_{n-r}$ are unitary, and $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$. (e) Let $Z = V \oplus W$. Then $Z^*(U^*AU)Z = \begin{bmatrix} 0 & \Sigma_r \\ 0 & 0 \end{bmatrix} \oplus 0_{n-2r}$,which is similar to (2.6.8) via a permutation matrix.

24. Let $A \in M_n$ be given. Suppose that $\operatorname{rank} A = r \geq 1$ and $A\bar{A} = 0$. Provide details for the following outline of a proof that A is unitarily *congruent* to (2.6.8), in which $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the positive singular values of A. (a) $\operatorname{range} \bar{A} \subseteq \operatorname{nullspace} A$ and hence $2r \leq n$. (b) Let the columns of $U_2 \in M_{n,n-r}$ be an orthonormal basis for the null space of A^T, so $U_2^T A = 0$. Let $U = [U_1\ U_2] \in M_n$ be unitary. Explain why the columns of $U_1 \in M_{n,r}$ are an orthonormal basis for the range of $\bar{A}$ and $AU_1 = 0$. (c) $U^TAU = \begin{bmatrix} 0 & B \\ 0 & 0 \end{bmatrix}$, in which $B \in M_{r,n-r}$ and $\operatorname{rank} B = r$. (d) $B = V[\Sigma_r\ \ 0_{r,n-2r}]W^*$, in which $V \in M_r$ and $W \in M_{n-r}$ are unitary, and $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$. (e) Let $Z = \bar{V} \oplus W$. Then $Z^T(U^TAU)Z = \begin{bmatrix} 0 & \Sigma_r \\ 0 & 0 \end{bmatrix} \oplus 0_{n-2r}$, which is unitarily congruent to (2.6.8) via a permutation matrix.

25. Let $A \in M_n$ and suppose that $\operatorname{rank} A = r < n$. Let $\sigma_1 \geq \cdots \geq \sigma_r > 0$ be the positive singular values of A and let $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$. Show that there is a unitary $U \in M_n$, $K \in M_r$, and $L \in M_{r,n-r}$ such that

$$A = U \begin{bmatrix} \Sigma_r K & \Sigma_r L \\ 0 & 0_{n-r} \end{bmatrix} U^*, \quad KK^* + LL^* = I_r \tag{2.6.9}$$

Hint: Let $A = V\Sigma W^*$ be a singular value decomposition with $\Sigma = \Sigma_r \oplus 0_{n-r}$, write $A = V\Sigma(W^*V)V^*$, and partition $W^*V = \begin{bmatrix} K & L \\ M & N \end{bmatrix}$.

26. Let $A \in M_n$, suppose that $1 \leq \operatorname{rank} A = r < n$, and consider the representation (2.6.9). Show that: (a) A is normal if and only if $L = 0$ and $\Sigma_r K = K\Sigma_r$. *Hint*: $L = 0 \Rightarrow M = 0 \Rightarrow K$ is unitary. (b) $A^2 = 0$ if and only if $K = 0$ (in which case $LL^* = I_r$). (c) $A^2 = 0$ if and only if A is unitarily similar to the direct sum (2.6.8). *Hint*: Proceed as in Problem 23(c).

Further Readings and Notes. The special singular value decomposition (2.6.6a) for complex symmetric matrices was published by L. Autonne in 1915; it has been rediscovered many times since then. Autonne's proof used a version of the uniqueness theorem (2.6.4), but his approach required that the matrix be nonsingular; Problem 20 shows how to deduce the singular case from the nonsingular case. See Section 3.0 of [HJ] for a history of the singular value decomposition, including an account of Autonne's contributions. The principle in Problem 10 ($A^*A = B^*B$ if and only if $A = UB$ for some unitary U) has many applications and generalizations; see R. Horn and I. Olkin, When Does $A^*A = B^*B$ and Why Does One Want to Know?, *Amer. Math. Monthly* 103 (1996) 470-482.

Index