

# Matrix Analysis (Lecture 3)

*Yikun Zhang\**

April 19, 2018

## Abstract

In the last two lectures, we analyze (left and right) eigenvalues and eigenvectors of a matrix, introduce a similarity transformation  $A \rightarrow S^{-1}AS$  for a specified matrix  $A$ , and arise the diagonalizability of a matrix in  $M_n$ . From now on, we restrict our study to a special case of similarity called unitary similarity. This certain type of similarity requires the nonsingular  $S$  to embrace a simple property:  $S^{-1} = S^*$ . In this lecture, we are supposed to introduce basic properties of unitary matrices and some classical unitary matrices that are of greater use in the subsequent lectures. More importantly, the well-known *QR factorization*, which is of considerable theoretical and computational importance, will also be scrutinized.

## 1 Introduction (Page 83)

The chapter 2 of Horn's book begins with a general introduction of transformations involved unitary matrices or conjugate transposes of nonsingular matrices. As defined in Section 1.1 in Lecture 1 and Definition 1.1 in Lecture 2, a similarity transformation  $A \rightarrow S^{-1}AS$  is conducted via a nonsingular matrix  $S$ . When the inverse of this matrix has a special form, that is,  $S^{-1} = S^*$ , the corresponding similarity transformation becomes  $A \rightarrow S^*AS$ , where  $S$  is a so-called unitary matrix. It turns out that similarity via a unitary matrix is not only conceptually simpler than general similarity (the conjugate transpose is much easier to compute than the inverse), but also exhibits superior

---

\*School of Mathematics, Sun Yat-sen University

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  $A$  (Schur form or Schur triangularization; see Lecture 4).

The transformation  $A \rightarrow S^*AS$ , in which  $S$  is nonsingular but not necessarily unitary, is called *\*congruence*.

**Remark.** \*congruence is not necessarily a similarity transformation. However, similarity by a unitary matrix is both a similarity and a \*congruence.

For  $A \in M_{n,m}$ , the transformation  $A \rightarrow UAV$ , in which  $U \in M_n$  and  $V \in M_m$  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 rectangular matrices: Every  $A \in M_{m,n}$  is unitarily equivalent to a nonnegative diagonal matrix whose diagonal entries (the singular values of  $A$ ) are of great significance.

## 2 Unitary matrices and the $QR$ factorization

(Page 83-91, Page 15-16, Page 19-20)

We have come across orthogonal vectors when examining left and right eigenvectors associated with different eigenvalues. Here we reiterate the definition of orthogonality.

**Definition 2.1.** A list of vectors  $x_1, \dots, x_k \in \mathbf{C}^n$  is orthogonal if  $x_i^*x_j = 0$  for all  $i \neq j, i, j \in \{1, \dots, k\}$ . If, in addition,  $x_i^*x_i = 1$  for all  $i = 1, \dots, k$  (that is, the vectors are normalized), then the list is orthonormal.

**Convention.** It is often convenient to say that “ $x_1, \dots, x_k$  are orthogonal (respectively, orthonormal)” instead of the more formal statement “the list of vectors  $v_1, \dots, v_k$  is orthogonal (orthonormal, respectively).”

**Example 2.2** (normalization). If  $y_1, \dots, y_k \in \mathbf{C}^n$  are orthogonal and nonzero, the vectors  $x_1, \dots, x_k$  defined by  $x_i = (y_i^*y_i)^{-\frac{1}{2}}y_i, i = 1, \dots, k$  are orthonormal.

**Definition 2.3.** Given any set  $S \subset \mathbf{C}^n$ , its orthogonal complement is the set  $S^\perp = \{x \in \mathbf{C}^n : x^*y = 0 \text{ for all } y \in S\}$  if  $S$  is nonempty; if  $S$  is empty, then  $S^\perp = \mathbf{C}^n$ .

**Remark.** (a) In either case,  $S^\perp = (\text{span } S)^\perp$ . Even if  $S$  is not a subspace,  $S^\perp$  is always a subspace. We have  $(S^\perp)^\perp = \text{span } S$ , and  $(S^\perp)^\perp = S$  if  $S$  is a

subspace.

(b) It is always the case that  $\dim S^\perp + \dim(S^\perp)^\perp = n$ . If  $S_1$  and  $S_2$  are subspaces, then  $(S_1 + S_2)^\perp = S_1^\perp \cap S_2^\perp$ .

An orthogonal list of vectors embraces many benign properties, which make them computationally simple.

**Theorem 2.4.** *Every orthogonal list of vectors in  $\mathbf{C}^n$  is linearly independent.*

*Proof.* Suppose that  $\{y_1, \dots, y_k\}$  is an orthogonal set. Normalize them as Example 2.2 did and obtain an orthonormal list of vectors  $\{x_1, \dots, x_k\}$ . Assume that  $0 = \alpha_1 x_1 + \dots + \alpha_k x_k$ . Then  $0 = (\alpha_1 x_1 + \dots + \alpha_k x_k)^* (\alpha_1 x_1 + \dots + \alpha_k x_k) = \sum_{i,j} \bar{\alpha}_i \alpha_j x_i^* x_j = \sum_{i=1}^k |\alpha_i|^2 x_i^* x_i = \sum_{i=1}^k |\alpha_i|^2$  because the vectors  $x_i$  are orthonormal. Thus, all  $\alpha_i = 0$  and hence  $\{x_1, \dots, x_k\}$  is a linearly independent set, which in turn means that  $\{y_1, \dots, y_k\}$  is linearly independent.  $\square$

**Example 2.5.** The fact that an orthogonal list of vectors  $x_1, \dots, x_k \in \mathbf{C}^n$  is linearly independent allows for two cases, either  $k \leq n$  or at least  $k - n$  of the vectors  $x_i$  are zero vectors. This is because there are at most  $n$  linearly independent vectors in  $\mathbf{C}^n$ , so the cardinality of a nonzero orthogonal set must satisfy  $k \leq n$ . If  $k > n$ , we can choose  $n$  vector from the list such that  $\text{span}\{x_1, \dots, x_n\} = \mathbf{C}^n$ , since  $x_1, \dots, x_k$  are linearly independent. Then  $\{x_{n+1}, \dots, x_k\} \subseteq (\text{span}\{x_1, \dots, x_n\})^\perp = (\mathbf{C}^n)^\perp = \{0\}$ , yielding that  $x_{n+1} = \dots = x_n = 0$ .

A linearly independent list need not be orthonormal, but one can apply the Gram-Schmidt orthonormalization procedure to it and obtain an orthonormal list with the same span.

**Example 2.6** (Gram-Schmidt orthonormalization). We first define the scalar  $\langle x, y \rangle = y^* x$  as the *Euclidean inner product* (*standard inner product*, *usual inner product*, *scalar product*, *dot product*) of  $x, y \in \mathbf{C}^n$ . And the *Euclidean norm* (*usual norm*, *Euclidean length*) function on  $\mathbf{C}^n$  is the real-valued function  $\|x\|_2 = \langle x, x \rangle^{\frac{1}{2}} = (x^* x)^{\frac{1}{2}}$ .

The *Gram-Schmidt process* starts with a list of vectors  $v_1, \dots, v_n$  and (if the given list is linearly independent) produces an orthonormal list of vectors  $z_1, \dots, z_n$  such that  $\text{span}\{z_1, \dots, z_k\} = \text{span}\{v_1, \dots, v_k\}$  for each  $k = 1, \dots, n$ . The vectors  $z_i$  may be calculated in turn as follows: Let  $y_1 = v_1$  and normalize

it:  $z_1 = \frac{y_1}{\|y_1\|_2}$ . Let  $y_2 = x_2 - \langle x_2, z_1 \rangle z_1$  ( $y_2$  is orthogonal to  $z_1$ ) and normalize it:  $z_2 = \frac{y_2}{\|y_2\|_2}$ . Once  $z_1, \dots, z_{k-1}$  have been determined, the vector

$$y_k = x_k - \langle x_k, z_{k-1} \rangle z_{k-1} - \langle x_k, z_{k-2} \rangle z_{k-2} - \cdots - \langle x_k, z_1 \rangle z_1$$

is orthogonal to  $z_1, \dots, z_{k-1}$ ; normalize it:  $z_k = \frac{y_k}{\|y_k\|_2}$ . Continue until  $k = n$ . If we denote  $Z = [z_1 \cdots z_n]$  and  $X = [x_1 \cdots x_n]$ , the Gram-Schmidt process gives a factorization  $X = ZR$ , where the square matrix  $R = [r_{ij}]$  is nonsingular and upper triangular; that is,  $r_{ij} = 0$  whenever  $i > j$ .

The Gram-Schmidt process may be applied to any finite list of vectors, independent or not. If  $x_1, \dots, x_n$  are linearly independent, the Gram-Schmidt process produces a vector  $y_k = 0$  for the least value of  $k$  for which  $x_k$  is a linear combination of  $x_1, \dots, x_{k-1}$ .

Since any nonzero subspace of  $\mathbf{R}^n$  or  $\mathbf{C}^n$  always has a linearly independent list of vectors that spans the subspace, one can apply the Gram-Schmidt orthonormalization to obtain its orthonormal basis.

We are now prepared to make a formal definition for the main concept of this lecture, unitary matrices, and investigate their properties.

**Definition 2.7.** A matrix  $U \in M_n$  is unitary if  $U^*U = I$ . A matrix  $U \in M_n(\mathbf{R})$  is real orthogonal if  $U^T U = I$ .

**Example 2.8.** The matrices  $U = \frac{1}{\sqrt{2}} \begin{bmatrix} I_n & iI_n \\ iI_n & I_n \end{bmatrix}$ ,  $V = \frac{1}{\sqrt{2}} \begin{bmatrix} -iI_n & -iI_n \\ I_n & -I_n \end{bmatrix}$ , and  $Q = \frac{1}{\sqrt{2}} \begin{bmatrix} I_n & I_n \\ I_n & -I_n \end{bmatrix}$  are all unitary. In particular,  $Q$  is also real orthogonal.

We list some of the basic equivalent conditions for  $U$  to be unitary in the next theorem.

**Theorem 2.9.** If  $U \in M_n$ , the following statements 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  $U$  are orthonormal.
- (f) The rows of  $U$  are orthonormal.

(g) For all  $x \in \mathbf{C}^n$ ,  $\|x\|_2 = \|Ux\|_2$ , that is,  $x$  and  $Ux$  have the same Euclidean norm.

$$(e) \Leftrightarrow (a) \Leftrightarrow (b) \Leftrightarrow (c) \Leftrightarrow (d) \Leftrightarrow (f)$$

*Proof.*

$$\begin{array}{c} \Updownarrow \\ (g) \end{array}$$

(a) implies (b) since  $U^{-1}$  (when it exists) is the unique matrix; the definition of unitary says that  $U$  is nonsingular and  $U^{-1} = U^*$ . Since  $BA = I$  if and only if  $AB = I$  (for  $A, B \in M_n$ ), (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  $U = [u_1 \cdots u_n]$  according to its columns. Then  $U^*U = I$  indicates 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). Likewise, (d) and (f) are equivalent.

If  $U$  is unitary 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  $z, w \in \mathbf{C}^n$  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  $\operatorname{Re} ie_p^T e_q = 0 = \operatorname{Re} ie_p^T A e_q = \operatorname{Re} i a_{pq} = -\operatorname{Im} a_{pq}$ , so every entry of  $A$  is real. Finally, take  $z = e_p$  and  $w = e_q$  and compute  $e_p^T e_q = \operatorname{Re} e_p^T e_q = \operatorname{Re} e_p^T A e_q = a_{pq}$ , which tells us that  $A = I$  and  $U$  is unitary.  $\square$

**Remark.** When proving (g)  $\Rightarrow$  (a), we choose  $x = w + z$  instead of just taking arbitrary  $x$  because if we take  $x = e_1$ , then  $1 = x^*x = x^*Ax = a_{11}$  and only the diagonal entries of  $A$  can be determined by consecutively changing the value of  $x$ .

An important geometrical fact is that any two lists containing equal numbers of orthonormal vectors are related via a unitary transformation.

**Theorem 2.10.** If  $X = [x_1 \cdots x_n] \in M_{n,k}$  and  $Y = [y_1 \cdots y_n] \in M_{n,k}$  have orthonormal columns, then there is a unitary  $U \in M_n$  such that  $Y = UX$ . If  $X$  and  $Y$  are real, then  $U$  may be taken to be real.

*Proof.* Extend each of the orthonormal lists  $x_1, \dots, x_k$  and  $y_1, \dots, y_k$  to orthonormal bases of  $\mathbf{C}^n$  by furnishing each list with linearly independent vectors and applying Gram-Schmidt orthonormalization. That is, construct

unitary matrices  $V = [X \ X_2]$  and  $W = [Y \ Y_2] \in M_n$ . Then  $U = WV^*$  is unitary and  $[Y \ Y_2] = W = UV = [UX \ UX_2]$ , so  $Y = UX$ . If  $X$  and  $Y$  are real, the matrices  $[X \ X_2]$  and  $[Y \ Y_2]$  may be chosen to be real orthogonal (their columns are orthonormal bases of  $\mathbf{R}^n$ ).  $\square$

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. To prove this result, we rely on the *law of complementary nullities*.

**Lemma 2.11.** *Suppose that  $A \in M_n$  is nonsingular, let  $\alpha$  and  $\beta$  be nonempty subsets of  $\{1, \dots, n\}$ , and write  $|\alpha| = r$  and  $|\beta| = s$  for the cardinalities of  $\alpha$  and  $\beta$ . The law of complementary nullities is*

$$\text{nullity}(A[\alpha, \beta]) = \text{nullity}(A^{-1}[\beta^c, \alpha^c]) \quad (1)$$

which is equivalent to the rank identity

$$\text{rank}(A[\alpha, \beta]) = \text{rank}(A^{-1}[\beta^c, \alpha^c]) + r + s - n \quad (2)$$

*Proof.* Since we can permute rows and columns to place first the  $r$  rows indexed by  $\alpha$  and the  $s$  columns indexed by  $\beta$ , it suffices to consider the presentations

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ and } A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

where  $A_{11}$  and  $B_{11}^T$  are  $r$ -by- $s$  and  $A_{22}$  and  $B_{22}^T$  are  $(n-r)$ -by- $(n-s)$ .

The underlying principle here is very simple. Suppose that the nullity of  $A_{11}$  is  $k$ . If  $k \geq 1$ , let the columns of  $X \in M_{s,k}$  is a basis for the null space of  $A_{11}$ . Since  $A$  is nonsingular,

$$A \begin{bmatrix} X \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}X \\ A_{21}X \end{bmatrix} = \begin{bmatrix} 0 \\ A_{21}X \end{bmatrix}$$

has full rank, so  $A_{21}X$  has  $k$  independent columns. But

$$\begin{bmatrix} B_{12}(A_{21}X) \\ B_{22}(A_{21}X) \end{bmatrix} = A^{-1} \begin{bmatrix} 0 \\ A_{21}X \end{bmatrix} = A^{-1}A \begin{bmatrix} X \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ 0 \end{bmatrix}$$

so  $B_{22}(A_{21}X) = 0$  and hence  $\text{nullity } B_{22} \geq k = \text{nullity } A_{11}$ , a statement that is trivially correct if  $k = 0$ . Symmetrically, a similar argument starting with  $B_{22}$  shows that  $\text{nullity } A_{11} \geq \text{nullity } B_{22}$ .

By the *rank-nullity* theorem, we obtain that  $s - \text{rank } A_{11} = n - r - \text{rank } B_{22}$ .  $\square$

**Theorem 2.12.** *Let a unitary  $U \in M_n$  be partitioned as  $U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$ , where  $U_{11} \in M_k$ . Then  $\text{rank } U_{12} = \text{rank } U_{21}$  and  $\text{rank } U_{22} = \text{rank } U_{11} + n - 2k$ . In particular,  $U_{12} = 0$  if and only if  $U_{21} = 0$ , in which case  $U_{11}$  and  $U_{22}$  are unitary.*

*Proof.* Applying the law of complementary nullities (1) and using the fact that  $U^{-1} = \begin{bmatrix} U_{11}^* & U_{12}^* \\ U_{21}^* & U_{22}^* \end{bmatrix}$ , we have that  $\text{nullity } U_{12} = \text{nullity } U_{21}^*$ ,  $\text{nullity } U_{11} = \text{nullity } U_{22}^*$ . Since  $U_{11} \in M_k, U_{12} \in M_{k, n-k}, U_{21} \in M_{n-k, k}, U_{22} \in M_{n-k, n-k}$  and conjugate transposition preserves the rank, we conclude that  $\text{rank } U_{12} = \text{rank } U_{21}$  and  $k - \text{rank } U_{11} = n - k - \text{rank } U_{22}$ .

In the case when  $U_{12} = U_{21} = 0$ ,  $U^*U = I$  ensures that

$$\begin{bmatrix} U_{11}^* & 0 \\ 0 & U_{22}^* \end{bmatrix} \begin{bmatrix} U_{11} & 0 \\ 0 & U_{22} \end{bmatrix} = I,$$

yielding that  $U_{11}^*U_{11} = U_{22}^*U_{22} = I$ .  $\square$

**Corollary 2.13.** *A unitary matrix is upper triangular if and only if it is diagonal.*

*Proof.* ( $\Leftarrow$ ) A diagonal matrix is obviously upper triangular.

( $\Rightarrow$ ) Partition the upper triangular unitary matrix  $U$  into

$$U = \begin{bmatrix} u_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix}, \quad u_{11} \in M_1, U_{12} \in M_{1, n-1}, U_{22} \in M_{n-1}.$$

By Theorem 2.12,  $U_{12} = 0$ . Since  $U_{22}$  is still upper triangular, by repeating the same argument  $n - 1$  times, we prove that  $U$  is diagonal.  $\square$

Besides these properties, the set of unitary matrices forms a group under the standard matrix multiplication.

**Theorem 2.14.** *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, real orthogonal) group, a subgroup of  $GL(n, \mathbf{C})$ .*

*Proof.* (Closure): If  $U, V \in M_n$  are unitary, i.e.,  $U^*U = V^*V = I$ , then  $(UV)^*(UV) = V^*U^*UV = V^*IV = V^*V = I$ , yielding that  $UV$  is also unitary.

(Associativity): The standard matrix multiplication are associative.

(Identity): Since the identity matrix  $I_n$  is unitary, it serves as the identity for this subgroup.

(Inverse): The inverse of a unitary matrix  $U$  is its conjugate transpose  $A^*$ , which is also unitary by (d) in Theorem 2.9.

The argument can be inherited to the case when matrices are real orthogonal without any effort.  $\square$

The group of unitary matrices in  $M_n$  has another very important property. 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 that it is a *bounded* subset. (In fact, it is bounded by  $n$ .) If  $U_k = [u_{ij}^{(k)}]$  is an infinite sequence of unitary matrices,  $k = 1, 2, \dots$ , such that  $\lim_{k \rightarrow \infty} u_{ij}^{(k)} = u_{ij}$  exists for all  $i, j = 1, 2, \dots, n$ , then from the identity  $U_k^*U_k = I$  for all  $k = 1, 2, \dots$ , we see that  $\lim_{k \rightarrow \infty} U_k^*U_k = U^*U = I$ , where  $U = [u_{ij}]$  and the limit is taken componentwisely. 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, 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.

**Lemma 2.15.** *Let  $U_1, U_2, \dots \in M_n$  be a given infinite sequence of unitary matrices. There exists an infinite subsequence  $U_{k_1}, U_{k_2}, \dots, 1 \leq k_1 < k_2 < \dots$ , such that all of the entries of  $U_{k_i}$  converge (as sequences of complex numbers) to the entries of a unitary matrix as  $i \rightarrow \infty$ .*

*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.  $\square$

**Remark.** (a) The selection principle (2.15) applies as well to the real orthogonal group; that is, an infinite sequence of real orthogonal matrices has an infinite subsequences that converges to a real orthogonal matrix.

(b) The unitary limit guaranteed by the lemma need not be unique; it can depend on the subsequence chosen.



**Example 2.16.** Consider the sequence of unitary matrices  $U_k = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^k$ ,  $k =$

$$1, 2, \dots \text{ Then } U_k = \begin{cases} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & \text{if } k \text{ is odd,} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & \text{if } k \text{ is even;} \end{cases} \quad \text{that is, there are two possible lim-}$$

its of subsequences.

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  $A \rightarrow A^{-1}A^*$  for all nonsingular  $A \in M_n$ .

**Theorem 2.17.** *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.* ( $\Leftarrow$ ) 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  $B^*$ .

( $\Rightarrow$ ) 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  $S_\theta = e^{i\theta}S$  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  $H_\theta = A^*H_\theta A$ , where  $H_\theta = S_\theta + S_\theta^*$  is Hermitian.

If  $H_\theta$  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  $H = H_{\theta_0}$  is nonsingular and has the property that  $H = A^*HA$ .

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

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

trix factorizations. More importantly, Householder matrices are considered as the fundamental ingredient of *QR factorization*.

**Example 2.18** (Plane rotations). Let  $1 \leq i < j \leq n$  and let

$$U(\theta; i, j) = \begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \cos \theta & & & -\sin \theta & \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & \\ & & & \sin \theta & & & \cos \theta & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix}$$

denote the result of replacing the  $(i, i)$  and  $(j, j)$  entries of  $n$ -by- $n$  identity matrix by  $\cos \theta$ , replacing its  $(i, j)$  entry by  $-\sin \theta$  and replacing its  $(j, i)$  entry by  $\sin \theta$ . The matrix  $U(\theta; i, j)$  is called a *plane rotation* or *Givens rotation*. It is easy to verify that  $U(\theta; i, j)$  is real orthogonal for any pair of indices  $i, j$  with  $1 \leq i < j \leq n$  and any parameter  $\theta \in [0, 2\pi)$ , i.e.,  $U(\theta; i, j)^{-1} = U(\theta; i, j)^T = U(-\theta; i, j)$ . Indeed, the matrix  $U(\theta; i, j)$  carries out a rotation (through an angle  $\theta$ ) 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.

**Example 2.19** (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  $U_w = I - 2ww^*$ . A Householder matrix embrace the following important properties.

(a) A Householder matrix  $U_w$  is both unitary and Hermitian, so  $U_w^{-1} = U_w$ .

*Proof.* By definition,  $U_w^* = I - 2(w^*w)^{-1}(ww^*)^* = I - 2(w^*w)^{-1}ww^* = U_w$ ,

$$\begin{aligned} U_w^* U_w &= U_w^2 = [I - 2(w^*w)^{-1}ww^*][I - 2(w^*w)^{-1}ww^*] \\ &= I - 4(w^*w)^{-1}(ww^*) + 4(w^*w)^{-2}[w(w^*w)w^*] \\ &= I - 4(w^*w)^{-1}(ww^*) + 4(w^*w)^{-1}(ww^*) \\ &= I, \end{aligned}$$

showing that  $U_w^{-1} = U_w^* = U_w$ .  $\square$

**Remark.** In a similar fashion, we can show that the Householder matrix  $U_w$  is real orthogonal and symmetric when  $w \in \mathbf{R}^n$  is a nonzero vector.

(b) 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  $U_w w = -w$ .

*Proof.* If  $x \perp w$ , namely,  $w^*x = 0$ , we have that  $U_w x = [I - 2(w^*w)^{-1}ww^*]x = x - 2(w^*w)^{-1}w(w^*x) = x$ . Additionally,  $U_w w = [I - 2(w^*w)^{-1}ww^*]w = w - 2(w^*w)^{-1}w(w^*w) = -w$ .  $\square$

(c) The eigenvalues of a Householder matrix  $U_w \in M_n$  are always  $-1, \overbrace{1, \dots, 1}^{n-1}$  and  $\det U_w = -1$  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  $+1$ ).

*Proof.* Using Formula (19) in Lecture 1 or Example 1.26 in Lecture 2 and the fact that  $\text{adj}(\alpha I) = \alpha^{n-1}I$ , we compute

$$\begin{aligned} p_{U_w}(t) &= \det(tI - U_w) = \det[(t-1)I + 2(w^*w)^{-1}ww^*] \\ &= \det[(t-1)I] + 2(w^*w)^{-1}w^* \text{adj}[(t-1)I]w \\ &= (t-1)^n + 2(t-1)^{n-1}(w^*w)^{-1}w^*w \\ &= (t-1)^{n-1}(t+1), \end{aligned}$$

yielding that  $U_w \in M_n$  has eigenvalues  $(-1)$  with multiplicity 1 and 1 with multiplicity  $(n-1)$  and thus  $\det(U_w) = \prod_{i=1}^n \lambda_i(U_w) = -1$ , where  $\lambda_i(U_w)$  is the  $i^{\text{th}}$  eigenvalue of  $U_w$ .  $\square$

**Definition 2.20.** A linear transformation  $T : \mathbf{C}^n \rightarrow \mathbf{C}^m$  is called a Euclidean isometry if  $\|x\|_2 = \|Tx\|_2$  for all  $x \in \mathbf{C}^n$ .

Theorem 2.9 (g) says that a square complex matrix  $U \in M_n$  is a Euclidean isometry (via  $U : x \rightarrow Ux$ ) if and only if it is unitary. However, the proof of Theorem 2.9 (g) did not explicitly give us a construction of the unitary matrix that takes any given vector in  $\mathbf{C}^n$  into any other in  $\mathbf{C}^n$  that has the same Euclidean norm. Fortunately, Householder matrices and unitary scalar matrices shed light on such an elegant construction.

**Theorem 2.21.** Let  $x, y \in \mathbf{C}^n$  be given and suppose that  $\|x\|_2 = \|y\|_2 > 0$ . If  $y = e^{i\theta}x$  for some real  $\theta$ , 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  $U(y, x) = e^{i\phi}U_w$ , where  $U_w = I - 2(w^*w)^{-1}ww^*$  is a Householder matrix. 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  $\theta$ .

If  $x$  and  $y$  are linearly independent, the Cauchy-Schwarz inequality  $|y^*x| \leq \|x\|_2\|y\|_2$  and its equality condition guarantee that  $x^*x \neq |x^*y|$ . In order to verify that  $U(y, x)x = y$ , and  $U(y, x)z \perp y$  whenever  $z \perp x$ , we first compute some small items involved in the whole calculating process:

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

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

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

Therefore,

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

If  $z \perp x$ , then  $w^*z = -y^*z$  and

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

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.  $\square$

**Remark.** A matrix  $A \in M_n$  is said to be *essentially Hermitian* if  $e^{i\theta}A$  is Hermitian for some  $\theta \in \mathbf{R}^n$ .

**Example 2.22.** 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. We construct  $U(y, e_1)$  using the recipe in the preceding Theorem 2.21 and conclude that its first column should be  $y$ , since  $y = U(y, e_1)e_1$ . More generally, let  $x \in \mathbf{C}^n$  be a given nonzero vector and therefore, the matrix  $U(\|x\|_2 e_1, x)$  constructed in the preceding Theorem 2.21 is an essentially Hermitian unitary matrix that takes  $x$  into  $\|x\|_2 e_1$ .

We now apply the construction of the Euclidean isometry in Theorem 2.21 to deduce the well-known *QR factorization*.

**Theorem 2.23.** (*QR factorization*) Let  $A \in M_{n,m}$  be given.

- (a) If  $n \geq m$ , there is a  $Q \in M_{n,m}$  with orthonormal columns and an upper triangular  $R \in M_m$  with nonnegative main diagonal entries such that  $A = QR$ . In particular, if  $m = n$ , then the factor  $Q$  is unitary.
- (b) If  $\text{rank } A = m$ , then the factors  $Q$  and  $R$  in (a) are uniquely determined and the main diagonal entries of  $R$  are all positive.
- (c) There is a unitary  $Q \in M_n$  and an upper triangular  $R \in M_{n,m}$  with nonnegative diagonal entries such that  $A = QR$ .
- (d) If  $A$  is real, then the factors  $Q$  and  $R$  in (a), (b), and (c) may be taken to be real.

*Proof.* (a) 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 2.21 gives an explicit construction for such a matrix, namely,  $U(r_1 e_1, a_1)$ . Partition

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

where  $A_2 \in M_{n-1, m-1}$ . Let  $a_2 \in \mathbf{C}^{n-1}$  be the first column of  $A_2$  and let  $r_2 = \|a_2\|_2$ . Use Theorem 2.21 again to construct a unitary  $V_2 = U(r_2 e_1, a_2) \in M_{n-1}$  such that  $V_2 a_2 = r_2 e_1$  and let  $U_2 = I_1 \oplus V_2$ . Then

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

Repeat this construction  $m$  times to obtain

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

where  $R \in M_n$  is upper triangular. Its main diagonal entries are  $r_1, \dots, r_m$ ; they are all nonnegative. Let  $U = U_m U_{m-1} \cdots U_2 U_1$ .

Partition  $U^* = U_1^* U_2^* \cdots U_{m-1}^* U_m^* = [Q \ Q_2]$ , where  $Q \in M_{n, m}$  has orthonormal columns (it contains the first  $m$  columns of a unitary matrix). Then  $A = QR$ , as desired. In particular, if  $m = n$ ,  $Q$  becomes a square matrix with orthonormal columns, which is thus unitary.

(b) If  $A$  has full column rank, then  $R$  is nonsingular, so its main diagonal entries are all positive. Suppose that  $\text{rank } A = m$  and  $A = QR = \tilde{Q}\tilde{R}$ , where  $R$  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  $D = \tilde{R}R^{-1} = \tilde{R}^{-*}R^* = (DR)^{-*}R^* = D^{-1}R^{-*}R^* = D^{-1}$ , showing that  $D^2 = I$  and hence  $D = I$ . We conclude that  $\tilde{R} = R$  and hence  $\tilde{Q} = Q$ .

(c) If  $n \geq m$ , we may start with the factorization in (a), let  $Q' = [Q \ Q_2] \in M_n$  be unitary, let  $R' = \begin{bmatrix} R \\ 0 \end{bmatrix} \in M_{n, m}$ , and observe that  $A = QR = Q'R'$ .

If  $n < m$ , we may undertake the construction in (a) (left multiplication by a sequence of scalar multiples of Householder transformations) and stops after  $n$  steps, when the factorization  $U_n \cdots U_1 A = [R \ \star]$  is achieved and  $R$  is upper triangular. Entries in the  $\star$  block need not be zero.

(d) The assertion follows from the assurance in Theorem 2.21 that the unitary matrices  $U_i$  involved in the constructions in (a) and (c) may all be chosen to be real.  $\square$

**Corollary 2.24.** (*Cholesky factorization*) Any  $B \in M_n$  of the form  $B = A^*A$ ,  $A \in M_n$ , may be written as  $B = LL^*$ , where  $L \in M_n$  is lower triangular and has nonnegative diagonal entries.

*Proof.* By Theorem 2.23, we factorize  $A$  into  $A = QR$ , where  $Q \in M_n$  is unitary and  $R \in M_n$  with nonnegative diagonal entries is upper triangular. Then  $B = A^*A = (QR)^*QR = R^*Q^*QR = R^*IR = R^*R$ . Choose  $L = R^*$ , which is lower triangular and has nonnegative diagonal entries.  $\square$

**Remark.** (a) Theorem 2.23 (b) illuminates that if  $A$  is nonsingular, all the diagonal entries of  $L$  should be positive and  $L$  is unique.  
(b) Every positive definite or semidefinite matrix may be factored in this way.

Some variants of the QR factorization of  $A \in M_{n,m}$  can be useful in practice and we are supposed to discuss them in detail.

**Corollary 2.25.** Let  $A \in M_{n,m}$  be given. If  $n \leq m$ , then there is a  $Q \in M_{n,m}$  with orthonormal rows and a lower triangular  $L \in M_n$  with nonnegative main diagonal entries such that  $A = LQ$ .

*Proof.* If  $n \leq m$ , then  $A^* \in M_{m,n}$  and  $A^* = \tilde{Q}R$  by Theorem 2.23, where  $\tilde{Q} \in M_{m,n}$  has orthonormal columns and  $R \in M_n$  is upper triangular. Then  $A = R^*\tilde{Q}^*$  and the results follow by replacing  $R^*$  by  $L$  and  $\tilde{Q}^*$  by  $Q$ .  $\square$

**Remark.** If  $Q' = \begin{bmatrix} Q \\ Q_2 \end{bmatrix}$  is unitary, we have the factorization of the form  $A = [L \ 0]Q'$ .

**Corollary 2.26.** Let  $A \in M_{n,m}$  be given.

- (a) There is a  $Q \in M_{n,m}$  with orthonormal columns and a lower triangular  $L \in M_n$  such that  $A = QL$ . If  $\tilde{Q} = [Q \ Q_2]$  is unitary, we have a factorization of the form  $A = \tilde{Q} \begin{bmatrix} L \\ 0 \end{bmatrix}$ .
- (b) If  $n \leq m$ , there is an upper triangular  $R \in M_n$ ,  $Q \in M_{n,m}$  with orthonormal rows, and a unitary  $\tilde{Q} = \begin{bmatrix} Q \\ Q_2 \end{bmatrix} \in M_n$  such that  $A = RQ = [R \ 0]\tilde{Q}$ .

*Proof.* (a) Let  $K_p$  be the (real orthogonal and symmetric)  $p$ -by- $p$  reversal matrix

$$\begin{bmatrix} & & & 1 \\ & & \ddots & \\ & & & \\ 1 & & & \end{bmatrix}$$

which has the pleasant property that  $K_p^2 = I_p$ . For square matrices  $R \in M_p$ , the matrix  $L = K_p R K_p$  is lower triangular if  $R$  is upper triangular; the main diagonal entries of  $L$  are those of  $R$ , with the order reversed.

If  $n \geq m$  and  $AK_m = Q'R$  as in Theorem 2.23 (a), then  $A = (Q'K_m)(K_m R K_m)$ , which is a factorization of the form with  $Q' \in M_{n,m}$  whose columns are orthonormal and an upper triangular  $R \in M_n$ . Thus,  $Q = Q'K_m$  has the reversed order of columns of  $Q'$ , which are still orthonormal, and  $L = (K_m R K_m)$  is lower triangular.

If  $n \leq m$  and we apply Theorem 2.23 (d) to  $AK_m$ , we obtain that  $A = (QK_n)(K_n[R \star]K_m)$ , which is a factorization of the form

$$A = \tilde{Q}L,$$

where  $\tilde{Q} \in M_n$  is unitary and  $L \in M_{n,m}$  is lower triangular.

(b) If  $n \leq m$  and we apply Theorem 2.23 (a) to  $A^*$ , we have that  $A^*K_n = \tilde{Q}\tilde{R}$  and  $A^* = (\tilde{Q}K_n)(K_n\tilde{R}K_n)$ , i.e.,  $A = (K_n\tilde{R}K_n)^*(\tilde{Q}K_n)^*$ . Thus,  $R = (K_n\tilde{R}K_n)^*$  is upper triangular and  $Q = (\tilde{Q}K_n)^*$  has orthonormal rows.  $\square$

## References

- [1] Roger A. Horn, Charles R. Johnson (2012) *Matrix Analysis, Second Edition*. Cambridge University Press.