Section 8.6 Singular Value Decomposition

SINGULAR VALUE DECOMPOSITION OF SQUARE MATRICES

We know from our work in Section 8.3 that symmetric matrices are orthogonally diagonalizable and are the only matrices with this property (Theorem 8.3.4). The orthogonal diagonalizability of an $n \times n$ symmetric matrix A means it can be factored as

$$A = PDP^T \tag{1}$$

where P is an $n \times n$ orthogonal matrix of eigenvectors of A, and D is the diagonal matrix whose diagonal entries are the eigenvalues corresponding to the column vectors of P. In this section we will call (1) an ***eigenvalue decomposition*** of A (abbreviated EVD of A).

Theorem 8.3.4

(a) A matrix is orthogonally diagonalizable if and only if it is symmetric.

There are two main paths that one might follow in looking for other kinds of factorizations of a general square matrix A: One might look for factorizations of the form

$$A = PJP^{-1}$$

in which P is invertible but not necessarily orthogonal, or one might look for factorizations of the form

$$A = U\Sigma V^T$$

in which U and V are orthogonal but not necessarily the same.

The first path leads to factorizations in which J is either diagonal (Theorem 8.2.6) or a certain kind of block diagonal matrix, called a *Jordan canonical form* in honor of the French mathematician Camille Jordan (1838–1922).

A block diagonal matrix is a partitioned ('block') square matrix that the main-diagonal partitions are square matrices and all off-diagonal partitions are zero matrices.

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{bmatrix}$$

Jordan canonical forms, which we will not consider in this text, are important theoretically and in certain applications, but they are of lesser importance numerically because of the roundoff difficulties that result from the lack of orthogonality in P.

Our discussion in this section will focus on the second path, starting with the following diagonalization theorem.

Theorem 8.2.6 *An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.*

Theorem 8.6.1 *If A is an $n \times n$ matrix of rank k, then A can be factored as*

$$A = U\Sigma V^T$$

where U and V are $n \times n$ orthogonal matrices and Σ is an $n \times n$ diagonal matrix whose main diagonal has k positive entries and $n - k$ zeros.

Proof The matrix A^TA is symmetric, so it has an eigenvalue decomposition

$$A^TA = VDV^T$$

Keep in mind that V is orthogonal ($V^{-1} = V^T$ so $V^TV = I$).

where the column vectors of V are unit eigenvectors of A^TA and D is the diagonal matrix whose diagonal entries are the corresponding eigenvalues of A^TA.

These eigenvalues are nonnegative, for if λ is an eigenvalue of A^TA and $\mathbf{x}$ is a corresponding eigenvector, then Formula (12) of Section 3.2 implies that

$$\|A\mathbf{x}\|^2 = A\mathbf{x} \cdot A\mathbf{x} = \mathbf{x} \cdot A^TA\mathbf{x} = \mathbf{x} \cdot \lambda\mathbf{x} = \lambda(\mathbf{x} \cdot \mathbf{x}) = \lambda\|\mathbf{x}\|^2$$

from which it follows that $\lambda \geq 0$.

$$\boxed{A\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^T\mathbf{v} \qquad (12)}$$

Since Theorems 7.5.8 and 8.2.3 imply that

$$\text{rank}(A) = \text{rank}(A^TA) = \text{rank}(D)$$

and since A has rank k, it follows that there are k positive entries and $n - k$ zeros on the main diagonal of D.

For convenience, suppose that the column vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of V have been ordered so that the corresponding eigenvalues of A^TA are in nonincreasing order

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

Thus,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0 \quad \text{and} \quad \lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = 0 \tag{2}$$

Theorem 7.5.8 *If A is an $m \times n$ matrix, then:*

⋮

(d) A and A^TA have the same rank.

Theorem 8.2.3

(a) Similar matrices have the same determinant.

(b) Similar matrices have the same rank.

⋮

Now consider the set of image vectors

$$\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n\}$$

This is an orthogonal set, for if $i \neq j$, then the orthogonality of $\mathbf{v}_i$ and $\mathbf{v}_j$ implies that

$$A\mathbf{v}_i \cdot A\mathbf{v}_j = \mathbf{v}_i \cdot A^T A\mathbf{v}_j = \mathbf{v}_i \cdot \lambda_j \mathbf{v}_j = \lambda_j(\mathbf{v}_i \cdot \mathbf{v}_j) = 0 \tag{3}$$

Moreover,

$$\|A\mathbf{v}_i\|^2 = A\mathbf{v}_i \cdot A\mathbf{v}_i = \mathbf{v}_i \cdot A^T A\mathbf{v}_i = \mathbf{v}_i \cdot \lambda_i \mathbf{v}_i = \lambda_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \lambda_i \|\mathbf{v}_i\|^2 = \lambda_i$$

from which it follows that

$$\|A\mathbf{v}_i\| = \sqrt{\lambda_i} \qquad (i = 1, 2, \ldots, n) \tag{4}$$

Vectors **v** are unit eigenvectors of A^TA.

The product $A\mathbf{v}_i$ may be zero!

$$\boxed{A\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^T\mathbf{v} \qquad (12)}$$

i, not 1!

Since $\lambda_1 > 0$ for $i = 1, 2, \ldots, k$, it follows from (3) and (4) that

$$\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_k\} \tag{5}$$

is an orthogonal set of k nonzero vectors in the column space of A; and since we know that the column space of A has dimension k (since A has rank k), it follows that (5) is an orthogonal basis for the column space of A.

If we now normalize these vectors to obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ for the column space, then Theorem 7.9.7 guarantees that we can extend this to an orthonormal basis

$$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$$

for R^n. Since the first k vectors in this set result from normalizing the vectors in (5), we have

$$\mathbf{u}_j = \frac{A\mathbf{v}_j}{\|A\mathbf{v}_j\|} = \frac{1}{\sqrt{\lambda_j}} A\mathbf{v}_j \qquad (1 \le j \le k)$$

which implies that

$$A\mathbf{v}_1 = \sqrt{\lambda_1}\mathbf{u}_1, \quad A\mathbf{v}_2 = \sqrt{\lambda_2}\mathbf{u}_2, \ldots, \quad A\mathbf{v}_k = \sqrt{\lambda_k}\mathbf{u}_k \tag{6}$$

Now let U be the orthogonal matrix

$$U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_n \end{bmatrix}$$

and let Σ be the diagonal matrix

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} & & & & & & \mathbf{0} \\ & \sqrt{\lambda_2} & & & & & \\ & & \ddots & & & & \\ & & & \sqrt{\lambda_k} & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ \mathbf{0} & & & & & & 0 \end{bmatrix} \tag{7}$$

Theorem 7.9.7 *If W is a nonzero subspace of R^n, then:*

(*a*) *Every orthogonal set of nonzero vectors in W can be enlarged to an orthogonal basis for W.*

(*b*) *Every orthonormal set in W can be enlarged to an orthonormal basis for W.*

It follows from (2) and (4) that $A\mathbf{v}_j = \mathbf{0}$ for $j > k$, so

$$\begin{aligned} U\Sigma &= [\sqrt{\lambda_1}\mathbf{u}_1 \quad \sqrt{\lambda_2}\mathbf{u}_2 \quad \cdots \quad \sqrt{\lambda_k}\mathbf{u}_k \quad \mathbf{0} \quad \cdots \quad \mathbf{0}] \\ &= [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_k \quad A\mathbf{v}_{k+1} \quad \cdots \quad A\mathbf{v}_n] = AV \end{aligned}$$

which we can rewrite as $A = U\Sigma V^T$ using the orthogonality of V. ■

It is important to keep in mind that the positive entries on the main diagonal of Σ are not eigenvalues of A, but rather square roots of the nonzero eigenvalues of A^TA. These numbers are called the ***singular values*** of A and are denoted by

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_k = \sqrt{\lambda_k}$$

REMARK In the special case where the matrix A is invertible, it follows that $k = \text{rank}(A) = n$, so there are no zeros on the diagonal of Σ.

With this notation, the factorization obtained in the proof of Theorem 8.6.1 has the form

$$A = U\Sigma V^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \quad \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_n] \begin{bmatrix} \sigma_1 & & & & & & 0 \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \quad (8)$$

which is called the ***singular value decomposition*** of A (abbreviated SVD of A).*

The numbers denoted by $\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_k = \sqrt{\lambda_k}$

are called the ***singular values*** of A.

The vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are called ***left singular vectors*** of A

and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are called ***right singular vectors*** of A.

The following theorem is a restatement of Theorem 8.6.1 and spells out some of the results that were established in the course of proving that theorem.

Theorem 8.6.2 (***Singular Value Decomposition of a Square Matrix***) *If A is an $n \times n$ matrix of rank k, then A has a singular value decomposition $A = U\Sigma V^T$ in which:*

(*a*) $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ *orthogonally diagonalizes* A^TA.

(*b*) *The nonzero diagonal entries of Σ are*

$$\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of A^TA corresponding to the column vectors of V.

(*c*) *The column vectors of V are ordered so that* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(*d*) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i}A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(*e*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\operatorname{col}(A)$.

(*f*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_n\}$ *is an extension of* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *to an orthonormal basis for* R^n.

EXAMPLE 1 Singular Value Decomposition of a Square Matrix

Find the singular value decomposition of the matrix

$$A = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}$$

Solution The first step is to find the eigenvalues of the matrix

$$A^T A = \begin{bmatrix} \sqrt{3} & 0 \\ 2 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} = \begin{bmatrix} 3 & 2\sqrt{3} \\ 2\sqrt{3} & 7 \end{bmatrix}$$

The characteristic polynomial of $A^T A$ is

$$\lambda^2 - 10\lambda + 9 = (\lambda - 9)(\lambda - 1)$$

so the eigenvalues of $A^T A$ are $\lambda_1 = 9$ and $\lambda_2 = 1$, and the singular values of A are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{9} = 3, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1$$

We leave it for you to show that unit eigenvectors of A^TA corresponding to the eigenvalues $\lambda_1 = 9$ and $\lambda_2 = 1$ are

$$\mathbf{v}_1 = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

respectively. Thus,

$$\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{3} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix},$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2} A\mathbf{v}_2 = (1) \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

so

$$U = [\mathbf{u}_1 \quad \mathbf{u}_2] = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \quad \text{and} \quad V = [\mathbf{v}_1 \quad \mathbf{v}_2] = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$$

It now follows that the singular value decomposition of A is

$$\begin{array}{ccccc} \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix} & = & \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} & \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \\ A & = & U & \Sigma & V^T \end{array}$$

You may want to confirm the validity of this equation by multiplying out the matrices on the right side. ■

SINGULAR VALUE DECOMPOSITION OF SYMMETRIC MATRICES

A symmetric matrix A has both an eigenvalue decomposition $A = PDP^T$ and a singular value decomposition $A = U\Sigma V^T$, so it is reasonable to ask what relationship, if any, might exist between the two. To answer this question, suppose that A has rank k and that the nonzero eigenvalues of A are ordered so that

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_k| > 0$$

In the case where A is symmetric we have $A^TA = A^2$, so the eigenvalues of A^TA are the squares of the eigenvalues of A. Thus, the nonzero eigenvalues of A^TA in nonincreasing order are

$$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_k^2 > 0$$

and the singular values of A in nonincreasing order are

$$\sigma_1 = \sqrt{\lambda_1^2} = |\lambda_1|, \quad \sigma_2 = \sqrt{\lambda_2^2} = |\lambda_2|, \ldots, \quad \sigma_k = \sqrt{\lambda_k^2} = |\lambda_k|$$

This shows that the *singular values of a symmetric matrix A are the absolute values of the nonzero eigenvalues of A*; and it also shows that if A is a symmetric matrix with nonnegative eigenvalues, then the singular values of A are the same as its nonzero eigenvalues.

EXAMPLE 2 Obtaining a Singular Value Decomposition from an Eigenvalue Decomposition

It follows from the computations in Example 2 of Section 8.3 that the symmetric matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$$

has the eigenvalue decomposition

$$A = PDP^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

$$A = PDP^T = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}$$

We can find a singular value decomposition of A using the following procedure to "shift" the negative sign from the diagonal factor to the second orthogonal factor:

$$A = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} =$$

change of signal

$$= \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = U\Sigma V^T$$

Alternatively, we could have shifted the negative sign to the first orthogonal factor (verify). This technique works for any *symmetric* matrix. ■

Theorem 8.6.3 **(*Polar Decomposition*)** *If A is an $n \times n$ matrix of rank k, then A can be factored as*

$$A = PQ \tag{9}$$

where P is an $n \times n$ positive semidefinite matrix of rank k, and Q is an $n \times n$ orthogonal matrix. Moreover, if A is invertible (rank n), then there is a factorization of form (9) in which P is positive definite.

REMARK A factorization $A = PQ$ in which Q is orthogonal and P is positive semidefinite is called a ***polar decomposition**** of A.

Proof Rewrite the singular value decomposition of A as

$$A = U\Sigma V^T = U\Sigma U^T UV^T = (U\Sigma U^T)(UV^T) = PQ \tag{10}$$

The matrix $Q = UV^T$ is orthogonal because it is a product of orthogonal matrices (Theorem 6.2.3), and the matrix $P = U\Sigma U^T$ is symmetric (verify).

Theorem 6.2.3

⋮

(c) *A product of orthogonal matrices is orthogonal.*

Also, the matrices Σ and $P = U\Sigma U^T$ are orthogonally similar, so they have the same rank and same eigenvalues. This implies that P has rank k and that its eigenvalues are nonnegative (since this is true of Σ).

Thus, P is a positive semidefinite matrix of rank k (see the remark following Theorem 8.4.3). Furthermore, if A is invertible, then there are no zeros on the diagonal of Σ (see the remark preceding Example 1), so the eigenvalues of P are positive, which means that P is positive definite. ■

Theorem 8.4.3 *If A is a symmetric matrix, then:*

(*a*) $\mathbf{x}^T A\mathbf{x}$ *is positive definite if and only if all eigenvalues of A are positive.*

(*b*) $\mathbf{x}^T A\mathbf{x}$ *is negative definite if and only if all eigenvalues of A are negative.*

(*c*) $\mathbf{x}^T A\mathbf{x}$ *is indefinite if and only if A has at least one positive eigenvalue and at least one negative eigenvalue.*

REMARK The three classifications in Definition 8.4.2 do not exhaust all of the possibilities. For example, a quadratic form for which $\mathbf{x}^T A\mathbf{x} \geq 0$ if $\mathbf{x} \neq \mathbf{0}$ is called ***positive semidefinite***, and one for which $\mathbf{x}^T A\mathbf{x} \leq 0$ if $\mathbf{x} \neq \mathbf{0}$ is called ***negative semidefinite***.

By adjusting the proof of Theorem 8.4.3 appropriately, one can prove that $\mathbf{x}^T A\mathbf{x}$ is positive semidefinite if and only if all eigenvalues of A are nonnegative and is negative semidefinite if and only if all eigenvalues of A are nonpositive.

EXAMPLE 3 Polar Decomposition

Find a polar decomposition of the matrix

$$A = \begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}$$

and interpret it geometrically.

Solution We found a singular value decomposition of A in Example 1. Using the matrices U, V, and Σ in that example and the expressions for P and Q in Formula (10) we obtain

$$P = U\Sigma U^T = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} = \begin{bmatrix} \frac{5}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{3}{2} \end{bmatrix}$$

and

$$Q = UV^T = \begin{bmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$$

Thus, a polar decomposition of A is

$$\underset{A}{\begin{bmatrix} \sqrt{3} & 2 \\ 0 & \sqrt{3} \end{bmatrix}} = \underset{P}{\begin{bmatrix} \frac{5}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{3}{2} \end{bmatrix}} \underset{Q}{\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}}$$

To understand what this equation says geometrically, let us rewrite it as

$$\begin{bmatrix} 1 & \frac{2}{\sqrt{3}} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{3} \end{bmatrix} = \begin{bmatrix} \frac{5}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{3}{2} \end{bmatrix}\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \tag{11}$$

$$A \text{ (factored)} \qquad = \qquad P \qquad\qquad Q$$

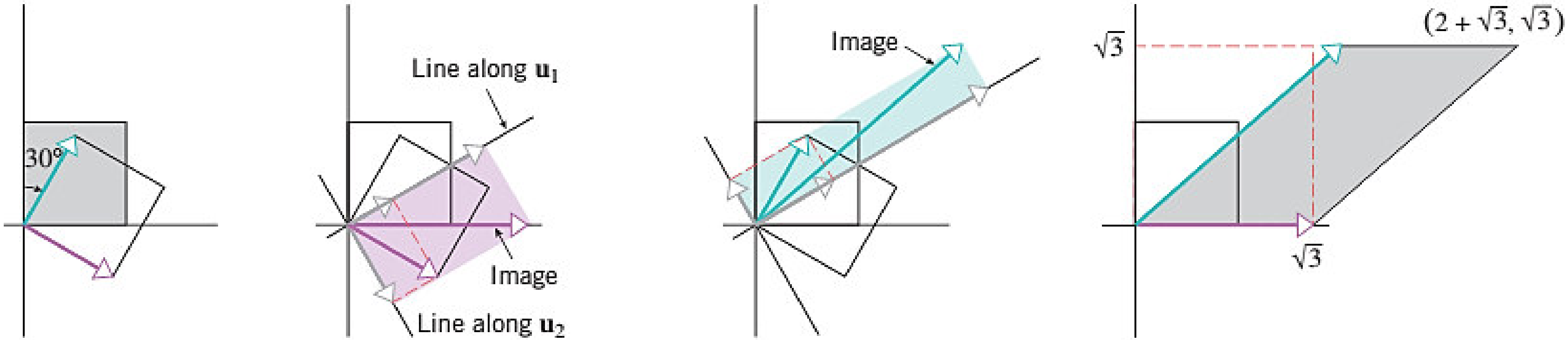
The right side of this equation tells us that multiplication by A is the same as multiplication by Q followed by multiplication by P.

In the exercises we will ask you to show that the multiplication by the orthogonal matrix Q produces a rotation about the origin through an angle of $-30°$ (or $330°$) and that the multiplication by the symmetric matrix P stretches R^2 by a factor of $\lambda_1 = 3$ in the direction of its unit eigenvector $\mathbf{u}_1 = (\sqrt{3}/2, 1/2)$ and by a factor of $\lambda_2 = 1$ in the direction of its unit eigenvector $\mathbf{u}_2 = (-1/2, \sqrt{3}/2)$ (i.e., no stretching).

On the other hand, the left side of (11) tells us that multiplication by A produces a dilation of factor $\sqrt{3}$ followed by a shear of factor $2/\sqrt{3}$ in the x-direction.

Geometrical interpretation.

Thus, the dilation followed by the shear must have the same effect as the rotation followed by the expansions along the eigenvectors (Figure 8.6.1). ■



Unit square rotated $-30°$

Image of purple edge obtained by projecting onto the lines along $\mathbf{u}_1$ and $\mathbf{u}_2$, scaling by factor 3 in the direction of $\mathbf{u}_1$, by factor 1 in the direction of $\mathbf{u}_2$, then adding.

Image of green edge obtained by projecting onto the lines along $\mathbf{u}_1$ and $\mathbf{u}_2$, scaling by factor 3 in the direction of $\mathbf{u}_1$, by factor 1 in the direction of $\mathbf{u}_2$, then adding.

The rotation and scalings produce the same image of the unit square as dilating by a factor of $\sqrt{3}$ and then shearing by a factor of $2/\sqrt{3}$ in the x-direction.

Figure 8.6.1

Graphical effects in different vectors.

SINGULAR VALUE DECOMPOSITION OF NONSQUARE MATRICES

Thus far we have focused on singular value decompositions of square matrices. However, the real power of singular value decomposition rests with the fact that it can be extended to general $m \times n$ matrices. To make this extension we define the ***main diagonal*** of an $m \times n$ matrix $A = [a_{ij}]$ to be the line of entries for which $i = j$. In the case of a square matrix, this line runs from the upper left corner to the lower right corner, but if $n > m$ or $m > n$, then the main diagonal is as pictured in Figure 8.6.2.

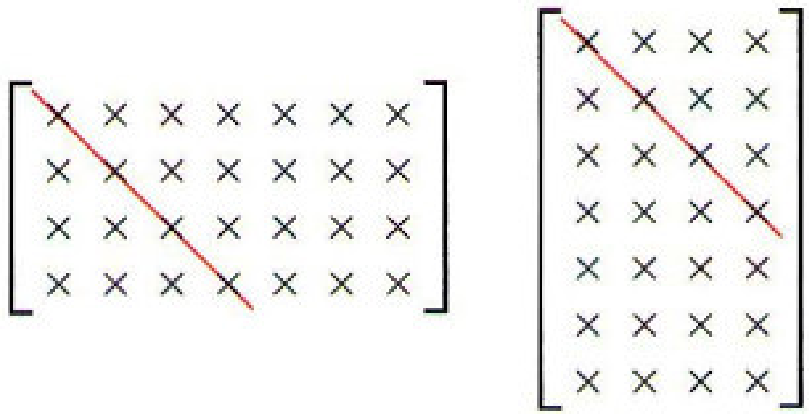


Figure 8.6.2 Main diagonal

If A is an $m \times n$ matrix, then $A^T A$ is an $n \times n$ symmetric matrix and hence has an eigenvalue decomposition, just as in the case where A is square. Except for appropriate size adjustments to account for the possibility that $n > m$ or $m < n$, the proof of Theorem 8.6.1 carries over without change and yields the following generalization of Theorem 8.6.2.

Theorem 8.6.4 (***Singular Value Decomposition of a General Matrix***) *If A is an $m \times n$ matrix of rank k, then A can be factored as*

$$A = U\Sigma V^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \mid \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & O_{k\times(n-k)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & O_{(m-k)\times k} & & & O_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array}\right] \tag{12}$$

in which U, Σ, and V have sizes $m \times m$, $m \times n$, and $n \times n$, respectively, and in which:

$$A = U\Sigma V^T = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \mid \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_m] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & O_{k\times(n-k)} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & O_{(m-k)\times k} & & & O_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array}\right] \tag{12}$$

(*a*) $V = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n]$ *orthogonally diagonalizes* A^TA.

(*b*) *The nonzero diagonal entries of* Σ *are* $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$, *where* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *are the nonzero eigenvalues of* A^TA *corresponding to the column vectors of* V.

(*c*) *The column vectors of* V *are ordered so that* $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.

(*d*) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i}A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(*e*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\mathrm{col}(A)$.

(*f*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ *is an extension of* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *to an orthonormal basis for* R^m.

EXAMPLE 4 Singular Value Decomposition of a Matrix That Is Not Square

Find the singular value decomposition of the matrix

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

Solution The first step is to find the eigenvalues of the matrix

$$A^TA = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

The characteristic polynomial of A^TA is

$$\lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

so the eigenvalues of A^TA are $\lambda_1 = 3$ and $\lambda_2 = 1$ and the singular values of A in order of decreasing size are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1$$

We leave it for you to show that unit eigenvectors of A^TA corresponding to the eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$ are

$$\mathbf{v}_1 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$$

respectively. These are the column vectors of V, and

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{\sqrt{3}}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = (1)\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix}$$

are two of the three column vectors of U.

Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal, as expected. We could extend the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ to an orthonormal basis for R^3 using the method of Example 2 of Section 7.4 and the Gram–Schmidt process directly.

However, the computations will be easier if we first remove the messy radicals by multiplying $\mathbf{u}_1$ and $\mathbf{u}_2$ by appropriate scalars. Thus, we will look for a unit vector $\mathbf{u}_3$ that is orthogonal to

$$\sqrt{6}\,\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \sqrt{2}\,\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

To satisfy these two orthogonality conditions, the vector $\mathbf{u}_3$ must be a solution of the homogeneous linear system

$$\begin{bmatrix} 2 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

We leave it for you to show that a general solution of this system is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = t \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

Normalizing the vector on the right yields

$$\mathbf{u}_3 = \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$$

Thus, the singular value decomposition of A is

$$\underset{A}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}} \underset{=}{=} \underset{U}{\begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}}$$

You may want to confirm the validity of this equation by multiplying out the matrices on the right side. ■

SINGULAR VALUE DECOMPOSITION AND THE FUNDAMENTAL SPACES OF A MATRIX

Theorem 8.6.5 *If A is an $m \times n$ matrix with rank k, and if $A = U\Sigma V^T$ is the singular value decomposition given in Formula* (12), *then:*

(*a*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\operatorname{col}(A)$.

(*b*) $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ *is an orthonormal basis for* $\operatorname{col}(A)^\perp = \operatorname{null}(A^T)$.

(*c*) $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ *is an orthonormal basis for* $\operatorname{row}(A)$.

(*d*) $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ *is an orthonormal basis for* $\operatorname{row}(A)^\perp = \operatorname{null}(A)$.

Proofs (a) and (b) We already know from Theorem 8.6.4 that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is a basis for $\operatorname{col}(A)$ and that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is an extension of that basis to an orthonormal basis for R^m. Since each of the vectors in the set $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ is orthogonal to each of the vectors in the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$, it follows that each of the vectors in $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ is orthogonal to $\operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} = \operatorname{col}(A)$. Thus, $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal set of $m-k$ vectors in $\operatorname{col}(A)^\perp = \operatorname{null}(A^T)$. But the dimension of $\operatorname{null}(A^T)$ is $m-k$ [see Formula (5) of Section 7.5], so $\{\mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ must be an orthonormal basis for $\operatorname{null}(A^T)$.

Specifically, if A is an $m \times n$ matrix with rank k, then

$$\begin{aligned} &\dim(\operatorname{row}(A)) = k, \quad \dim(\operatorname{null}(A)) = n - k \\ &\dim(\operatorname{col}(A)) = k, \quad \dim(\operatorname{null}(A^T)) = m - k \end{aligned} \tag{5}$$

Proofs (c) and (d) The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ form an orthonormal set of eigenvectors of A^TA and are ordered so that the corresponding eigenvalues of A^TA (all of which are nonnegative) are in the nonincreasing order

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

$(A^TA)\mathbf{v}_{k+1} = \lambda_{k+1}\mathbf{v}_{k+1} = 0$,
since $\lambda_{k+1} = 0$.

We know from Theorem 8.6.4 that the first k of these eigenvalues are positive and the subsequent $n-k$ are zero. Thus, $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal set of $n-k$ vectors in the null space of A^TA, which is the same as the null space of A (Theorem 7.5.8). Since the dimension of null(A) is $n-k$ [see Formula (5) of Section 7.5], it follows that $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for null(A).

Moreover, since each of the vectors in the set $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is orthogonal to each of the vectors in the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, it follows that each of the vectors in the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is orthogonal to span$\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\} =$ null(A). Thus, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthonormal set of k vectors in null$(A)^\perp =$ row(A). But row(A) has dimension k, so $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ must be an orthonormal basis for row(A). ■

Theorem 7.5.8 *If A is an $m \times n$ matrix, then:*

(*a*) *A and A^TA have the same null space.*

(*b*) *A and A^TA have the same row space.*

(*c*) *A^T and A^TA have the same column space.*

(*d*) *A and A^TA have the same rank.*