can then be used to store the entries of L, thereby reducing the amount of memory required to solve the system.

- If A is a large matrix consisting mostly of zeros, and if the nonzero entries are concentrated in a "band" around the main diagonal, then there are techniques that can be used to reduce the cost of LU-decomposition, giving it an advantage over Gauss–Jordan elimination.

Exercise Set 9.3

1. A certain computer can execute 10 gigaflops per second. Use Formula (5) to find the time required to solve the system using Gauss–Jordan elimination.

a. A system of 1000 equations in 1000 unknowns.

b. A system of 10,000 equations in 10,000 unknowns.

c. A system of 100,000 equations in 100,000 unknowns.

2. A certain computer can execute 100 gigaflops per second. Use Formula (5) to find the time required to solve the system using Gauss–Jordan elimination.

a. A system of 10,000 equations in 10,000 unknowns.

b. A system of 100,000 equations in 100,000 unknowns.

c. A system of 1,000,000 equations in 1,000,000 unknowns.

3. A certain computer can execute 70 gigaflops per second. Use **Table 1** to estimate the time required to perform the following operations on the invertible $10{,}000 \times 10{,}000$ matrix A.

a. Execute the forward phase of Gauss–Jordan elimination.

b. Execute the backward phase of Gauss–Jordan elimination.

c. LU-decomposition of A.

d. Find A^{-1} by reducing $[A \mid I]$ to $[I \mid A^{-1}]$.

4. The IBM Sequoia computer can operate at speeds in excess of 16 petaflops per second (1 petaflop $= 10^{15}$ flops). Use Table 1 to estimate the time required to perform the following operations on an invertible $100{,}000 \times 100{,}000$ matrix A.

a. Execute the forward phase of Gauss–Jordan elimination.

b. Execute the backward phase of Gauss–Jordan elimination.

c. LU-decomposition of A.

d. Find A^{-1} by reducing $[A \mid I]$ to $[I \mid A^{-1}]$.

5. **a.** Approximate the time required to execute the forward phase of Gauss–Jordan elimination for a system of 100,000 equations in 100,000 unknowns using a computer that can execute 1 gigaflop per second. Do the same for the backward phase. (See Table 1.)

b. How many gigaflops per second must a computer be able to execute to find the LU-decomposition of a matrix of size $10{,}000 \times 10{,}000$ in less than 0.5 s? (See Table 1.)

6. About how many teraflops per second must a computer be able to execute to find the inverse $100{,}000 \times 100{,}000$ matrix in less than 0.5 s? (1 teraflop $= 10^{12}$ flops.)

In Exercises 7–10, suppose A and B are $n \times n$ matrices and c is a real number.

7. How many flops are required to compute cA?

8. How many flops are required to compute $A + B$?

9. How many flops are required to compute AB?

10. If A is a diagonal matrix and k is a positive integer, how many flops are required to compute A^k?

9.4 Singular Value Decomposition

In this section we will discuss an extension of the diagonalization theory for $n \times n$ symmetric matrices to general $m \times n$ matrices. The results that we will develop in this section have applications to compression, storage, and transmission of digitized information and form the basis for many of the best computational algorithms that are currently available for solving linear systems.

Decompositions of Square Matrices

We saw in Formula (2) of Section 7.2 that every symmetric matrix A with real entries can be expressed as

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

where P is an orthogonal matrix whose columns are 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).

If an $n \times n$ matrix A is not symmetric, then it does not have an eigenvalue decomposition, but it does have a ***Hessenberg decomposition***

$$A = PHP^T$$

in which P is an orthogonal matrix and H is in upper Hessenberg form (Theorem 7.2.4). Moreover, if A has real eigenvalues, then it has a ***Schur decomposition***

$$A = PSP^T$$

in which P is an orthogonal matrix and S is upper triangular (Theorem 7.2.3).

The eigenvalue, Hessenberg, and Schur decompositions are important in numerical algorithms not only because the matrices D, H, and S have simpler forms than A, but also because the orthogonal matrices that appear in these factorizations do not magnify roundoff error. To see why this is so, suppose that $\hat{\mathbf{x}}$ is a column vector whose entries are known exactly and that

$$\mathbf{x} = \hat{\mathbf{x}} + \mathbf{e}$$

is the vector that results when roundoff error is present in the entries of $\hat{\mathbf{x}}$. If P is an orthogonal matrix, then the length-preserving property of orthogonal transformations implies that

$$\|P\mathbf{x} - P\hat{\mathbf{x}}\| = \|\mathbf{x} - \hat{\mathbf{x}}\| = \|\mathbf{e}\|$$

which tells us that the error in approximating $P\hat{\mathbf{x}}$ by $P\mathbf{x}$ has the same magnitude as the error in approximating $\hat{\mathbf{x}}$ by $\mathbf{x}$.

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

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

in which P is invertible but not necessarily orthogonal, or one might look for decompositions 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 decompositions in which J is either diagonal or a certain kind of block diagonal matrix, called a ***Jordan canonical form*** in honor of the French mathematician Camille Jordan (see p. 538). 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. In this section we will focus on the second path.

Singular Values

Since matrix products of the form A^TA will play an important role in our work, we will begin with two basic theorems about them.

Theorem 9.4.1

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.

We will prove part (*a*) and leave the remaining proofs for the exercises.

Proof (a) We must show that every solution of $A\mathbf{x} = \mathbf{0}$ is a solution of $A^TA\mathbf{x} = \mathbf{0}$, and conversely. If $\mathbf{x}_0$ is any solution of $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x}_0$ is also a solution of $A^TA\mathbf{x} = \mathbf{0}$ since

$$A^TA\mathbf{x}_0 = A^T(A\mathbf{x}_0) = A^T\mathbf{0} = \mathbf{0}$$

Conversely, if $\mathbf{x}_0$ is any solution of $A^TA\mathbf{x} = \mathbf{0}$, then $\mathbf{x}_0$ is in the null space of A^TA and hence is orthogonal to all vectors in the row space of A^TA by part (s) of Theorem 8.2.4. However, A^TA is symmetric, so $\mathbf{x}_0$ is also orthogonal to every vector in the column space of A^TA. In particular, $\mathbf{x}_0$ must be orthogonal to the vector $(A^TA)\mathbf{x}_0$; that is,

$$\mathbf{x}_0 \cdot (A^TA)\mathbf{x}_0 = 0$$

Using the first row in Table 1 of Section 3.2 and properties of the transpose operation we can rewrite this as

$$\mathbf{x}_0^T(A^TA)\mathbf{x}_0 = (A\mathbf{x}_0)^T(A\mathbf{x}_0) = (A\mathbf{x}_0)\cdot(A\mathbf{x}_0) = \|A\mathbf{x}_0\|^2 = 0$$

which implies that $A\mathbf{x}_0 = \mathbf{0}$, thereby proving that $\mathbf{x}_0$ is a solution of $A\mathbf{x}_0 = \mathbf{0}$. ■

Theorem 9.4.2

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

(a) A^TA is orthogonally diagonalizable.

(b) The eigenvalues of A^TA are nonnegative real numbers.

Proof (a) The matrix A^TA, being symmetric, is orthogonally diagonalizable by Theorem 7.2.1.

Proof (b) Since A^TA is orthogonally diagonalizable, there is an orthonormal basis for R^n consisting of eigenvectors of A^TA, say $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. If we let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the corresponding eigenvalues, then for $1 \le i \le n$ we have

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= A\mathbf{v}_i \cdot A\mathbf{v}_i = \mathbf{v}_i \cdot A^TA\mathbf{v}_i \qquad \textbf{[Formula (26) of Section 3.2]} \\ &= \mathbf{v}_i \cdot \lambda_i\mathbf{v}_i = \lambda_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i \end{aligned}$$

It follows from this relationship that $\lambda_i \ge 0$. ■

We will assume throughout this section that the eigenvalues of A^TA are named so that

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$$

and hence that

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n \ge 0$$

Definition 1

If A is an $m \times n$ matrix, and if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A^TA, then the numbers

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \ldots, \quad \sigma_n = \sqrt{\lambda_n}$$

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

EXAMPLE 1 | Singular Values

Find the singular values of the matrix

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

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

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

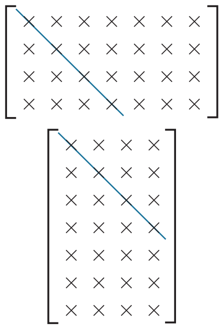
so the eigenvalues of A^TA are $\lambda_1 = 3$ and $\lambda_2 = 1$ and the singular values of A in order of decreasing size are

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

Singular Value Decomposition

Before turning to the main result in this section, we will find it useful to extend the notion of a "main diagonal" to matrices that are not square. We define the ***main diagonal*** of an $m \times n$ matrix to be the line of entries shown in **Figure 9.4.1**—it starts at the upper left corner and extends diagonally as far as it can go. We will refer to the entries on the main diagonal as the ***diagonal entries***.

We are now ready to consider the main result in this section, which is concerned with a specific way of factoring a general $m \times n$ matrix A. This factorization, called ***singular value decomposition*** (abbreviated SVD) will be given in two forms, a brief form that captures the main idea, and an expanded form that spells out the details. The proof is given at the end of this section.



Main diagonals

FIGURE 9.4.1

Theorem 9.4.3

Singular Value Decomposition (Brief Form)

If A is an $m \times n$ matrix of rank k, then A can be expressed in the form $A = U\Sigma V^T$, where Σ has size $m \times n$ and can be expressed in partitioned form as

$$\Sigma = \left[\begin{array}{c|c} D & 0_{k\times(n-k)} \\ \hline 0_{(m-k)\times k} & 0_{(m-k)\times(n-k)} \end{array}\right]$$

in which D is a diagonal $k \times k$ matrix whose successive entries are the first k singular values of A in nonincreasing order, U is an $m \times n$ orthogonal matrix, and V is an $n \times n$ orthogonal matrix.

Historical Note

Harry Bateman (1882–1946)

The term *singular value* is apparently due to the British-born mathematician Harry Bateman, who used it in a research paper published in 1908. Bateman emigrated to the United States in 1910, teaching at Bryn Mawr College, Johns Hopkins University, and finally at the California Institute of Technology. Interestingly, he was awarded his Ph.D. in 1913 by Johns Hopkins at which point in time he was already an eminent mathematician with 60 publications to his name.

[Image: Courtesy of the Archives, California Institute of Technology]

Theorem 9.4.4

Singular Value Decomposition (Expanded Form)

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

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

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

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

EXAMPLE 2 | Singular Value Decomposition if A Is Not Square

Find a singular value decomposition of the matrix

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

Solution We showed in Example 1 that the eigenvalues of A^TA are $\lambda_1 = 3$ and $\lambda_2 = 1$ and that the corresponding singular values of A are $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$. We leave it for you to verify that

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

are eigenvectors corresponding to λ_1 and λ_2, respectively, and that $V = [\mathbf{v}_1 \mid \mathbf{v}_2]$ orthogonally diagonalizes A^TA. From part (*d*) of Theorem 9.4.4, the vectors

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{\sqrt{3}}{3}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} \\ \frac{\sqrt{6}}{6} \\ \frac{\sqrt{6}}{6} \end{bmatrix}$$

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

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. 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

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

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.

OPTIONAL: We conclude this section with an optional proof of Theorem 9.4.4.

Proof of Theorem 9.4.4 For notational simplicity we will prove this theorem in the case where A is an $n \times n$ matrix. To modify the argument for an $m \times n$ matrix you need only make the notational adjustments required to account for the possibility that $m > n$ or $n > m$.

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

$$A^TA = VDV^T$$

in which the column vectors of

$$V = [\mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n]$$

are unit eigenvectors of A^TA, and D is a diagonal matrix whose successive diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A^TA corresponding in succession to the column vectors of V. Since A is assumed to have rank k, it follows from Theorem 9.4.1 that A^TA also has rank k. It follows as well that D has rank k, since it is similar to A^TA and rank is a similarity invariant. Thus, the diagonal matrix D can be expressed in the form

$$D = \begin{bmatrix} \lambda_1 & & & & & & 0 \\ & \lambda_2 & & & & & \\ & & \ddots & & & & \\ & & & \lambda_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix} \tag{2}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$. Now let us consider the set of image vectors

$$\{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_n\} \tag{3}$$

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

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

Moreover, the first k vectors in (3) are nonzero since we showed in the proof of Theorem 9.4.2(b) that $\|A\mathbf{v}_i\|^2 = \lambda_i$ for $i = 1, 2, \ldots, n$, and we have assumed that the first k diagonal entries in (2) are positive. Thus,

$$S = \{A\mathbf{v}_1, A\mathbf{v}_2, \ldots, A\mathbf{v}_k\}$$

is an orthogonal set of *nonzero* vectors in the column space of A. But the column space of A has dimension k since

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

and hence S, being a linearly independent set of k vectors, must be an orthogonal basis for col(A). If we now normalize the vectors in S, we will obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ for col(A) in which

$$\mathbf{u}_i = \frac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \frac{1}{\sqrt{\lambda_i}}A\mathbf{v}_i \qquad (1 \leq i \leq k)$$

or, equivalently, in which

$$A\mathbf{v}_1 = \sqrt{\lambda_1}\mathbf{u}_1 = \sigma_1\mathbf{u}_1, \quad A\mathbf{v}_2 = \sqrt{\lambda_2}\mathbf{u}_2 = \sigma_2\mathbf{u}_2, \ldots, \quad A\mathbf{v}_k = \sqrt{\lambda_k}\mathbf{u}_k = \sigma_k\mathbf{u}_k \tag{4}$$

It follows from Theorem 6.3.6 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. Now let U be the orthogonal matrix

$$U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k \quad \mathbf{u}_{k+1} \quad \cdots \quad \mathbf{u}_n]$$

and let Σ be the diagonal matrix

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & 0 \\ & \sigma_2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_k & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & & & & 0 \end{bmatrix}$$

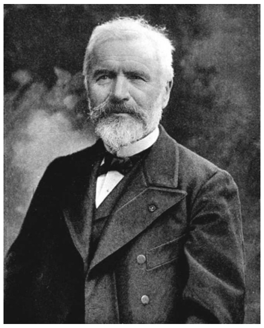
It follows from (4), and the fact that $A\mathbf{v}_i = 0$ for $i > k$, that

$$\begin{aligned} U\Sigma &= [\sigma_1\mathbf{u}_1 \quad \sigma_2\mathbf{u}_2 \quad \cdots \quad \sigma_k\mathbf{u}_k \quad 0 \quad \cdots \quad 0] \\ &= [A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_k \quad A\mathbf{v}_{k+1} \quad \cdots \quad A\mathbf{v}_n] \\ &= AV \end{aligned}$$

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

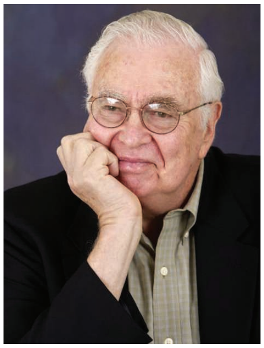
Historical Note

Eugenio Beltrami (1835–1900)



Camille Jordan (1838–1922)

Herman Klaus Weyl (1885–1955)



Gene H. Golub (1932–2007)

The theory of singular value decompositions can be traced back to the work of five people: the Italian mathematician Eugenio Beltrami, the French mathematician Camille Jordan, the English mathematician James Sylvester (see p. 36), and the German mathematicians Erhard Schmidt (see p. 369) and the mathematician Herman Weyl. More recently, the pioneering efforts of the American mathematician Gene Golub produced a stable and efficient algorithm for computing it. Beltrami and Jordan were the progenitors of the decomposition—Beltrami gave a proof of the result for real, invertible matrices with distinct singular values in 1873. Subsequently, Jordan refined the theory and eliminated the unnecessary restrictions imposed by Beltrami. Sylvester, apparently unfamiliar with the work of Beltrami and Jordan, rediscovered the result in 1889 and suggested its importance. Schmidt was the first person to show that the singular value decomposition could be used to approximate a matrix by another matrix with lower rank, and, in so doing, he transformed it from a mathematical curiosity to an important practical tool. Weyl showed how to find the lower rank approximations in the presence of error.

[Images: http://www-history.mcs.st-andrews.ac.uk/history/PictDisplay/Beltrami.html (Beltrami); The Granger Collection, New York (Jordan); Courtesy Electronic Publishing Services, Inc., New York City (Weyl); Courtesy of Hector Garcia-Molina (Golub)]

Exercise Set 9.4

In Exercises 1–4, find the distinct singular values of A.

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

2. $A = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$

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

4. $A = \begin{bmatrix} \sqrt{2} & 0 \\ 1 & \sqrt{2} \end{bmatrix}$

In Exercises 5–12, find a singular value decomposition of A.

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

6. $A = \begin{bmatrix} -3 & 0 \\ 0 & -4 \end{bmatrix}$

7. $A = \begin{bmatrix} 4 & 6 \\ 0 & 4 \end{bmatrix}$

8. $A = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix}$