Chapter 15

# The Singular Value Decomposition

*You should be familiar with*

- Eigenvalues and eigenvectors
- Singular values
- Orthogonal matrices
- Matrix column space, row space, and null space
- The matrix 2-norm and Frobenius norm
- The matrix inverse

Matrix decompositions play a critical role in numerical linear algebra, and we have already see the *QR* decomposition, one of the great accomplishments in the field. The *singular value decomposition* (SVD) is also among the greatest results in linear algebra. Just like the *QR* decomposition, the SVD is a matrix decomposition that applies to any matrix, real, or complex. The SVD is a powerful tool for many matrix computations because it reveals a great deal about the structure of a matrix. A number of its many applications are listed in Section 15.7.2.

This chapter proves the SVD theorem but does not develop a useable algorithm for its computation. We will use the SVD from this point forward in the book and will see some of its powerful applications. Unfortunately, computing the SVD efficiently is quite difficult, and we will present two methods for its computation in Chapter 23. In the meantime, we will use the built-in MATLAB command `svd` to compute it.

## 15.1 THE SVD THEOREM

Recall that if $A$ is an $m \times n$ matrix, then $A^{\mathrm{T}}A$ is an $n \times n$ symmetric matrix with nonnegative eigenvalues (Lemma 7.4). The singular values of an $m \times n$ matrix are the square roots of the eigenvalues of $A^{\mathrm{T}}A$, and the 2-norm of a matrix is the largest singular value. The SVD factors $A$ into a product of two orthogonal matrices and a diagonal matrix of its singular values.

**Theorem 15.1.** *Let $A \in \mathbb{R}^{m\times n}$ be a matrix having r positive singular values, $m \geq n$. Then there exist orthogonal matrices $U \in R^{m\times m}, V \in \mathbb{R}^{n\times n}$, and a diagonal matrix $\tilde{\Sigma} \in \mathbb{R}^{m\times n}$such that*

$$\begin{aligned} A &= U\tilde{\Sigma}V^{\mathrm{T}} \\ \tilde{\Sigma} &= \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

*where $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the positive singular values of A.*

**Overview:**

*The proof is by construction. Build the $m \times n$ matrix $\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}$ by placing the positive singular values on the diagonal of $\Sigma$, so*

$$\Sigma = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}.$$

*Find an orthonormal basis $v_i$, $1 \leq i \leq n$ of eigenvectors of $A^{\mathrm{T}}A$. This can be done because $A^{\mathrm{T}}A$ is symmetric (Theorem 7.6, the spectral theorem). Then, $V = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$. Build the orthogonal matrix U using A, $v_i$, and $\sigma_i$.*

Numerical Linear Algebra with Applications. http://dx.doi.org/10.1016/B978-0-12-394435-1.00015-6

*Proof.* The matrix $A^{\mathrm{T}}A$ is symmetric and by Lemma 7.4, its eigenvalues are real and nonnegative. Listing the eigenvalues in descending order we obtain

$$\sigma_1^2 \geq \sigma_2^2 \geq \cdots \geq \sigma_r^2 \geq \sigma_{r+1}^2 \geq \cdots \geq \sigma_n^2 \geq 0.$$

Assume that the first $r$ eigenvalues are positive, and the eigenvalues $\sigma_{r+1}^2 = \sigma_{r+2}^2 = \cdots \sigma_n^2 = 0$. Let

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix},$$

where $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ is a diagonal matrix of singular values. The spectral theorem (Theorem 7.6) guarantees that $A^{\mathrm{T}}A$ has an orthonormal basis for $\mathbb{R}^n$ of eigenvectors. Let

$$V = \begin{bmatrix} v_1 & v_2 & \ldots & v_{n-1} & v_n \end{bmatrix},$$

and it turns out this is the matrix we are looking for. We now need to find $U$. The matrix $U$ must be orthogonal, so its columns will form a basis for $\mathbb{R}^m$. Let

$$u_i = \frac{Av_i}{\sigma_i}, \quad 1 \leq i \leq r.$$

These vectors are orthonormal since

$$\langle u_i, u_j \rangle = \frac{(Av_i)^{\mathrm{T}}(Av_j)}{\sigma_i \sigma_j} = \frac{(\sigma_i v_i)^{\mathrm{T}}(\sigma_j v_j)}{\sigma_i \sigma_j} = v_i^{\mathrm{T}} v_j = \langle v_i, v_j \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}.$$

If $r < m$, we still need $m - r$ additional vectors $\{ u_{r+1} \;\; u_{r+1} \;\; \ldots \;\; u_{m-1} \;\; u_m \}$ so that $\{ u_1 \;\; u_2 \;\; \ldots \;\; u_{m-1} \;\; u_m \}$ forms an orthonormal set. Beginning with

$$u_1, \; u_2, \; \ldots \; u_{r-1}, \; u_r$$

use a Gram-Schmidt algorithm step to add the standard basis vector $e_{r+1}$ to the set to obtain the orthonormal set

$$u_1, \; u_2, \; \ldots \; u_{r-1}, \; u_r, \; u_{r+1}\;.$$

Continue by adding in the same fashion $e_{r+2}, e_{r+3}, \ldots, e_m$ to obtain the basis

$$u_1, \; u_2, \; \ldots \; u_{m-1}, \; u_m$$

for $\mathbb{R}^m$ and the matrix $U = \begin{bmatrix} u_1 & u_2 & \ldots & u_{m-1} & u_m \end{bmatrix}$. Now we need to show that $A = U\tilde{\Sigma}V^{\mathrm{T}}$, or $U^{\mathrm{T}}AV = \tilde{\Sigma}$. Before we begin, note two things:

**a.** The vectors $v_1, \; v_2, \; \ldots \; , v_r$ are the nonzero eigenvectors of $A^{\mathrm{T}}A$, so $A^{\mathrm{T}}Av_i = 0, \; r+1 \leq i \leq n$. Multiply by $v_i^{\mathrm{T}}$ to get

$$v_i^{\mathrm{T}} A^{\mathrm{T}} A v_i = 0,$$

so

$$(Av_i)^{\mathrm{T}}(Av_i) = \|Av_i\|_2^2 = 0,$$

and $Av_i = 0, \; r+1 \leq i \leq n$.

**b.** Write $U^{\mathrm{T}} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \\ u_{r+1} \\ \vdots \\ u_m \end{bmatrix}$, where each $u_i$ is a row of $U^{\mathrm{T}}$. Since $u_1, u_2, \ldots, u_r$ are an orthonormal set, it follows that $U^{\mathrm{T}}u_i = e_i, \; 1 \leq i \leq r$, and so $\sigma_i U^{\mathrm{T}} u_i = \sigma_i e_i, \; 1 \leq i \leq r$.

Now continue.

$$U^{\mathrm{T}}AV = U^{\mathrm{T}}A \begin{bmatrix} v_1 & v_2 & \ldots & v_{n-1} & v_n \end{bmatrix} = U^{\mathrm{T}} \begin{bmatrix} Av_1 & Av_2 & \ldots & Av_{n-1} & Av_n \end{bmatrix} = \tag{15.1}$$

$$U^{\mathrm{T}} \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \ldots & \sigma_r u_r & Av_{r+1} & \ldots & Av_n \end{bmatrix} = \tag{15.2}$$

$$U^{\mathrm{T}}\begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r & 0 & \dots & 0 \end{bmatrix} = \tag{15.3}$$

$$\begin{bmatrix} \sigma_1 U^{\mathrm{T}} u_1 & \sigma_2 U^{\mathrm{T}} u_2 & \dots & \sigma_r U^{\mathrm{T}} u_r & 0 & \dots & 0 \end{bmatrix} = \tag{15.4}$$

$$\begin{bmatrix} \sigma_1 e_1 & \sigma_2 e_2 & \dots & \sigma_r e_r & 0 & \dots & 0 \end{bmatrix} = \tilde{\Sigma} \tag{15.5}$$

□

*Remark* 15.1.

**a.** There is no loss of generality in assuming that $m \geq n$, for if $m < n$, find the SVD for $A^{\mathrm{T}}$ and transpose back. We have

$$A^{\mathrm{T}} = U\tilde{\Sigma}V^{\mathrm{T}}$$

so

$$A = V\tilde{\Sigma}U^{\mathrm{T}},$$

and we have an SVD for $A$.

The columns of $U$ and $V$ are called the *left* and *right singular vectors*, respectively. The largest and smallest singular values are denoted, respectively, as $\sigma_{\max}$ and $\sigma_{\min}$.

**Example 15.1.** The matrix $A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$ has SVD

$$\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

We see that the columns of $U$ and $V$ have unit length since $U = I$, and

$$\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 = 1.$$

A simple calculation of inner products will show the columns of $U$ and $V$ are mutually orthogonal. ■

**Example 15.2.** Let $A = \begin{bmatrix} 1 & -1 & 3 \\ 1 & 0 & 1 \\ 1 & 2 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 2 & 1 & 1 \end{bmatrix}$. Here are SVDs for each matrix:

$$A = \begin{bmatrix} -0.9348 & 0.0194 & 0.3546 \\ 0.3465 & -0.2684 & -0.8988 \\ 0.0778 & -0.9631 & 0.2577 \end{bmatrix} \begin{bmatrix} 3.5449 & 0 & 0 \\ 0 & 2.3019 & 0 \\ 0 & 0 & 0.3676 \end{bmatrix} \begin{bmatrix} -0.3395 & 0.3076 & -0.8889 \\ -0.5266 & -0.8452 & -0.0913 \\ -0.7794 & 0.4371 & 0.4489 \end{bmatrix},$$

$$r = 3,\ \sigma_1 = 3.5449,\ \sigma_2 = 2.3019,\ \sigma_3 = 0.3676$$

$$B = \begin{bmatrix} -0.1355 & 0.8052 & -0.5774 \\ -0.6295 & -0.5199 & -0.5774 \\ -0.7651 & 0.2852 & 0.5774 \end{bmatrix} \begin{bmatrix} 3.1058 & 0 & 0 \\ 0 & 2.0867 & 0 \\ 0 & 0 & 0.0000 \end{bmatrix} \begin{bmatrix} -0.7390 & -0.2900 & -0.6081 \\ 0.4101 & 0.5226 & -0.7475 \\ 0.5345 & -0.8018 & -0.2673 \end{bmatrix},$$

$$r = 2,\ \sigma_1 = 3.1058,\ \sigma_2 = 2.0867$$

*Note*: The rank of $A$ is 3, and the rank of $B$ is 2. ■

**Example 15.3.** Consider the matrix

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

An SVD is

$$U = \begin{bmatrix} -0.5000 & -0.8660 & 0 & 0.0000 & 0 & 0.0000 & 0 \\ 0.0000 & 0.0000 & 1 & 0.0000 & 0 & 0.0000 & 0 \\ 0.5000 & -0.2887 & 0 & 0.8165 & 0 & 0.0000 & 0 \\ 0.0000 & 0.0000 & 0 & 0.0000 & -1 & 0.0000 & 0 \\ -0.5000 & 0.2887 & 0 & 0.4082 & 0 & 0.7071 & 0 \\ 0.5000 & -0.2887 & 0 & -0.4082 & 0 & 0.7071 & 0 \\ 0.0000 & 0.0000 & 0 & 0.0000 & 0 & 0.0000 & 1 \end{bmatrix},$$

$$\tilde{\Sigma} = \begin{bmatrix} 4.4721 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$V = \begin{bmatrix} -.4472 & -0.3651 & -0.6712 & 0.4614 & 0.0581 \\ -.4472 & -0.3651 & -0.0671 & -0.8074 & -0.1016 \\ -.4472 & 0.5477 & 0.0000 & -0.0883 & 0.7016 \\ -.4472 & -0.3651 & 0.7383 & 0.3460 & 0.0435 \\ -.4472 & 0.5477 & 0.0000 & 0.0883 & -0.7016 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 4.4721 \end{bmatrix}, \ \sigma_1 = 4.4721, \ r = 1$$

*Note*: The rank of $A$ is 1. ■

If the singular values of $A$ are placed in descending order in $\tilde{\Sigma}$, then $\tilde{\Sigma}$ is unique; however, $U$ and $V$ are not in general. In the proof of Theorem 15.1, we extended the orthonormal basis

$$\left\{ u_1 \ \ u_2 \ \ \ldots \ \ u_{r-1} \ \ u_r \right\}$$

to an orthonormal basis for $\mathbb{R}^m$. This can be done in many ways. For the matrix of Example 15.3, the matrices

$$U = \begin{bmatrix} -0.5000 & 0.8660 & 0.0000 & 0 & 0.0000 & 0.0000 & 0 \\ 0.0000 & 0.0000 & -1.0000 & 0 & 0.0000 & 0.0000 & 0 \\ 0.5000 & 0.2887 & 0.0000 & 0 & 0.5774 & -0.5774 & 0 \\ 0.0000 & 0.0000 & 0.0000 & 1 & 0.0000 & 0.0000 & 0 \\ -0.5000 & -0.2887 & 0.0000 & 0 & 0.7887 & 0.2113 & 0 \\ 0.5000 & 0.2887 & 0.0000 & 0 & 0.2113 & 0.7887 & 0 \\ 0.0000 & 0.0000 & 0.0000 & 0 & 0.0000 & 0.0000 & 1 \end{bmatrix}, \ V = \begin{bmatrix} -0.4472 & 0.3651 & 0.0000 & -0.5774 & -0.5774 \\ -0.4472 & 0.3651 & 0.0000 & -0.2113 & -0.7887 \\ -0.4472 & -0.5477 & 0.7071 & 0.0000 & 0.0000 \\ -0.4472 & 0.3651 & 0.0000 & 0.7887 & -0.2113 \\ -0.4472 & -0.5477 & -0.7071 & 0.0000 & 0.0000 \end{bmatrix}$$

also produce a valid SVD.

*Remark* 15.2. If $A$ is square and the $\sigma_i$ are distinct, then $u_i$ and $v_i$ are uniquely determined except for sign.

## 15.2 USING THE SVD TO DETERMINE PROPERTIES OF A MATRIX

The *rank* of a matrix is the number of linearly independent columns or rows. Notice that in Example 15.2, the matrix $A$ has three nonzero singular values, and the matrix $B$ has two. The matrix of Example 15.3 has only one nonzero singular value. The rank of the matrices is 3, 2, and 1, respectively. This is not an accident. The rank of a matrix is the number of nonzero singular values in $\tilde{\Sigma}$.

We now present a theorem dealing with rank that we have not needed until now. The result will allow us to show the relationship between the rank of a matrix and its singular values.

**Theorem 15.2.** *If $A$ is an $m \times n$ matrix, $X$ is an invertible $m \times m$ matrix, and $Y$ is an invertible $n \times n$ matrix, then* rank $(XAY)$ = rank $(A)$.

*Proof.* Since $X$ is invertible, it can be written as a product of elementary row matrices, so $X = E_k^{(X)}E_{k-1}^{(X)}\dots E_2^{(X)}E_1^{(X)}$. Similarly, $Y$ is a product of elementary row matrices, $Y = E_p^{(Y)}E_{p-1}^{(Y)}\dots E_2^{(Y)}E_1^{(Y)}$, and so

$$XAY = E_k^{(X)}E_{k-1}^{(X)}\dots E_2^{(X)}E_1^{(X)}AE_p^{(Y)}E_{p-1}^{(Y)}\dots E_2^{(Y)}E_1^{(Y)}.$$

The product of the elementary row matrices on the left performs elementary row operations on $A$, and this does not change the rank of $A$. The product of elementary row matrices on the left perform elementary column operations, which also do not alter rank. Thus, $\text{rank}\,(XAY) = \text{rank}\,(A)$. $\square$

**Theorem 15.3.** *The rank of a matrix A is the number of nonzero singular values.*

*Proof.* Let $A = U\tilde{\Sigma}V^{\mathrm{T}}$ be the SVD of $A$. Orthogonal matrices are invertible, so by Theorem 1.2,

$$\text{rank}\,(A) = \text{rank}\left(U\tilde{\Sigma}V^{\mathrm{T}}\right) = \text{rank}\left(\tilde{\Sigma}\right).$$

The rank of $\tilde{\Sigma}$ is r, since

$$\begin{bmatrix}\sigma_1 & 0 & 0 & \dots & 0\end{bmatrix}^{\mathrm{T}},\quad \begin{bmatrix}0 & \sigma_2 & 0 & \dots & 0\end{bmatrix}^{\mathrm{T}},\quad \begin{bmatrix}0 & 0 & \sigma_3 & \dots & 0\end{bmatrix}^{\mathrm{T}},\quad \begin{bmatrix}0 & 0 & \dots & \sigma_r & 0 & \dots & 0\end{bmatrix}^{\mathrm{T}}$$

is a basis for the column space of $\tilde{\Sigma}$. $\square$

From the components of the SVD, we can determine other properties of the original matrix. Recall that the *null space* of a matrix $A$, *written* null $(A)$, is the set of *vectors x* for which $Ax = 0$, and the *range* of $A$ is the set all linear combinations of the columns of $A$ (the column space of $A$). Let $u_i$, $1 \leq i \leq m$ and $v_i$ , $1 \leq i \leq n$ be the column vectors of $U$ and $V$, respectively. Then

$$Av_i = U\tilde{\Sigma}V^{\mathrm{T}}v_i.$$

The matrix $V^{\mathrm{T}}$ can be written as $\begin{bmatrix} v_1^{\mathrm{T}} \\ \vdots \\ v_i^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix}$, where the $v_i$ are the orthonormal columns of $V$. The product $V^{\mathrm{T}}v_i = \begin{bmatrix} v_1^{\mathrm{T}} \\ \vdots \\ v_i^{\mathrm{T}} \\ \vdots \\ v_n^{\mathrm{T}} \end{bmatrix} v_i = e_i$, where $e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$ is the $i$th standard basis vector in $\mathbb{R}^n$. Now,

$$\tilde{\Sigma}e_i = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & \sigma_i & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & \ddots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \sigma_r & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} e_i = \begin{bmatrix} 0 \\ \vdots \\ \sigma_i \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}, \; i \leq r,$$

and

$$U\begin{bmatrix} 0 \\ \vdots \\ \vdots \\ \sigma_i \\ \vdots \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1r} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2r} & \cdots & u_{2m} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & u_{rr} & \cdots & u_{rm} \\ \vdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mr} & \cdots & u_{mm} \end{bmatrix}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_i \\ \vdots \\ 0 \end{bmatrix} = \sigma_i u_i, \quad 1 \le i \le r.$$

For $v_i$, $r+1 \le i \le m$, we have

$$Av_i = U\widetilde{\Sigma}e_i = \begin{bmatrix} u_{11} & \cdots & u_{1i} & \cdots & u_{1m} \\ u_{21} & \cdots & u_{2i} & \cdots & u_{2m} \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \vdots & \cdots & u_{ii} & \cdots & u_{im} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{m1} & \cdots & u_{mi} & \cdots & u_{mm} \end{bmatrix}\begin{bmatrix} \sigma_1 & \cdots & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \sigma_r & \ddots & \vdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & 0 & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & \vdots \\ \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = 0$$

and $Av_i = 0$, $r+1 \le i \le m$.

In summary, we have

$$Av_i = \sigma_i u_i, \quad \sigma_i \ne 0,\ 1 \le i \le r$$

$$Av_i = 0, \quad r+1 \le i \le n$$

Since $U$ and $V$ are orthogonal matrices, all $u_i$ and $v_i$ are linearly independent. For $1 \le i \le r$, $Av_i = \sigma_i u_i$, $\sigma_i \ne 0$, and $u_i$, $1 \le i \le r$ is in the range of A. Since by Theorem 15.2 the rank of $A$ is $r$, the $u_i$ are a basis for the range of A. For $r+1 \le i \le n$, $Av_i = 0$, so $v_i$ is in null $(A)$. Since rank $(A)$ + nullity $(A) = n$, nullity $(A) = n - r$. There are $n - (r+1) + 1 = n - r$ orthogonal vectors $v_i$, so the $v_i$, $r+1 \le i \le n$, are a basis for the null space of $A$.

**Example 15.4.** Let $B = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 2 & 1 & 1 \end{bmatrix}$ be the matrix in Example 15.2. From the SVD, the vectors $\begin{bmatrix} -0.1355 \\ -0.6295 \\ -0.7651 \end{bmatrix}$ and $\begin{bmatrix} 0.8052 \\ -0.5199 \\ 0.2852 \end{bmatrix}$ are a basis for the range of $B$, and the vector $\begin{bmatrix} 0.5345 \\ -0.8018 \\ -0.2673 \end{bmatrix}$ is a basis for the null space of $B$. Remember when looking at the decomposition of $B$, $V^{\mathrm{T}}$ appears, not $V$. ■

### 15.2.1 The Four Fundamental Subspaces of a Matrix

There are four fundamental subspaces associated with an $m \times n$ matrix. We have seen two, the range and the null space, and have determined how compute a basis for each using the SVD. The other two subspaces are the range and the null space of

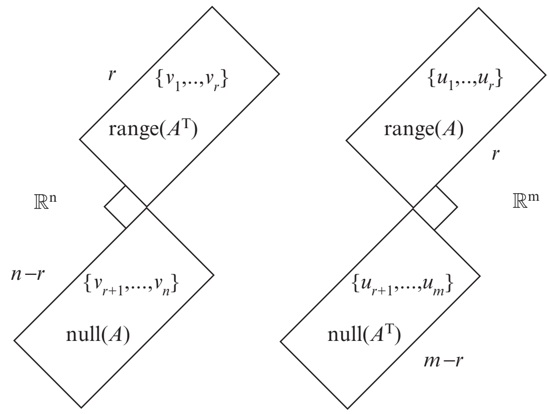


**FIGURE 15.1** The four fundamental subspaces of a matrix.

$A^{\mathrm{T}}$. If we take the transpose of the SVD for $A$, the result is

$$A^{\mathrm{T}} = V\widetilde{\Sigma}U^{\mathrm{T}}. \tag{15.6}$$

Applying the same procedure that we used to determine an orthonormal basis for the range of $A$ to Equation 15.6, it follows that $v_i$, $1 \leq i \leq r$ is a basis for the range of $A^{\mathrm{T}}$. Note that the range of $A^{\mathrm{T}}$ is the row space of $A$. We already know that $v_i$, $r+1 \leq i \leq n$, is a basis for null $(A)$. Since all the $v_i$ are orthogonal, it follows that the vectors in range $\left(A^{\mathrm{T}}\right)$ are orthogonal to the vectors in null $(A)$. Again using Equation 15.6, we see that $u_i$, $r+1 \leq i \leq m$ is an orthonormal basis for null $\left(A^{\mathrm{T}}\right)$. We have shown that $u_i$, $1 \leq i \leq r$ is a basis for range $(A)$. Thus, range $(A)$ is orthogonal to null $\left(A^{\mathrm{T}}\right)$. Table 15.1 summarizes the four fundamental subspaces, and Figure 15.1 provides a graphical depiction. In the figure, the symbol $\square$ indicates the subspaces are orthogonal.

We have stated a number of times that the dimension of the column space and row space are equal, and now we can prove it.

**Theorem 15.4.** *The dimension of the column space and the dimension of the row space of a matrix are equal and is called the rank of the matrix.*

*Proof.* Our discussion of the SVD has shown that if $r$ is the number of nonzero singular values, $u_i$, $1 \leq i \leq r$ is a basis for the range of $A$, and $v_i$, $1 \leq i \leq r$ is a basis for the range of $A^{\mathrm{T}}$, which is the row space of $A$. $\square$

**Example 15.5.** Let $A = \begin{bmatrix} 1 & 4 & 2 \\ -1 & 0 & 2 \\ 5 & -1 & -11 \\ 0 & 2 & 2 \\ 1 & 1 & -1 \end{bmatrix}$. The SVD of $A$ is

**TABLE 15.1** The Four Fundamental Subspaces of a Matrix

| | Range | Null Space |
|---|---|---|
| $A$ | $u_i, 1 \leq i \leq r$ | $v_i,\ r+1 \leq i \leq n$ |
| $A^{\mathrm{T}}$ | $v_i,\ 1 \leq i \leq r$ | $u_i,\ r+1 \leq i \leq m$ |

$$A = \begin{bmatrix} -0.1590 & -0.8589 & -0.3950 & 0.1461 & 0.2443 \\ -0.1740 & 0.0738 & 0.4983 & 0.6089 & 0.5876 \\ 0.9532 & -0.1726 & 0.0759 & 0.2364 & -0.0054 \\ -0.1665 & -0.3926 & 0.5709 & 0.1858 & -0.6766 \\ 0.0907 & -0.2701 & 0.5139 & -0.7194 & 0.3705 \end{bmatrix} \begin{bmatrix} 12.6906 & 0 & 0 \\ 0 & 4.7905 & 0 \\ 0 & 0 & 0.0000 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.3839 & -0.1443 & -0.9120 \\ -0.4312 & -0.9014 & -0.0389 \\ 0.8165 & -0.4082 & 0.4082 \end{bmatrix},$$

so $r = 2$. Using Table 15.1, the four fundamental subspaces have orthonormal bases as follows:

$$\text{range}\,(A) = \left\{ \begin{bmatrix} -0.1590 \\ -0.1740 \\ 0.9532 \\ -0.1665 \\ 0.0907 \end{bmatrix}, \begin{bmatrix} -0.8589 \\ 0.0738 \\ -0.1726 \\ -0.3926 \\ -0.2701 \end{bmatrix} \right\}, \quad \text{null}(A^{\mathrm{T}}) = \left\{ \begin{bmatrix} -0.3950 \\ 0.4983 \\ 0.0759 \\ 0.5709 \\ 0.5139 \end{bmatrix}, \begin{bmatrix} 0.1461 \\ 0.6089 \\ 0.2364 \\ 0.1858 \\ -0.7194 \end{bmatrix}, \begin{bmatrix} 0.2443 \\ 0.5876 \\ -0.0054 \\ -0.6766 \\ 0.3705 \end{bmatrix} \right\}$$

$$\text{range}(A^{\mathrm{T}}) = \left\{ \begin{bmatrix} 0.3839 \\ -0.1443 \\ -0.9120 \end{bmatrix}, \begin{bmatrix} -0.4312 \\ -0.9014 \\ -0.0389 \end{bmatrix} \right\}, \quad \text{null}\,(A) = \left\{ \begin{bmatrix} 0.8165 \\ -0.4082 \\ 0.4082 \end{bmatrix} \right\}$$

■

## 15.3 SVD AND MATRIX NORMS

The SVD provides a means of computing the 2-norm of a matrix, since $\|A\|_2 = \sqrt{\sigma_1}$. If $A$ is invertible, then $\left\|A^{-1}\right\|_2 = \sqrt{\frac{1}{\sigma_n}}$. The SVD can be computed accurately, so using it is an effective way to find the 2-norm. The SVD also provides a means of computing the Frobenius norm.

*Remark* 15.3. If $M$ is a general matrix (Problem 7.12),

$$\|A\|_F^2 = \text{trace}\left(A^{\mathrm{T}}A\right) = \text{trace}\left(AA^{\mathrm{T}}\right).$$

There is means of computing the Frobenius norm using the singular values of matrix $A$. Before developing the formula, we need to prove the invariance of the Frobenius norm under multiplication by orthogonal matrices.

**Lemma 15.1.** *If $U$ is an $m \times m$ orthogonal matrix, and $V$ is an $n \times n$ orthogonal matrix, then* $\|UAV\|_F^2 = \|A\|_F^2$.

*Proof.*

$$\|UA\|_F^2 = \text{trace}\left((UA)^{\mathrm{T}}(UA)\right) = \text{trace}\left(\left(A^{\mathrm{T}}U^{\mathrm{T}}\right)(UA)\right) = \text{trace}\left(A^{\mathrm{T}}IA\right) = \text{trace}\left(A^{\mathrm{T}}A\right) = \|A\|_F^2,$$

showing that the Frobenius norm is invariant under left multiplication by an orthogonal matrix. Now,

$$\|AV\|_F^2 = \text{trace}\left((AV)(AV)^{\mathrm{T}}\right) = \text{trace}\left((AV)\left(V^{\mathrm{T}}A^{\mathrm{T}}\right)\right) = \text{trace}\left(AA^{\mathrm{T}}\right) = \|A\|_F^2,$$

so the Frobenius norm is invariant under right multiplication by an orthogonal matrix. Now form the complete product.

$$\|UAV\|_F^2 = \|U(AV)\|_F^2 = \|AV\|_F^2 = \|A\|_F^2.$$ □

**Theorem 15.5.** $\|A\|_F = \left(\sum_{i=1}^{r} \sigma_i^2\right)^{\frac{1}{2}}$

*Proof.* By the SVD, there exist orthogonal matrices $U$ and $V$ such that $A = U\tilde{\Sigma}V^{\mathrm{T}}$. Then, $\|A\|_F = \left\|U\tilde{\Sigma}V^{\mathrm{T}}\right\|_F = \left\|\tilde{\Sigma}\right\|_F$ by Lemma 15.1. The only nonzero entries in $\tilde{\Sigma}$ are the singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$, so $\|A\|_F = \left(\sum_{i=1}^{r} \sigma_i^2\right)^{\frac{1}{2}}$. □

## 15.4 GEOMETRIC INTERPRETATION OF THE SVD

Multiplying a vector $x$ by a matrix $A$ stretches or contracts the vector. The singular values of $A$ significantly add to a geometric understanding of the linear transformation $Ax$. For $x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{R}^m$, the $m$-dimensional *unit sphere* is defined by

$$\sum_{i=1}^{m} x_i^2 = 1$$

If $A$ is an $m \times n$ matrix, the product $Ax$ takes a vector $x \in \mathbb{R}^n$ and produces a vector in $\mathbb{R}^m$. In particular, we will look at the subset of vectors $y = Ax$ in $\mathbb{R}^m$ as $x$ varies over the unit sphere, $\|x\|_2 = 1$. The matrix $A$ has $n$ singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, $\sigma_i = 0$, $r+1 \leq i \leq n$ and $A = U\widetilde{\Sigma}V^{\mathrm{T}}$, so we will investigate

$$Ax = U\widetilde{\Sigma}V^{\mathrm{T}}x, \quad \sum_{i=1}^{m} x_i^2 = 1.$$

If $K$ is an $n \times n$ orthogonal matrix and $x$ is on the unit sphere, then

$$\langle Kx, Kx\rangle = (Kx)^{\mathrm{T}} Kx = x^{\mathrm{T}}K^{\mathrm{T}}Kx = x^{\mathrm{T}}Ix = x^{\mathrm{T}}x = 1$$

Thus, the orthogonal linear transformation $Kx$ maps $x$ to another vector on the unit sphere. Now choose any $x \in \mathbb{R}^n$ on the unit sphere. Since the columns $k_i$, $1 \leq i \leq n$, of $K$ are an orthonormal basis for $\mathbb{R}^n$,

$$x = c_1k_1 + c_2k_2 + \cdots + c_pk_p = K\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}.$$

Then, $\|x\|_2^2 = \sum_{i=1}^{p} c_i^2 = 1$, so $\begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$ is on the unit sphere. Thus, $Kx$ maps out the unit sphere, $\|x\|_2 = 1$, in $\mathbb{R}^p$.

Assume $\|x\|_2 = 1$. Since the $n \times n$ matrix $V^{\mathrm{T}}$ in the SVD of $A$ is an orthogonal matrix, $y = V^{\mathrm{T}}x$ is on the unit sphere in $\mathbb{R}^n$, and

$$\widetilde{\Sigma}y = \begin{bmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ 0 & & & & 0 \end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sigma_1 y_1 \\ \vdots \\ \sigma_r y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} x_1' \\ \vdots \\ x_r' \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Now,

$$\left(\frac{x_1'}{\sigma_1}\right)^2 + \left(\frac{x_2'}{\sigma_2}\right)^2 + \cdots + \left(\frac{x_r'}{\sigma_r}\right)^2 = 1,$$

which is an $r$-dimensional *ellipsoid* in $\mathfrak{R}^m$ with semiaxes of length $\sigma_i$, $1 \leq i \leq r$. For instance, if $m = 3$ and $r = 2$, then we have a circle in 3-dimensional space. If $m = 3$ and $r = 3$, then the surface is an ellipsoid. We still have to account for multiplication by $U$. Since $U$ is an orthogonal matrix, $U$ causes a change of orthonormal basis, or a rotation.

**Summary** If $A$ is an $m \times n$ matrix, then $Ax$ applied to the unit sphere $\|x\|_2 \leq 1$ in $\mathbb{R}^n$ is a rotated ellipsoid in $\mathbb{R}^m$ with semiaxes $\sigma_i$, $1 \leq i \leq r$, where the $\sigma_i$ are the nonzero singular values of $A$.

**Example 15.6.** We will illustrate the geometric interpretation of the SVD with a MATLAB function `svdgeom` in the software distribution that draws the unit circle, $x_1^2 + x_2^2 = 1$, in the plane, computes $Ax$ for values of $x$, and draws the resulting ellipse. The function then outputs the SVD.

The linear transformation used for the example is $A = \begin{bmatrix} 1.50 & 0.75 \\ -0.50 & -1.00 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$. Note that the semiaxes of the ellipse are 1.9294 and 0.5831 (Figure 15.2).

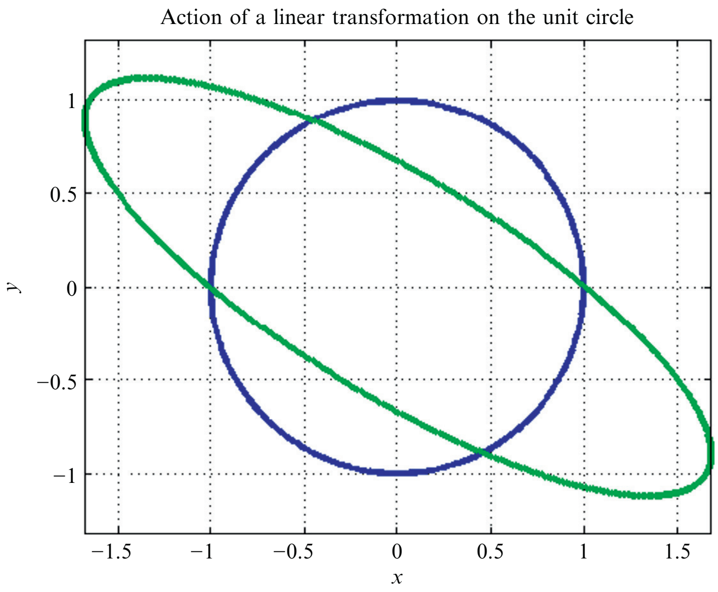


**FIGURE 15.2** SVD rotation and distortion.

```
>> svdgeom(A)
The singular value decomposition for A is

U =
   -0.8550    0.5187
    0.5187    0.8550

S =

    1.9294         0
         0    0.5831
V =

   -0.7991    0.6012
   -0.6012   -0.7991
```

■

## 15.5 COMPUTING THE SVD USING MATLAB

Although it is possible to compute the SVD by using the construction in Theorem 15.1, this is far too slow and prone to roundoff error, so it should not be used. The MATLAB function `svd` computes the SVD:

**a.** `[U S V] = svd(A)`
**b.** `S = svd(A)`

Form 2 returns only the singular values in descending order in the vector `S`.

**Example 15.7.** Find the SVD for the matrix of Example 15.5. Notice that $\sigma_3 = 4.8021 \times 10^{-16}$ and yet the rank is 2. In this case, the true value is 0, but roundoff error caused `svd` to return a very small singular value. MATLAB computes the rank using the SVD, so `rank` decided that $\sigma_3$ is actually 0.

```
>> [U,S,V] = svd(A)

U =
      0.15897       0.8589      -0.2112      -0.3642     -0.24447
      0.17398    -0.073783      -0.9063      0.22223      0.30581
     -0.95316      0.17263     -0.18552      0.14793    -0.073387
      0.16648      0.39256      0.21339      0.87898   -0.0063535
    -0.090746      0.27006      0.23251     -0.15324      0.91722
```

```
S =
       12.691          0          0
            0     4.7905          0
            0          0  4.8021e-16
            0          0          0
            0          0          0

V =
     -0.38387    0.43125     0.8165
       0.1443    0.90139   -0.40825
      0.91204   0.038895    0.40825

>> rank(A)

ans =
     2
```

■

The function `svd` applies equally well to a matrix of dimension $m \times n$, $m < n$. Of course, in this case the rank does not exceed $m$.

**Example 15.8.** Let $A = \begin{bmatrix} 7 & 9 & -5 & 10 & 10 & -8 \\ 9 & 3 & 1 & -7 & 0 & -2 \\ -8 & 8 & 10 & 10 & 6 & 9 \end{bmatrix}$

```
>> [U S V] = svd(A)

U =
     -0.42586   -0.89303   -0.14539
      0.26225   -0.27562     0.9248
     -0.86595    0.35571    0.35157

S =
       22.577          0          0          0          0          0
            0     20.176          0          0          0          0
            0          0     9.5513          0          0          0

V =
      0.27935   -0.57383     0.4704    0.60332   0.008204   0.085659
     -0.44176    -0.2983    0.44795   -0.37549   -0.51939   -0.32318
     -0.27763    0.38396    0.54102    0.10972    0.60466   -0.32426
     -0.65349    -0.1707   -0.46191    0.53138  0.0076852   -0.21915
     -0.41876   -0.33684   0.068636   -0.28352    0.38297     0.6924
     -0.21753     0.5401     0.2594    0.34673   -0.46673     0.5056

>> rank(A)

ans =
     3
```

Since $m = 3$ and the rank is 3, $A$ has full rank. ■

## 15.6 COMPUTING $A^{-1}$

We know that the inverse is often difficult to compute accurately and that, under most circumstances, its computation should be avoided. When it is necessary to compute $A^{-1}$, the SVD can be used. Since $A$ is invertible, the matrix $\widetilde{\Sigma}$ cannot have a 0 on its diagonal (rank would be $< n$), so $\widetilde{\Sigma} = \Sigma$. From $A = U\Sigma V^{\mathrm{T}}$, $A^{-1} = \left(V^{\mathrm{T}}\right)^{-1} \Sigma^{-1} U^{-1}$, where all the matrices have dimension $n \times n$. $U$ and $V$ are orthogonal, so

$$A^{-1} = V \begin{bmatrix} \frac{1}{\sigma_1} & & & & 0 \\ & \frac{1}{\sigma_2} & & & \\ & & \ddots & & \\ & & & \frac{1}{\sigma_{n-1}} & \\ 0 & & & & \frac{1}{\sigma_n} \end{bmatrix} U^{\mathrm{T}}.$$

**Example 15.9.** Let $A = \begin{bmatrix} 1 & -1 & 3 \\ 4 & 2 & 3 \\ 5 & 1 & -1 \end{bmatrix}$.

```
>> A = [1 -1 3;4 2 3;5 1 -1];
>> [U S V] = svd(A);
>> Ainv = V*diag(1./diag(S))*U'

Ainv =
    0.1190   -0.0476    0.2143
   -0.4524    0.3810   -0.2143
    0.1429    0.1429   -0.1429

>> inv(A)

ans =
    0.1190   -0.0476    0.2143
   -0.4524    0.3810   -0.2143
    0.1429    0.1429   -0.1429
```

■

## 15.7 IMAGE COMPRESSION USING THE SVD

Suppose you are given a fairly large image, at least $256 \times 256$ pixels. In any large image, some pixels will not be noticed by the human eye. By applying the SVD to a matrix representing the image, we can take advantage of this. The idea involves using only portions of the SVD that involve the larger singular values, since the smaller singular values do not contribute much to the image. Assume matrix $A$ contains the image in some format. Then,

$$A = U\tilde{\Sigma}V^{\mathrm{T}} = \begin{bmatrix} u_{11} & u_{12} & \cdots & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & \cdots & u_{2m} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ u_{m1} & \cdots & \cdots & \cdots & u_{mm} \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & 0 & \\ & & \sigma_r & & \\ & 0 & & \ddots & \\ & & & & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} & \cdots & \cdots & v_{n1} \\ v_{12} & v_{22} & \cdots & \cdots & v_{n2} \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ v_{1n} & \cdots & \cdots & & v_{nn} \end{bmatrix}.$$

**Definition 15.1.** A *rank 1 matrix* is a matrix with only one linearly independent column or row.

The primary idea in using the SVD for image compression is that we can write a matrix $A$ as a sum of rank 1 matrices.

**Lemma 15.2.** *After applying the SVD to an $m \times n$ matrix $A$, we can write $A$ as follows:*

$$A = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_r u_r v_r^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{T}},$$

*where each term in the sum is a rank 1 matrix.*

*Proof.* For $1 \le i \le r$, let $\Sigma_i$ be the $m \times n$ matrix

$$\Sigma_i = \begin{array}{c} \begin{bmatrix} 0 & & & & & & & \\ & \ddots & & & & & & \\ & & 0 & & & & & \\ & & & \sigma_i & & & & \\ & & & & \ddots & & & \\ & & & & & 0 & & \\ & & & & & & 0 & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{bmatrix} \begin{array}{c} \\ \\ \\ i \\ \\ r \\ \\ \\ \\ \end{array} \end{array}.$$

Then

$$\tilde{\Sigma} = \Sigma_1 + \Sigma_2 + \cdots + \Sigma_r,$$

and

$$A = U\Sigma_1 V^{\mathrm{T}} + U\Sigma_2 V^{\mathrm{T}} + \cdots + U\Sigma_r V^{\mathrm{T}} = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_r u_r v_r^{\mathrm{T}}.$$

Each product $\sigma_i u_i v_i^{\mathrm{T}}$ has dimension $(m \times 1)\,(1 \times n) = m \times n$, so the sum is an $m \times n$ matrix. Each portion of the sum, $u_i v_i^{\mathrm{T}}$ has the form

$$\sigma_i \begin{bmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \\ \vdots \\ u_{mi} \end{bmatrix} \begin{bmatrix} v_{1i} & v_{2i} & v_{3i} & \cdots & v_{ni} \end{bmatrix} = \sigma_i \begin{bmatrix} u_{1i}v_{1i} & u_{1i}v_{2i} & u_{1i}v_{3i} & \cdots & u_{1i}v_{ni} \\ u_{2i}v_{1i} & u_{2i}v_{2i} & u_{2i}v_{3i} & \cdots & u_{2i}v_{ni} \\ u_{3i}v_{1i} & u_{31i}v_{2i} & u_{31i}v_{3i} & \cdots & u_{3i}v_{ni} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ u_{mi}v_{1i} & u_{m1i}v_{2i} & u_{m1i}v_{3i} & \cdots & u_{mi}v_{ni} \end{bmatrix}. \tag{15.7}$$

Each column in Equation 15.7 is a multiple of the vector $\begin{bmatrix} u_{1i} \\ u_{2i} \\ u_{3i} \\ \vdots \\ u_{mi} \end{bmatrix}$, so each matrix $\sigma_i u_i v_i^{\mathrm{T}}$ has rank 1. □

Each term $\sigma_i u_i v_i^{\mathrm{T}}$ is called a *mode*, so we can view an image as a sum of modes. Because the singular values $\sigma_i$ are ordered $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, significant compression of the image is possible if the set of singular values has only a few large modes. Form the sum of those modes, and it will be indistinguishable from the original image. For instance if the rank of the matrix containing the image is 350, it is possible that only modes 1-25 are necessary to cleanly represent the image. If the first $k$ modes are summed, the rank of the matrix will be $k$, since the first $k$ singular values are in the sum. In other words, if $\overline{A} = \sum_{i=1}^{k} \sigma_i u_i v_i^{\mathrm{T}}$, then $\operatorname{rank}(\overline{A}) = k$. If the first $k$ modes dominate the set of $r$ modes, then $\overline{A}$ will be a good approximation to $A$.

### 15.7.1 Image Compression Using MATLAB

Image processing with MATLAB is somewhat complex, so we will confine ourselves to simple operations with gray scale images. In order to use the SVD for working with images, you must input the file using the command `imread`.

```
>> A = imread('filename.ext');
```

The extension "ext" can be one of many possibilities, including "tif/tiff," "bmp," "gif," "jpg," and so forth. Display the image with the commands

```
>> imagesc(A);
>> colormap(gray);
```

The first function "`imagesc`" scales the image so it uses the full `colormap` and displays it. A color map is a matrix that may have any number of rows, but it must have exactly 3 columns. Each row is interpreted as a color, with the first element specifying the intensity of red light, the second green, and the third blue (RGB). For gray scale, the MATLAB colormap is `gray`. After reading the file into the array $A$, the data type of its elements must be converted from `uint8` (unsigned 8-bit integers) to `double`.

```
>> A = double(A);
```

At this point, the SVD can be applied to the array $A$.

```
>> [U S V] = svd(A);
```

Compute the sum of the first $k$ modes using the following statement:

```
>> Aapprox = U(:,1:k)*S(1:k,1:k)*V(:,1:k)';
```

Now display the approximation using

```
>> imagesc(Aapprox);
```

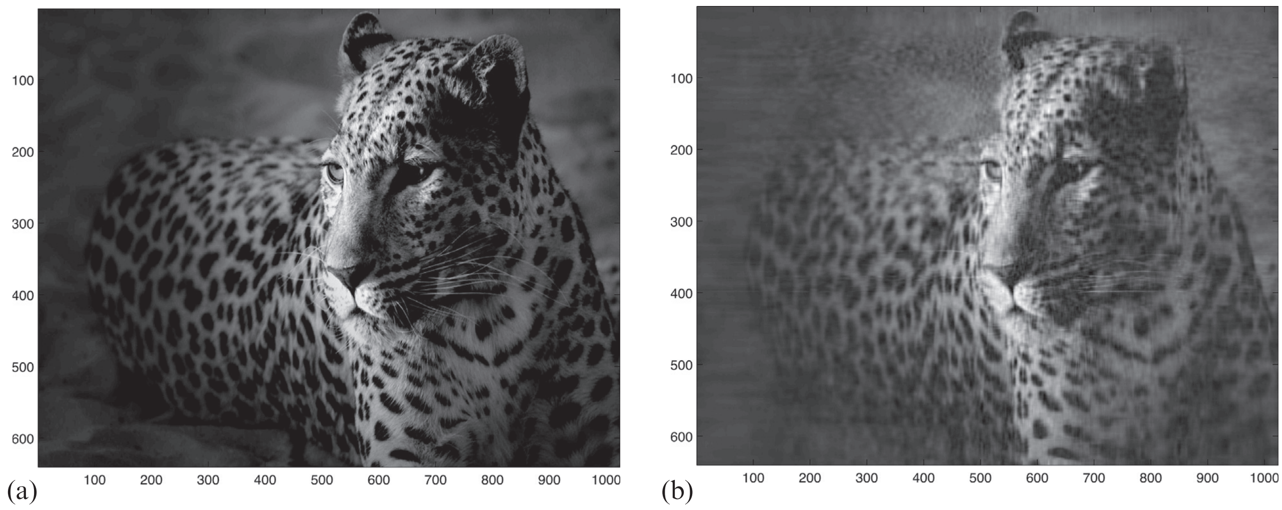


**FIGURE 15.3** (a) jaguar (640 × 1024) and (b) jaguar using 35 modes.

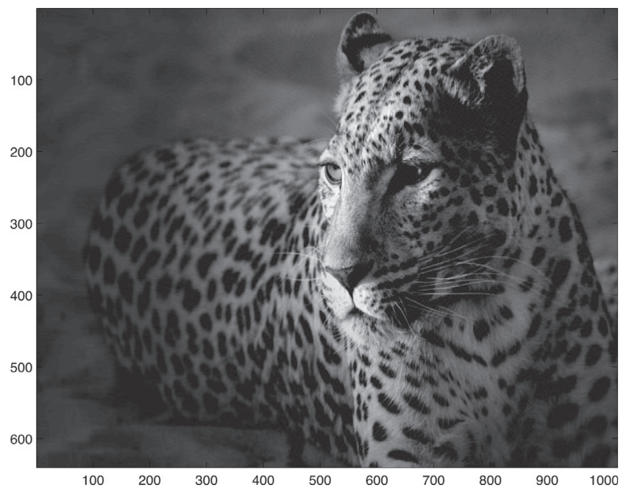

**FIGURE 15.4** Jaguar using 145 modes.

The following MATLAB sequence reads a picture of a jaguar, displays it, and then displays the image using 35 of the 640 modes (Figure 15.3). The 640 × 1024 image is in JPG format within the software distribution. Notice that the use of 35 modes is somewhat blurry but very distinguishable.

```
>> JAGUAR = imread('jaguar.jpg');
>> JAGUAR = double(JAGUAR);
>> imagesc(JAGUAR); colormap(gray);
>> [U S V] = svd(JAGUAR);
>> rank(JAGUAR)

ans = 640

>> figure(2);
>> JAGUAR35 = U(:,1:35)*S(1:35,1:35)*V(:,1:35)';
>> imagesc(JAGUAR35);colormap(gray);
```

Now let us use 145 modes and see the result in Figure 15.4.

```
>> JAGUAR145 = U(:,1:145)*S(1:145,1:145)*V(:,1:145)';
>> imagesc(JAGUAR145);colormap(gray);
```

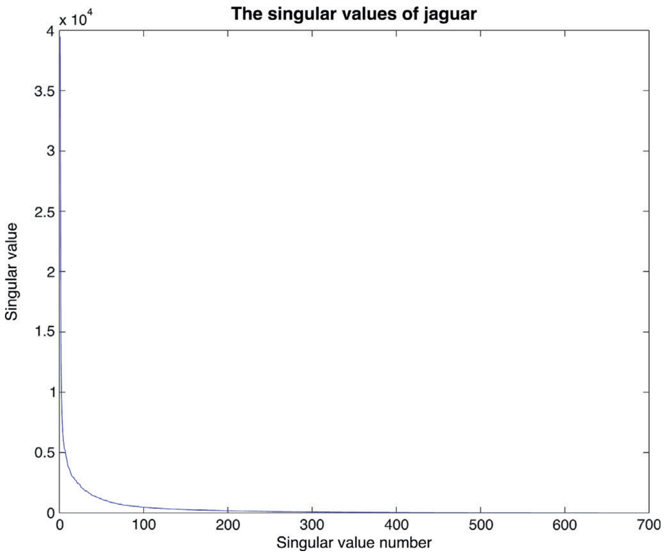


**FIGURE 15.5** Singular value graph of jaguar.

Figure 15.5 is a graph of the 640 singular value numbers versus the singular values. Note that the initial singular values are very large, and modes larger than number 145 contribute almost nothing to the image.

In general, if an image is stored in an $m \times n$ matrix, we need to retain $m \times n$ numbers. Suppose that after converting the image to a matrix, we perform the SVD on this matrix and discover that only the largest $k$ singular values capture the "important" information. Instead of keeping $m \times n$ numbers, we keep the $k$ singular values, plus the $k$ vectors $u_1, u_2, \ldots, u_k$ of dimension $m$, plus vectors $v_1, v_2, \ldots, v_k$ of dimension $n$, for a total of $k + km + kn$ numbers. For jaguar using $k = 145$, the ratio of the compressed image to the original image is 0.3684.

In addition to `jaguar.jpg`, the software distribution provides the files

*`black-hole.tif`, `horsehead-nebula.tif`, `planets.tif`, `saturn.tif`, and `whirlpool.tif`*

in the subdirectory SVD_compress. For each file, there is a corresponding file with the same name and extension "`.mat`." Each file name is in uppercase and contains the image converted to array format. By using the `load` command, you directly obtain the array in `double` format; for example,

```
>> load SATURN;
```

The software distribution contains a function `svdimage` that allows you to start with any mode of an image, add one mode per mouse click, and watch the image improve. Input to the function is the image in matrix format, the starting mode number, and the colormap, which should be `gray`. Here is an example using the image of Saturn with a starting mode of 50 (Figure 15.6). Be sure to terminate the function by typing "q," or you will receive a MATLAB error message.

```
>> load SATURN;
>> svdimage(SATURN, 50, gray);
```

### 15.7.2 Additional Uses

The SVD has many applications:

- Most accurate way to determine the rank of a matrix and the four fundamental subspaces.
- Determining the condition number of a matrix.
- Solve least-squares problems. Least-squares problems are discussed in Chapter 16.
- Used in computer graphics because its factors provide geometric information about the original matrix.
- Principal components analysis approximating a high-dimensional data set with a lower-dimensional subspace. Used in statistics.
- Image compression

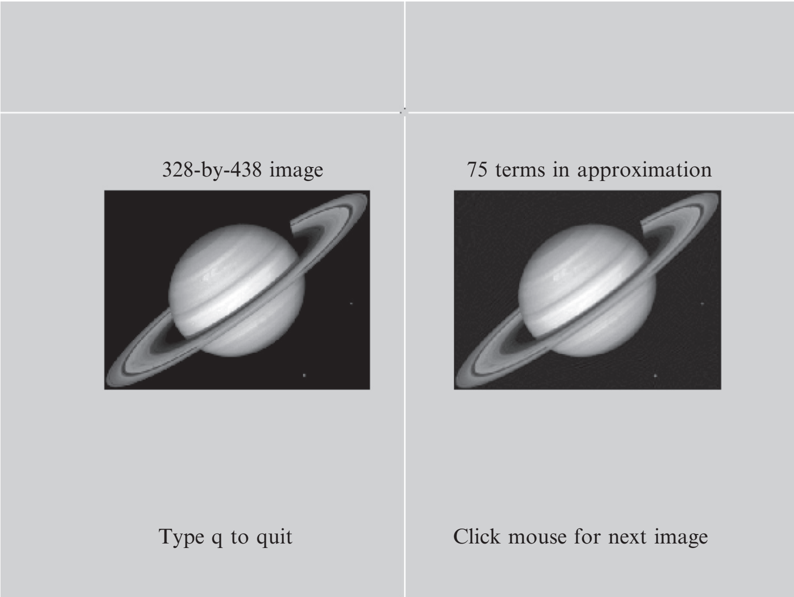


**FIGURE 15.6** SVD image capture.

- Image restoration, in which "noise" has caused an image to become blurry. The noises are caused by small singular values, and the SVD can be used to remove them.
- Applications in digital signal processing. For example, as a method for noise reduction. Let a matrix $A$ represents the noisy signal, compute the SVD, and then discard small singular values of $A$. The small singular values primarily represent the noise, and thus a rank-$k$ matrix, $k < \text{rank}(A)$, represents a filtered signal with less noise.

## 15.8 FINAL COMMENTS

The SVD of a matrix is a powerful technique for matrix computations. Despite its power, however, there are some disadvantages. The SVD is computationally expensive. Many real world problems involve very large matrices. In these cases, applying simpler techniques, such as the *QR* decomposition or another of a number of matrix decompositions may be indicated. The SVD operates on a fixed matrix, and hence it is not useful in problems that require adaptive procedures. An *adaptive procedure* is a procedure that changes its behavior based on the information available. A good example is adaptive quadrature, a very accurate method for approximating $\int_a^b f(x)\,\mathrm{d}x$. As an adaptive method runs, it estimates error, isolates regions where the error tolerance has not been met, and deals with those regions separately. Adaptive quadrature is particularly effective when $f(x)$ behaves badly near a point in the interval $a \le x \le b$.

## 15.9 CHAPTER SUMMARY

### The SVD Theorem

The SDV theorem states that there exist orthogonal matrices $U \in R^{m\times m}$, $V \in \mathbb{R}^{n\times n}$, $m \ge n$, and a diagonal matrix $\tilde{\Sigma} \in \mathbb{R}^{m\times n}$such that

$$\begin{aligned} A &= U\tilde{\Sigma}V^{\mathrm{T}} \\ \tilde{\Sigma} &= \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}, \end{aligned}$$

where $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$, and $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$ are the positive singular values of $A$. The proof of the theorem is by construction. The columns of $V = \begin{bmatrix} v_1 & v_2 & \ldots & v_{n-1} & v_n \end{bmatrix}$ are the eigenvalues of $A^{\mathrm{T}}A$, and $U$ is constructed from $Av_i$ and the use of Gram-Schmidt to fill out the matrix, if necessary. If $m < n$, then find the SVD of

$$A^{\mathrm{T}} = U\tilde{\Sigma}V^{\mathrm{T}},$$

and form

$$A = V\Sigma\tilde{U}^{\mathrm{T}}.$$

## Determining Matrix Properties Using the SVD

The SVD reveals substantial information about $A$:

The rank of $A$ is the number of nonzero singular values, $r$. The following table lists the bases of four subspaces immediately available from the SVD:

$$A = U\tilde{\Sigma}V^{T}$$

| | **Range** | **Null Space** |
|---|---|---|
| $A$ | $u_i, 1 \le i \le r$ | $v_i, r+1 \le i \le n$ |
| $A^{\mathrm{T}}$ | $v_i, 1 \le i \le r$ | $u_i, r+1 \le i \le m$ |

Note that the column space and the row space both have dimension $r$, proving that the row and column space of a matrix have the same dimension.

## The SVD and Matrix Norms

The SVD provides a means of computing the 2-norm of a matrix, since $\|A\|_2 = \sqrt{\sigma_1}$. If $A$ is invertible, then $\left\|A^{-1}\right\|_2 = \sqrt{\frac{1}{\sigma_r}}$. The SVD can be computed accurately, so using it is an effective way to find the 2-norm. The SVD also provides a means of computing the Frobenius norm, namely, $\|A\|_F = \left(\sum_{i=1}^{r} \sigma_i^2\right)^{\frac{1}{2}}$.

## Geometric Interpretation of the SVD

If $A$ is an $m \times n$ matrix, the SVD tells us that $Ax$ applied to the unit sphere $\|x\|_2 \le 1$ in $\mathbb{R}^n$ is a rotated ellipsoid in $\mathbb{R}^m$ with semiaxes $\sigma_i$, $1 \le i \le r$, where the $\sigma_i$ are the nonzero singular values of $A$. In $\mathbb{R}^2$, the image is a rotated ellipse centered at the origin and in $\mathbb{R}^3$, the image is a rotated ellipsoid centered at the origin.

## Computation of the SVD Using MATLAB

In MATLAB, compute just the singular values with the statement

```
S = svd(S);
```

and the full SVD with

```
[U,S,V] = svd(A);
```

## Using the SVD to Compute $A^{-1}$

Normally, we do not compute $A^{-1}$, but if it is required we can proceed as follows:

$$\begin{aligned}
A &= U\Sigma V^{\mathrm{T}} \\
A^{-1} &= V\Sigma^{-1}U^{\mathrm{T}} \\
A^{-1} &= V\begin{bmatrix} \frac{1}{\sigma_1} & & & & 0 \\ & \frac{1}{\sigma_2} & & & \\ & & \ddots & & \\ & & & \frac{1}{\sigma_{n-1}} & \\ 0 & & & & \frac{1}{\sigma_n} \end{bmatrix} U^{\mathrm{T}}.
\end{aligned}$$

Since $A$ is invertible, all the singular values are nonzero.

## Image Compression Using the SVD

After applying the SVD to an $m \times n$ matrix A, we can write A as follows:

$$A = \sigma_1 u_1 v_1^{\mathrm{T}} + \sigma_2 u_2 v_2^{\mathrm{T}} + \cdots + \sigma_r u_r v_r^{\mathrm{T}} = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{T}},$$

where each term in the sum is a rank 1 matrix. To use the SVD for image compression, convert the image to a matrix and discard the terms involving small singular values. For instance, if the first $k$ singular values dominate the remaining ones the approximation is

$$\tilde{A} = \sum_{i=1}^{k} \sigma_i u_i v_i^{\mathrm{T}}.$$

After applying the MATLAB function svd and obtaining U, S, and V, the approximation can be written using colon notation as

```
U(:,1:k)*S(1:k,1:k)*V(:,1:k)'.
```

The reader should follow Section 15.7 and use the compression technique on the images supplied in the subdirectory `SVD_compress` with the book software distribution.

## 15.10 PROBLEMS

**15.1** Prove that for all rank-one matrices, $\sigma_1^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$. Hint: Use Theorem 15.5.

**15.2** Suppose $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ are orthonormal bases for $\mathbb{R}^n$. Construct the matrix $A$ that transforms each $v_i$ into $u_i$ to give $Av_1 = u_1, Av_2 = u_2, \ldots, Av_n = u_n$.

**15.3** Let $A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$. Determine value(s) for the $a_{ij}$ so that A has distinct singular values.

**15.4** Prove that rank $\left(A^{\mathrm{T}}A\right) =$ rank $\left(AA^{\mathrm{T}}\right)$.

**15.5** Find the SVD of

**a.** $A^{\mathrm{T}}A$.

**b.** $\left(A^{\mathrm{T}}A\right)^{-1}$

**15.6** If $S$ is a subspace of $\mathbb{R}^n$, the *orthogonal complement* of S, written $S^{\perp}$, is the set of all vectors $x$ orthogonal to $S$; in other words

$$S^{\perp} = \left\{x \in \mathbb{R}^n \mid x^{\mathrm{T}}y = < x,\ y > = 0 \text{ for all } y \in S\right\}.$$

Prove that $S^{\perp}$ is a subspace. Find orthogonal complements among the four fundamental subspaces.

**15.7** Let $A \in \mathbb{R}^n$ be a rank-one matrix. Show that there exist two vectors $u$ and $v$ such that $A = uv^{\mathrm{T}}$.

**15.8** If $B = kA$, where $k$ is a positive integer, what is the SVD of $B$?

**15.9** Find an SVD of a column vector and a row vector.

**15.10**

**a.** What is the SVD for $A^{\mathrm{T}}$?

**b.** What is the SVD for $A^{-1}$?

**15.11** Suppose $P$ is an orthogonal matrix, and $B$ is an $n \times n$ matrix. Show that $A = P^{-1}BP$ has the same singular values as $B$.

**15.12** Prove that if $A$ is nonsingular, all its singular values are greater than zero.

**15.13** Prove that the null space of $A^{\mathrm{T}}A$ and $A$ are equal.

**15.14** Show that if $A$ is an $n \times n$ matrix and $A = U\widetilde{\Sigma}V^{\mathrm{T}}$ is its SVD, then

$$\left\|A^2\right\|_2 = \left\|\widetilde{\Sigma}V^{\mathrm{T}}U\widetilde{\Sigma}\right\|_2.$$

**15.15**

**a.** Prove that the matrices $A^{\mathrm{T}}A$ and $AA^{\mathrm{T}}$ have the same eigenvalues

$$\left\{\ \sigma_1^2\ \ \sigma_2^2\ \ \ldots\ \ \sigma_r^2\ \ 0\ \ \ldots\ \ 0\ \right\}.$$

**b.** Prove that the orthonormal column vectors of $V$ are orthonormal eigenvectors of $A^{\mathrm{T}}A$ and that the column vectors of $U$ are the orthonormal eigenvectors of $AA^{\mathrm{T}}$.

**c.** Using the results of (a) and (b), describe an algorithm for the computation of the SVD.

**d.** Using your algorithm, compute the SVD for the matrix $A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$. Do the computations with pencil and paper.

**15.16** Assume $A$ is nonsingular with SVD

$$A = U\Sigma V^{\mathrm{T}}.$$

**a.** Prove that

$$\sigma_n \|x\|_2 \leq \|Ax\|_2 \leq \sigma_1 \|x\|_2 .$$

Hint: To prove the left half of the inequality, use $A^{-1}$.

**b.** Show that $\frac{\|Ax\|_2}{\|x\|_2}$ attains its maximum value $\sigma_1$ at $x = v_1$.

**c.** Show that $\frac{\|A^{-1}x\|_2}{\|x\|_2}$ attains its maximum value $\frac{1}{\sigma_n}$ at $x = u_n$.

**15.17** An $n \times n$ matrix $X$ is said to be the *square root* of $A$ if $A = X^2$.

**a.** Show that if $A$ is positive definite and not diagonal, then the Cholesky factor $R$ is not a square root.

**b.** Let $A$ be positive definite and $A = S^{\mathrm{T}}S$ be the Cholesky decomposition of $A$. Let $S = U\Sigma V^{\mathrm{T}}$ be the SVD for $S$, and define $X = V\Sigma V^{\mathrm{T}}$. Show that $X$ is positive definite and $X^2 = A$, so that a positive definite matrix has a positive definite square root.

### 15.10.1 MATLAB Problems

**15.18** **a.** Find the SVD for the singular matrix

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

**b.** Find a basis for its range and null space.

**15.19** **a.** Find the SVD for the matrix

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

**b.** What is its rank and the dimension of its null space?

**15.20** Find the SVD for the matrices

**a.** $\begin{bmatrix} 2 & 5 & -1 & -7 \\ -4 & 1 & 5 & 0 \end{bmatrix}$

**b.** $\begin{bmatrix} -1 & 1 & 2 & 3 & 5 \\ 1 & 2 & 3 & 4 & 5 \\ -4 & -3 & -2 & -1 & 0 \end{bmatrix}$

In each case, find an orthonormal basis for the range, null space, row space, and the null space of its transpose.

**15.21** **a.** Execute the MATLAB function `svdgeom` with the matrix $\begin{bmatrix} 6 & 1 \\ -7 & 3 \end{bmatrix}$.

**b.** Do part (a) using the matrix $\begin{bmatrix} 0.092091 & -0.0043853 \\ 0.035082 & 0.052623 \end{bmatrix}$.

**15.22** Develop a method for building a $2 \times 2$ matrix with specified singular values $\sigma_1, \sigma_2$. Use your method to construct matrix $A$ with singular values $\sigma_1 = 55.63$, $\sigma_2 = 25.7$, and matrix $B$ with singular values $\sigma_1 = .2$, $\sigma_2 = .1$. In each case, use the function `svdgeom` to show how the linear transformation transforms the unit circle.

**15.23** The software distribution contains a graphics file "`black-hole.tif`." Use MATLAB to read the image, convert it to a matrix, and use the SVD to display the graphic using only large modes. Compute the percentage of image storage you save.

**15.24** The software distribution contains the files `SATURN.mat` and `WHIRLPOOL.mat`. Each of these is an image matrix stored in MATLAB format. Input each one using the `load` command and experiment with it using the SVD to compress the images. In each case, draw a graph of the singular value number vs. the singular value similar to Figure 15.5.

*Remark* 15.4. There is another image, HORSEHEAD.mat, with which you might want to experiment.

**15.25** The Hilbert matrices $H_{ij} = \frac{1}{i+j-1}$, $1 \leq i,j \leq n$ have notoriously bad properties. It can be shown that any Hilbert matrix is nonsingular and, as such, has rank $n$. MATLAB constructs a Hilbert matrix with the command `H = hilb(n)`.

**a.** Use MATLAB to verify that the rank of the $8 \times 8$ Hilbert matrix $H$ is 8.
**b.** Find the SVD of $H$.
**c.** Comment on the singular values.
**d.** Compute $\frac{\sigma_1}{\sigma_8}$, the condition number of $H$.

**15.26**

**a.** Construct `A = rand(m,n)` for $m = 5$, $n = 4$. Find the eigenvalues and eigenvectors for $A^{\mathrm{T}}A$ and $AA^{\mathrm{T}}$. Do the experiment again with $m = 3$, $n = 5$.
**b.** If you see a pattern of behavior, state a theorem a prove it.
The following problem is adapted from material in Ref. [46, Section 10.7].

**15.27** This problem is adapted from the material in Ref. [46], Section 10.7.

**a.** MATLAB provides a matrix, `gallery(5)`, for eigenvalue testing. Use the MATLAB poly function to determine the characteristic polynomial of $A$. Is the matrix singular? Would computation of the eigenvalues of $A$ be ill-conditioned? Explain.

The remainder of the problem deals with the singular values of `gallery(5)`.

**b.**

**i.** Show that $\tilde{\Sigma} + \delta\tilde{\Sigma} = U^{\mathrm{T}}(A + \delta A)V$.

**ii.** Using (b), part (i) show that $\|\delta A\|_2 = \left\|\delta\tilde{\Sigma}\right\|_2$. This says that the size of the errors in computing the singular values is the same as the errors involved in forming $A$. This type of perturbation result is ideal.

**iii.** The result of (b), part (ii) deals with all elements of the singular value problem. There can be very large and very small singular values, and the primary problem is with the small singular values. If $A$ is singular, one or more singular values will be zero, but may not actually be reported as zero due to rounding errors. It is hard to distinguish between a very small singular value and one that is actually 0. gallery (5) will help in understanding the problems with small singular values. Execute the statements

```
A = gallery(5);
format long e;
svd(A)
```

Comment on the distribution of singular values. In Ref. [46], it is stated that the small singular values that should be 0 lie somewhere between *eps* and $\|A\|_2$ *eps*. Is this the case here?

**iv.** Compute the SVD of a randomly perturbed matrix by running the following code. The function `randn(5,5)` creates a random $5 \times 5$ matrix, and `randn(5,5).*A` multiplies each entry $a_{ij}$ by $r_{ij}$. The sum `A+eps*randn(5,5).*A` perturbs `gallery(5)` by a small amount. Run

```
format long e
clc

for i = 1:5
   svd(A + eps*randn(5,5).*A)
   fprintf('----------------------------------');
end
fprintf('\n');
```

The MATLAB output is in the format $digit.d_1d_3d_3\ldots d_{15}e \pm e_1e_2e_3$. Write down one line of svd output; for example,

```
1.010353607103610e+005
1.679457384066240e+000
1.462838728085211e+000
```

```
1.080169069985495e+000
4.288402425161663e-014
```

Analyze the output of the `for` statement and place a star ('*') at every digit position in your written line of output that changes. The asterisks show the digits that change as as a result of random perturbations. Comment on the results.

**15.28**

**a.** Write a function, `buildtaumat`, that builds the matrix

$$T = \begin{bmatrix} 0 & \Gamma_n \\ \Gamma_n & 0 \end{bmatrix},$$

where $\Gamma_n$ is an $n \times n$ submatrix of ones, and 0 represents an $n \times n$ zero submatrix. Thus, T is a $2n \times 2n$ matrix. For instance, $T = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$.

**b.** Construct the matrix for $n = 2$, 4, 8, and 16. In each case, compute the eigenvalues and the singular values. Propose a theorem that explains what you observe. Prove it.

**15.29** If $A$ is nonsingular, the SVD can be used to solve a linear system $Ax = b$.

**a.** Explain why $\tilde{\Sigma}$ in the SVD of $A$ is invertible. What is its inverse?

**b.** Show that

$$x = V\tilde{\Sigma}^{-1}U^{\mathrm{T}}b.$$

**c.** Develop a one line MATLAB command to compute $\tilde{\Sigma}^{-1}$ that only uses the function `diag`.

**d.** Solve

$$\begin{bmatrix} 1 & -1 & 0 \\ 8 & 4 & 1 \\ -9 & 0 & 3 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 3 & 0 & 1 \\ 1 & 5 & -3 & 8 \\ 12 & 5 & 7 & 0 \\ 6 & 77 & 15 & 35 \end{bmatrix} x = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$$

using the results of parts (a)-(c).

**e.** Does using the SVD seems like a practical means of solving $Ax = b$? Explain your answer.

**15.30** There is a geometric interpretation of the condition number. You are given the matrix sequence

$$\left\{ \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1.5 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1.9 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1.99 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1.999 \end{bmatrix} \right\}.$$

**a.** Remove textual output from the function `svdgeom` introduced in Section 9.4 so it only plots a graph, and name the function `svdgeom1`.

**b.** For each matrix in the sequence, do the following

(a) Compute the condition number.
(b) Call the function `svdgeom1` and observe the graph.

**c.** Do you see a relationship between the condition number of a matrix and its action as a linear transformation?

**15.31** In discussing image compression using the SVD, we approximated the graphical representation by using the first $k$ singular values and discarding the rest. The image looked just as good as the original if we included sufficient singular values. This process is known as computing a rank $k$ approximation. There is theoretical justification for this technique. A proof of the following theorem can be found in Refs. [1, pp. 110-113] and [47, pp. 83-84]. It says that using the SVD we obtain the optimal rank $k$ approximation to a matrix.

***Theorem***. *Assume that the* $m \times n$ *matrix* $A$, $m \geq n$, *has the SVD* $A = U\tilde{\Sigma}V^{\mathrm{T}}$. *If* $k < \operatorname{rank}(A)$, *a matrix of rank k closest to A as measured by the 2-norm is* $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^{\mathrm{T}}$, *and* $\|A - A_k\|_2 = \sigma_{k+1}$. $A_k$ *can also be written as* $A_k = U\tilde{\Sigma}_k V^{\mathrm{T}}$, *where* $\tilde{\Sigma}_k = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma, 0, \ldots, 0)$. *Another way of putting this is*

$$\sigma_{k+1} = \min_{B \in \mathbb{R}^{m\times n},\, \operatorname{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 .$$

Run the following MATLAB program and explain the output.

```
load('HORSEHEAD.mat');
[U,S,V] = svd(HORSEHEAD);
HORSEHEAD250 = U(:,1:250)*S(1:250,1:250)*V(:,1:250)';
figure(2);
imagesc(HORSEHEAD250);
colormap(gray);
fprintf('sigma251 = %.12f\n\n', S(251,251));
deltaS = 0.1:-.005:0;
for i = 1:length(deltaS)
   B = HORSEHEAD;
   B = B + deltaS(i)*ones(566,500);
   B = makerank(B,250);
   fprintf('%.12f\n', abs(norm(HORSEHEAD - B)-S(251,251)));
end
```

**15.32** The *polar decomposition* of an $n \times n$ matrix $A$ is

$$A = UP,$$

where $U$ is orthogonal and $P$ is symmetric positive semidefinite ($x^{\mathrm{T}}Px \geq 0$ for all $x \neq 0$). Intuitively, the polar decomposition factors $A$ into a component $P$ that stretches $Ax$ along a set of orthogonal axes followed by a rotation $U$. This is analogous to the polar form of a complex number $z = r\mathrm{e}^{\mathrm{i}\theta}$. $P$ plays the role of $r$, and $U$ plays the role of $\mathrm{e}^{\mathrm{i}\theta}$. Applications of the polar decomposition include factor analysis and aerospace computations [48].

**a.** If $A = U\Sigma V^{\mathrm{T}}$ is the SVD for $A$, show that $A = \left(UV^{\mathrm{T}}\right)\left(V\Sigma V^{\mathrm{T}}\right)$ is a polar decomposition for $A$.

**b.** For what class of matrices can we guarantee that $P$ is positive definite?

**c.** Write a function `[U P] = polardecomp(A)` that computes a polar decomposition for the square matrix $A$. Test your function with matrices of dimensions $3 \times 3$, $5 \times 5$, $10 \times 10$, and $50 \times 50$.

**15.33** **a.** Using the result of Problem 15.17, write a function `sqrroot` that computes the square root of a positive definite matrix.

**b.** For $n = 5$, 10, 25, and 50, test your function using the matrices

`gallery('moler',n)`

**15.34** Generate a random $2 \times 2$ matrix `A = rand(2,2)`. Then type `eigshow(A)` at the MATLAB prompt. A window will open. Click on the svd button on the right side of the window. Your matrix `A` will appear (in MATLAB notation) in the menu bar above the graph. Underneath the graph the statement "Make $A{*}x$ perpendicular to $A{*}y$" should appear. The graph shows a pair of orthogonal unit vectors `x` and `y`, together with the image vectors `Ax` and `Ay`. Move the pointer onto the vector `x`, and then make the pair of vectors `x`, `y` go around in a circle. The transformed vectors `Ax` and `Ay` then move around an ellipse, as we expect from the discussion in Section 15.4. Generally `Ax` will not be perpendicular to `Ay`. Keep moving vector `x` until you find a position where `Ax` is perpendicular to `Ay`. When this happens, then the singular values $\sigma_1$ and $\sigma_2$ of `A` are the lengths of the vectors `Ax` and `Ay`. Estimate the lengths from the graph. Take note of the fact that $\|x\|_2 = \|y\|_2 = 1$. Confirm your estimates by using MATLAB's `svd` command.