4. $Q(\mathbf{x}) = 3x_1^2+3x_2^2+5x_3^2+6x_1x_2+2x_1x_3+2x_2x_3$ (See Exercise 2.)

5. $Q(\mathbf{x}) = x_1^2 + x_2^2 - 10x_1x_2$

6. $Q(\mathbf{x}) = 3x_1^2 + 9x_2^2 + 8x_1x_2$

7. Let $Q(\mathbf{x}) = -2x_1^2 - x_2^2 + 4x_1x_2 + 4x_2x_3$. Find a unit vector $\mathbf{x}$ in $\mathbb{R}^3$ at which $Q(\mathbf{x})$ is maximized, subject to $\mathbf{x}^T\mathbf{x} = 1$. [*Hint:* The eigenvalues of the matrix of the quadratic form Q are 2, −1, and −4.]

8. Let $Q(\mathbf{x}) = 7x_1^2 + x_2^2 + 7x_3^2 - 8x_1x_2 - 4x_1x_3 - 8x_2x_3$. Find a unit vector $\mathbf{x}$ in $\mathbb{R}^3$ at which $Q(\mathbf{x})$ is maximized, subject to $\mathbf{x}^T\mathbf{x} = 1$. [*Hint:* The eigenvalues of the matrix of the quadratic form Q are 9 and −3.]

9. Find the maximum value of $Q(\mathbf{x}) = 7x_1^2 + 3x_2^2 - 2x_1x_2$, subject to the constraint $x_1^2 + x_2^2 = 1$. (Do not go on to find a vector where the maximum is attained.)

10. Find the maximum value of $Q(\mathbf{x}) = -3x_1^2 + 5x_2^2 - 2x_1x_2$, subject to the constraint $x_1^2 + x_2^2 = 1$. (Do not go on to find a vector where the maximum is attained.)

11. Suppose $\mathbf{x}$ is a unit eigenvector of a matrix A corresponding to an eigenvalue 3. What is the value of $\mathbf{x}^TA\mathbf{x}$?

12. Let λ be any eigenvalue of a symmetric matrix A. Justify the statement made in this section that $m \le \lambda \le M$, where m and M are defined as in (2). [*Hint:* Find an $\mathbf{x}$ such that $\lambda = \mathbf{x}^TA\mathbf{x}$.]

13. Let A be an $n \times n$ symmetric matrix, let M and m denote the maximum and minimum values of the quadratic form $\mathbf{x}^TA\mathbf{x}$, where $\mathbf{x}^T\mathbf{x} = 1$, and denote corresponding unit eigenvectors by $\mathbf{u}_1$ and $\mathbf{u}_n$. The following calculations show that given any number t between M and m, there is a unit vector $\mathbf{x}$ such that $t = \mathbf{x}^TA\mathbf{x}$. Verify that $t = (1-\alpha)m + \alpha M$ for some number α between 0 and 1. Then let $\mathbf{x} = \sqrt{1-\alpha}\,\mathbf{u}_n + \sqrt{\alpha}\,\mathbf{u}_1$, and show that $\mathbf{x}^T\mathbf{x} = 1$ and $\mathbf{x}^TA\mathbf{x} = t$.

[M] In Exercises 14–17, follow the instructions given for Exercises 3–6.

14. $3x_1x_2 + 5x_1x_3 + 7x_1x_4 + 7x_2x_3 + 5x_2x_4 + 3x_3x_4$

15. $4x_1^2-6x_1x_2-10x_1x_3-10x_1x_4-6x_2x_3-6x_2x_4-2x_3x_4$

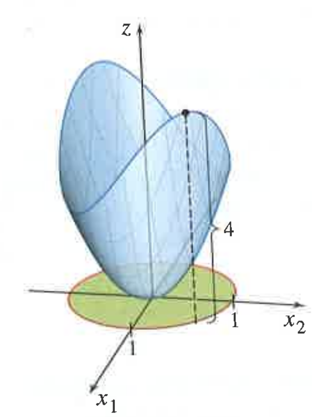
16. $-6x_1^2-10x_2^2-13x_3^2-13x_4^2-4x_1x_2-4x_1x_3-4x_1x_4+6x_3x_4$

17. $x_1x_2 + 3x_1x_3 + 30x_1x_4 + 30x_2x_3 + 3x_2x_4 + x_3x_4$

SOLUTIONS TO PRACTICE PROBLEMS

1. The matrix of the quadratic form is $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$. It is easy to find the eigenvalues, 4 and 2, and corresponding unit eigenvectors, $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. So the desired change of variable is $\mathbf{x} = P\mathbf{y}$, where $P = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. (A common error here is to forget to normalize the eigenvectors.) The new quadratic form is $\mathbf{y}^TD\mathbf{y} = 4y_1^2 + 2y_2^2$.

2. The maximum of $Q(\mathbf{x})$, for a unit vector $\mathbf{x}$, is 4 and the maximum is attained at the unit eigenvector $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. [A common incorrect answer is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This vector maximizes the quadratic form $\mathbf{y}^TD\mathbf{y}$ instead of $Q(\mathbf{x})$.]



The maximum value of $Q(\mathbf{x})$ subject to $\mathbf{x}^T\mathbf{x} = 1$ is 4.

7.4 THE SINGULAR VALUE DECOMPOSITION

The diagonalization theorems in Sections 5.3 and 7.1 play a part in many interesting applications. Unfortunately, as we know, not all matrices can be factored as $A = PDP^{-1}$ with D diagonal. However, a factorization $A = QDP^{-1}$ *is* possible for any $m \times n$ matrix A! A special factorization of this type, called the *singular value decomposition*, is one of the most useful matrix factorizations in applied linear algebra.

The singular value decomposition is based on the following property of the ordinary diagonalization that can be imitated for rectangular matrices: The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks

certain vectors (the eigenvectors). If $A\mathbf{x} = \lambda\mathbf{x}$ and $\|\mathbf{x}\| = 1$, then

$$\|A\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\| = |\lambda| \tag{1}$$

If λ_1 is the eigenvalue with the greatest magnitude, then a corresponding unit eigenvector $\mathbf{v}_1$ identifies a direction in which the stretching effect of A is greatest. That is, the length of $A\mathbf{x}$ is maximized when $\mathbf{x} = \mathbf{v}_1$, and $\|A\mathbf{v}_1\| = |\lambda_1|$, by (1). This description of $\mathbf{v}_1$ and $|\lambda_1|$ has an analogue for rectangular matrices that will lead to the singular value decomposition.

EXAMPLE 1 If $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$, then the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps the unit sphere $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$ in $\mathbb{R}^3$ onto an ellipse in $\mathbb{R}^2$, shown in Figure 1. Find a unit vector $\mathbf{x}$ at which the length $\|A\mathbf{x}\|$ is maximized, and compute this maximum length.

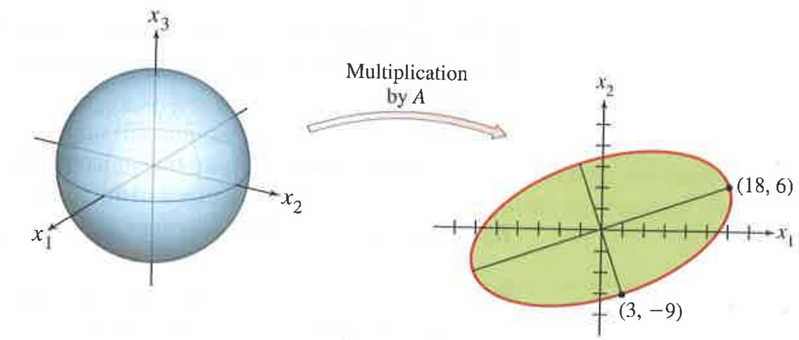


FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

SOLUTION The quantity $\|A\mathbf{x}\|^2$ is maximized at the same $\mathbf{x}$ that maximizes $\|A\mathbf{x}\|$, and $\|A\mathbf{x}\|^2$ is easier to study. Observe that

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

Also, A^TA is a symmetric matrix, since $(A^TA)^T = A^TA^{TT} = A^TA$. So the problem now is to maximize the quadratic form $\mathbf{x}^T(A^TA)\mathbf{x}$ subject to the constraint $\|\mathbf{x}\| = 1$. By Theorem 6 in Section 7.3, the maximum value is the greatest eigenvalue λ_1 of A^TA. Also, the maximum value is attained at a unit eigenvector of A^TA corresponding to λ_1.

For the matrix A in this example,

$$A^TA = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

The eigenvalues of A^TA are $\lambda_1 = 360$, $\lambda_2 = 90$, and $\lambda_3 = 0$. Corresponding unit eigenvectors are, respectively,

$$\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

The maximum value of $\|A\mathbf{x}\|^2$ is 360, attained when $\mathbf{x}$ is the unit vector $\mathbf{v}_1$. The vector $A\mathbf{v}_1$ is a point on the ellipse in Figure 1 farthest from the origin, namely,

$$A\mathbf{v}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}\begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}$$

For $\|\mathbf{x}\| = 1$, the maximum value of $\|A\mathbf{x}\|$ is $\|A\mathbf{v}_1\| = \sqrt{360} = 6\sqrt{10}$. ■

Example 1 suggests that the effect of A on the unit sphere in $\mathbb{R}^3$ is related to the quadratic form $\mathbf{x}^T(A^TA)\mathbf{x}$. In fact, the entire geometric behavior of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is captured by this quadratic form, as we shall see.

The Singular Values of an $m \times n$ Matrix

Let A be an $m \times n$ matrix. Then A^TA is symmetric and can be orthogonally diagonalized. Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A^TA, and let $\lambda_1, \dots, \lambda_n$ be the associated eigenvalues of A^TA. Then, for $1 \le i \le n$,

$$\begin{aligned} \|A\mathbf{v}_i\|^2 &= (A\mathbf{v}_i)^TA\mathbf{v}_i = \mathbf{v}_i^TA^TA\mathbf{v}_i \\ &= \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) && \text{Since } \mathbf{v}_i \text{ is an eigenvector of } A^TA \\ &= \lambda_i && \text{Since } \mathbf{v}_i \text{ is a unit vector} \end{aligned} \tag{2}$$

So the eigenvalues of A^TA are all nonnegative. By renumbering, if necessary, we may assume that the eigenvalues are arranged so that

$$\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_n \ge 0$$

The **singular values** of A are the square roots of the eigenvalues of A^TA, denoted by $\sigma_1, \dots, \sigma_n$, and they are arranged in decreasing order. That is, $\sigma_i = \sqrt{\lambda_i}$ for $1 \le i \le n$. By equation (2), *the singular values of A are the lengths of the vectors* $A\mathbf{v}_1, \dots, A\mathbf{v}_n$.

EXAMPLE 2 Let A be the matrix in Example 1. Since the eigenvalues of A^TA are 360, 90, and 0, the singular values of A are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \quad \sigma_2 = \sqrt{90} = 3\sqrt{10}, \quad \sigma_3 = 0$$

From Example 1, the first singular value of A is the maximum of $\|A\mathbf{x}\|$ over all unit vectors, and the maximum is attained at the unit eigenvector $\mathbf{v}_1$. Theorem 7 in Section 7.3 shows that the second singular value of A is the maximum of $\|A\mathbf{x}\|$ over all unit vectors that are *orthogonal to* $\mathbf{v}_1$, and this maximum is attained at the second unit eigenvector, $\mathbf{v}_2$ (Exercise 22). For the $\mathbf{v}_2$ in Example 1,

$$A\mathbf{v}_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}\begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}$$

This point is on the minor axis of the ellipse in Figure 1, just as $A\mathbf{v}_1$ is on the major axis. (See Figure 2.) The first two singular values of A are the lengths of the major and minor semiaxes of the ellipse. ■

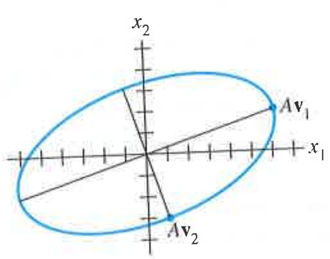


FIGURE 2

The fact that $A\mathbf{v}_1$ and $A\mathbf{v}_2$ are orthogonal in Figure 2 is no accident, as the next theorem shows.

THEOREM 9

Suppose $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA, arranged so that the corresponding eigenvalues of A^TA satisfy $\lambda_1 \ge \dots \ge \lambda_n$, and suppose A has r nonzero singular values. Then $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A, and rank $A = r$.

PROOF Because $\mathbf{v}_i$ and $\lambda_j\mathbf{v}_j$ are orthogonal for $i \ne j$,

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

Thus $\{A\mathbf{v}_1, \dots, A\mathbf{v}_n\}$ is an orthogonal set. Furthermore, since the lengths of the vectors $A\mathbf{v}_1, \dots, A\mathbf{v}_n$ are the singular values of A, and since there are r nonzero singular values, $A\mathbf{v}_i \ne \mathbf{0}$ if and only if $1 \le i \le r$. So $A\mathbf{v}_1, \dots, A\mathbf{v}_r$ are linearly independent vectors, and they are in Col A. Finally, for any $\mathbf{y}$ in Col A—say, $\mathbf{y} = A\mathbf{x}$—we can write $\mathbf{x} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$, and

$$\begin{aligned} \mathbf{y} = A\mathbf{x} &= c_1A\mathbf{v}_1 + \dots + c_rA\mathbf{v}_r + c_{r+1}A\mathbf{v}_{r+1} + \dots + c_nA\mathbf{v}_n \\ &= c_1A\mathbf{v}_1 + \dots + c_rA\mathbf{v}_r + 0 + \dots + 0 \end{aligned}$$

Thus $\mathbf{y}$ is in Span$\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$, which shows that $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an (orthogonal) basis for Col A. Hence rank $A = \dim \operatorname{Col} A = r$. ■

NUMERICAL NOTE

In some cases, the rank of A may be very sensitive to small changes in the entries of A. The obvious method of counting the number of pivot columns in A does not work well if A is row reduced by a computer. Roundoff error often creates an echelon form with full rank.

In practice, the most reliable way to estimate the rank of a large matrix A is to count the number of nonzero singular values. In this case, extremely small nonzero singular values are assumed to be zero for all practical purposes, and the *effective rank* of the matrix is the number obtained by counting the remaining nonzero singular values.[1]

The Singular Value Decomposition

The decomposition of A involves an $m \times n$ "diagonal" matrix Σ of the form

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{array}{l} \\ \leftarrow m - r \text{ rows} \end{array} \tag{3}$$
$$\qquad\qquad \uparrow\, n - r \text{ columns}$$

where D is an $r \times r$ diagonal matrix for some r not exceeding the smaller of m and n. (If r equals m or n or both, some or all of the zero matrices do not appear.)

THEOREM 10

The Singular Value Decomposition

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix Σ as in (3) for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_r > 0$, and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

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

Any factorization $A = U\Sigma V^T$, with U and V orthogonal, Σ as in (3), and positive diagonal entries in D, is called a **singular value decomposition** (or **SVD**) of A. The matrices U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A. See Exercise 19. The columns of U in such a decomposition are called **left singular vectors** of A, and the columns of V are called **right singular vectors** of A.

[1] In general, rank estimation is not a simple problem. For a discussion of the subtle issues involved, see Philip E. Gill, Walter Murray, and Margaret H. Wright, *Numerical Linear Algebra and Optimization*, vol. 1 (Redwood City, CA: Addison-Wesley, 1991), Sec. 5.8.

PROOF Let λ_i and $\mathbf{v}_i$ be as in Theorem 9, so that $\{A\mathbf{v}_1, \dots, A\mathbf{v}_r\}$ is an orthogonal basis for Col A. Normalize each $A\mathbf{v}_i$ to obtain an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$, where

$$\mathbf{u}_i = \frac{1}{\|A\mathbf{v}_i\|}A\mathbf{v}_i = \frac{1}{\sigma_i}A\mathbf{v}_i$$

and

$$A\mathbf{v}_i = \sigma_i \mathbf{u}_i \qquad (1 \le i \le r) \tag{4}$$

Now extend $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ of $\mathbb{R}^m$, and let

$$U = [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_m\,] \quad \text{and} \quad V = [\,\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \cdots \;\; \mathbf{v}_n\,]$$

By construction, U and V are orthogonal matrices. Also, from (4),

$$AV = [\,A\mathbf{v}_1 \;\; \cdots \;\; A\mathbf{v}_r \;\; \mathbf{0} \;\; \cdots \;\; \mathbf{0}\,] = [\,\sigma_1\mathbf{u}_1 \;\; \cdots \;\; \sigma_r\mathbf{u}_r \;\; \mathbf{0} \;\; \cdots \;\; \mathbf{0}\,]$$

Let D be the diagonal matrix with diagonal entries $\sigma_1, \dots, \sigma_r$, and let Σ be as in (3) above. Then

$$U\Sigma = [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_m\,]\left[\begin{array}{cccc|c} \sigma_1 & & & 0 & \\ & \sigma_2 & & & 0 \\ & & \ddots & & \\ 0 & & & \sigma_r & \\ \hline & & 0 & & 0 \end{array}\right]$$

$$= [\,\sigma_1\mathbf{u}_1 \;\; \cdots \;\; \sigma_r\mathbf{u}_r \;\; \mathbf{0} \;\; \cdots \;\; \mathbf{0}\,]$$
$$= AV$$

Since V is an orthogonal matrix, $U\Sigma V^T = AVV^T = A$. ∎

The next two examples focus attention on the internal structure of a singular value decomposition. An efficient and numerically stable algorithm for this decomposition would use a different approach. See the Numerical Note at the end of the section.

EXAMPLE 3 Use the results of Examples 1 and 2 to construct a singular value decomposition of $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$.

SOLUTION A construction can be divided into three steps.

SG Computing an SVD 7-10

Step 1. Find an orthogonal diagonalization of A^TA. That is, find the eigenvalues of A^TA and a corresponding orthonormal set of eigenvectors. If A had only two columns, the calculations could be done by hand. Larger matrices usually require a matrix program.[2] However, for the matrix A here, the eigendata for A^TA are provided in Example 1.

Step 2. Set up V and Σ. Arrange the eigenvalues of A^TA in decreasing order. In Example 1, the eigenvalues are already listed in decreasing order: 360, 90, and 0. The corresponding unit eigenvectors, $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$, are the right singular vectors of A. Using Example 1, construct

$$V = [\,\mathbf{v}_1 \;\; \mathbf{v}_2 \;\; \mathbf{v}_3\,] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$$

[2] See the *Study Guide* for software and graphing calculator commands. MATLAB, for instance, can produce both the eigenvalues and the eigenvectors with one command, `eig`.

The square roots of the eigenvalues are the singular values:

$$\sigma_1 = 6\sqrt{10}, \quad \sigma_2 = 3\sqrt{10}, \quad \sigma_3 = 0$$

The nonzero singular values are the diagonal entries of D. The matrix Σ is the same size as A, with D in its upper left corner and with 0's elsewhere.

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}, \qquad \Sigma = [\,D \;\; 0\,] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

Step 3. Construct U. When A has rank r, the first r columns of U are the normalized vectors obtained from $A\mathbf{v}_1, \dots, A\mathbf{v}_r$. In this example, A has two nonzero singular values, so rank $A = 2$. Recall from equation (2) and the paragraph before Example 2 that $\|A\mathbf{v}_1\| = \sigma_1$ and $\|A\mathbf{v}_2\| = \sigma_2$. Thus

$$\mathbf{u}_1 = \frac{1}{\sigma_1}A\mathbf{v}_1 = \frac{1}{6\sqrt{10}}\begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{3\sqrt{10}}\begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

Note that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is already a basis for $\mathbb{R}^2$. Thus no additional vectors are needed for U, and $U = [\,\mathbf{u}_1 \;\; \mathbf{u}_2\,]$. The singular value decomposition of A is

$$A = \underset{\substack{\uparrow \\ U}}{\begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}} \underset{\substack{\uparrow \\ \Sigma}}{\begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}} \underset{\substack{\uparrow \\ V^T}}{\begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}}$$ ∎

EXAMPLE 4 Find a singular value decomposition of $A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$.

SOLUTION First, compute $A^TA = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$. The eigenvalues of A^TA are 18 and 0, with corresponding unit eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

These unit vectors form the columns of V:

$$V = [\,\mathbf{v}_1 \;\; \mathbf{v}_2\,] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

The singular values are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$. Since there is only one nonzero singular value, the "matrix" D may be written as a single number. That is, $D = 3\sqrt{2}$. The matrix Σ is the same size as A, with D in its upper left corner:

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

To construct U, first construct $A\mathbf{v}_1$ and $A\mathbf{v}_2$:

$$A\mathbf{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad A\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

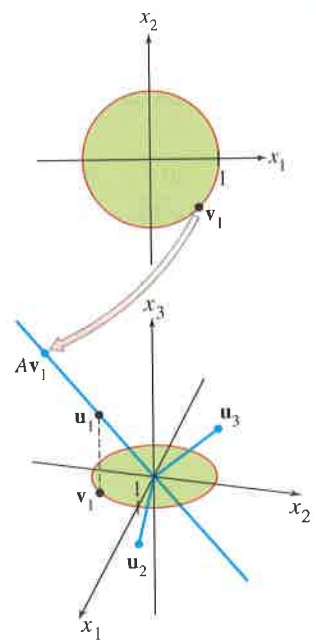


FIGURE 3

As a check on the calculations, verify that $\|A\mathbf{v}_1\| = \sigma_1 = 3\sqrt{2}$. Of course, $A\mathbf{v}_2 = \mathbf{0}$ because $\|A\mathbf{v}_2\| = \sigma_2 = 0$. The only column found for U so far is

$$\mathbf{u}_1 = \frac{1}{3\sqrt{2}}A\mathbf{v}_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

The other columns of U are found by extending the set $\{\mathbf{u}_1\}$ to an orthonormal basis for $\mathbb{R}^3$. In this case, we need two orthogonal unit vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ that are orthogonal to $\mathbf{u}_1$. (See Figure 3.) Each vector must satisfy $\mathbf{u}_1^T\mathbf{x} = 0$, which is equivalent to the equation $x_1 - 2x_2 + 2x_3 = 0$. A basis for the solution set of this equation is

$$\mathbf{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

(Check that $\mathbf{w}_1$ and $\mathbf{w}_2$ are each orthogonal to $\mathbf{u}_1$.) Apply the Gram–Schmidt process (with normalizations) to $\{\mathbf{w}_1, \mathbf{w}_2\}$, and obtain

$$\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

Finally, set $U = [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \mathbf{u}_3\,]$, take Σ and V^T from above, and write

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

■

Applications of the Singular Value Decomposition

The SVD is often used to estimate the rank of a matrix, as noted above. Several other numerical applications are described briefly below, and an application to image processing is presented in Section 7.5.

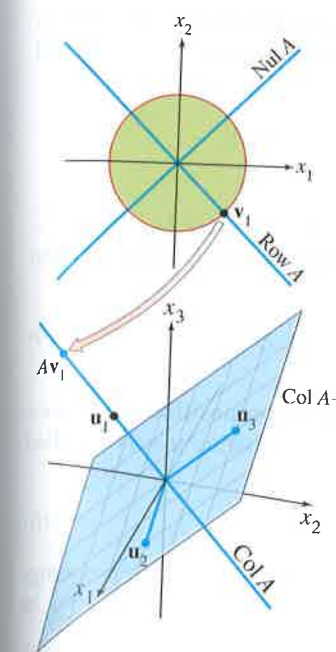
EXAMPLE 5 (The Condition Number) Most numerical calculations involving an equation $A\mathbf{x} = \mathbf{b}$ are as reliable as possible when the SVD of A is used. The two orthogonal matrices U and V do not affect lengths of vectors or angles between vectors (Theorem 7 in Section 6.2). Any possible instabilities in numerical calculations are identified in Σ. If the singular values of A are extremely large or small, roundoff errors are almost inevitable, but an error analysis is aided by knowing the entries in Σ and V.

If A is an invertible $n \times n$ matrix, then the ratio σ_1/σ_n of the largest and smallest singular values gives the **condition number** of A. Exercises 41–43 in Section 2.3 showed how the condition number affects the sensitivity of a solution of $A\mathbf{x} = \mathbf{b}$ to changes (or errors) in the entries of A. (Actually, a "condition number" of A can be computed in several ways, but the definition given here is widely used for studying $A\mathbf{x} = \mathbf{b}$.) ■

EXAMPLE 6 (Bases for Fundamental Subspaces) Given an SVD for an $m \times n$ matrix A, let $\mathbf{u}_1, \ldots, \mathbf{u}_m$ be the left singular vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ the right singular vectors, and $\sigma_1, \ldots, \sigma_n$ the singular values, and let r be the rank of A. By Theorem 9,

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \tag{5}$$

is an orthonormal basis for Col A.



The fundamental subspaces in Example 4.

Recall from Theorem 3 in Section 6.1 that $(\text{Col } A)^\perp = \text{Nul } A^T$. Hence

$$\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\} \tag{6}$$

is an orthonormal basis for Nul A^T.

Since $\|A\mathbf{v}_i\| = \sigma_i$ for $1 \le i \le n$, and σ_i is 0 if and only if $i > r$, the vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ span a subspace of Nul A of dimension $n - r$. By the Rank Theorem, dim Nul $A = n -$ rank A. It follows that

$$\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\} \tag{7}$$

is an orthonormal basis for Nul A, by the Basis Theorem (in Section 4.5).

From (5) and (6), the orthogonal complement of Nul A^T is Col A. Interchanging A and A^T, note that $(\text{Nul } A)^\perp = \text{Col } A^T = \text{Row } A$. Hence, from (7),

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} \tag{8}$$

is an orthonormal basis for Row A.

Figure 4 summarizes (5)–(8), but shows the orthogonal basis $\{\sigma_1\mathbf{u}_1, \ldots, \sigma_r\mathbf{u}_r\}$ for Col A instead of the normalized basis, to remind you that $A\mathbf{v}_i = \sigma_i\mathbf{u}_i$ for $1 \le i \le r$. Explicit orthonormal bases for the four fundamental subspaces determined by A are useful in some calculations, particularly in constrained optimization problems. ■

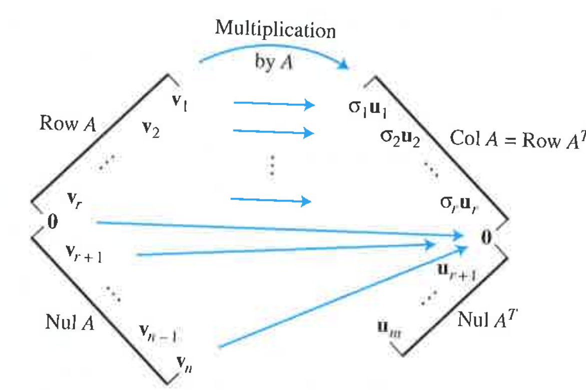


FIGURE 4 The four fundamental subspaces and the action of A.

The four fundamental subspaces and the concept of singular values provide the final statements of the Invertible Matrix Theorem. (Recall that statements about A^T have been omitted from the theorem, to avoid nearly doubling the number of statements.) The other statements were given in Sections 2.3, 2.9, 3.2, 4.6, and 5.2.

THEOREM

The Invertible Matrix Theorem (concluded)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.

u. $(\text{Col } A)^\perp = \{\mathbf{0}\}$.
v. $(\text{Nul } A)^\perp = \mathbb{R}^n$.
w. Row $A = \mathbb{R}^n$.
x. A has n nonzero singular values.

EXAMPLE 7 (Reduced SVD and the Pseudoinverse of A) When Σ contains rows or columns of zeros, a more compact decomposition of A is possible. Using the notation established above, let $r = \text{rank } A$, and partition U and V into submatrices whose first blocks contain r columns:

$$U = [\,U_r \quad U_{m-r}\,], \quad \text{where } U_r = [\,\mathbf{u}_1 \ \cdots \ \mathbf{u}_r\,]$$
$$V = [\,V_r \quad V_{n-r}\,], \quad \text{where } V_r = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_r\,]$$

Then U_r is $m \times r$ and V_r is $n \times r$. (To simplify notation, we consider U_{m-r} or V_{n-r} even though one of them may have no columns.) Then partitioned matrix multiplication shows that

$$A = [\,U_r \quad U_{m-r}\,]\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_r^T \\ V_{n-r}^T \end{bmatrix} = U_r D V_r^T \tag{9}$$

This factorization of A is called a **reduced singular value decomposition** of A. Since the diagonal entries in D are nonzero, D is invertible. The following matrix is called the **pseudoinverse** (also, the **Moore–Penrose inverse**) of A:

$$A^+ = V_r D^{-1} U_r^T \tag{10}$$

Supplementary Exercises 12–14 at the end of the chapter explore some of the properties of the reduced singular value decomposition and the pseudoinverse. ■

EXAMPLE 8 (Least-Squares Solution) Given the equation $A\mathbf{x} = \mathbf{b}$, use the pseudoinverse of A in (10) to define

$$\hat{\mathbf{x}} = A^+\mathbf{b} = V_r D^{-1} U_r^T \mathbf{b}$$

Then, from the SVD in (9),

$$\begin{aligned} A\hat{\mathbf{x}} &= (U_r D V_r^T)(V_r D^{-1} U_r^T \mathbf{b}) \\ &= U_r D D^{-1} U_r^T \mathbf{b} \qquad \text{Because } V_r^T V_r = I_r \\ &= U_r U_r^T \mathbf{b} \end{aligned}$$

It follows from (5) that $U_r U_r^T \mathbf{b}$ is the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto Col A. (See Theorem 10 in Section 6.3.) Thus $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. In fact, this $\hat{\mathbf{x}}$ has the smallest length among all least-squares solutions of $A\mathbf{x} = \mathbf{b}$. See Supplementary Exercise 14. ■

NUMERICAL NOTE

Examples 1–4 and the exercises illustrate the concept of singular values and suggest how to perform calculations by hand. In practice, the computation of A^TA should be avoided, since any errors in the entries of A are squared in the entries of A^TA. There exist fast iterative methods that produce the singular values and singular vectors of A accurately to many decimal places.

Further Reading

Horn, Roger A., and Charles R. Johnson, *Matrix Analysis* (Cambridge: Cambridge University Press, 1990).

Long, Cliff, "Visualization of Matrix Singular Value Decomposition." *Mathematics Magazine* **56** (1983), pp. 161–167.

Moler, C. B., and D. Morrison, "Singular Value Analysis of Cryptograms." *Amer. Math. Monthly* **90** (1983), pp. 78–87.

Strang, Gilbert, *Linear Algebra and Its Applications*, 4th ed. (Belmont, CA: Brooks/Cole, 2005).

Watkins, David S., *Fundamentals of Matrix Computations* (New York: Wiley, 1991), pp. 390–398, 409–421.

PRACTICE PROBLEMS

WEB

1. Given a singular value decomposition, $A = U\Sigma V^T$, find an SVD of A^T. How are the singular values of A and A^T related?
2. For any $n \times n$ matrix A, use the SVD to show that there is an $n \times n$ orthogonal matrix Q such that $A^TA = Q^T(A^TA)Q$.

Remark: Practice Problem 2 establishes that for any $n \times n$ matrix A, the matrices AA^T and A^TA are *orthogonally similar*.

7.4 EXERCISES

Find the singular values of the matrices in Exercises 1–4.

1. $\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix}$
2. $\begin{bmatrix} -3 & 0 \\ 0 & 0 \end{bmatrix}$
3. $\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$
4. $\begin{bmatrix} 3 & 0 \\ 8 & 3 \end{bmatrix}$

Find an SVD of each matrix in Exercises 5–12. [*Hint:* In Exercise 11, one choice for U is $\begin{bmatrix} -1/3 & 2/3 & 2/3 \\ 2/3 & -1/3 & 2/3 \\ 2/3 & 2/3 & -1/3 \end{bmatrix}$. In Exercise 12, one column of U can be $\begin{bmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.]

5. $\begin{bmatrix} -2 & 0 \\ 0 & 0 \end{bmatrix}$
6. $\begin{bmatrix} -3 & 0 \\ 0 & -2 \end{bmatrix}$
7. $\begin{bmatrix} 2 & -1 \\ 2 & 2 \end{bmatrix}$
8. $\begin{bmatrix} 4 & 6 \\ 0 & 4 \end{bmatrix}$
9. $\begin{bmatrix} 3 & -3 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}$
10. $\begin{bmatrix} 7 & 1 \\ 5 & 5 \\ 0 & 0 \end{bmatrix}$
11. $\begin{bmatrix} -3 & 1 \\ 6 & -2 \\ 6 & -2 \end{bmatrix}$
12. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}$
13. Find the SVD of $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$ [*Hint:* Work with A^T.]
14. In Exercise 7, find a unit vector $\mathbf{x}$ at which $A\mathbf{x}$ has maximum length.
15. Suppose the factorization below is an SVD of a matrix A, with the entries in U and V rounded to two decimal places.

$$A = \begin{bmatrix} .40 & -.78 & .47 \\ .37 & -.33 & -.87 \\ -.84 & -.52 & -.16 \end{bmatrix}\begin{bmatrix} 7.10 & 0 & 0 \\ 0 & 3.10 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} .30 & -.51 & -.81 \\ .76 & .64 & -.12 \\ .58 & -.58 & .58 \end{bmatrix}$$

   a. What is the rank of A?
   b. Use this decomposition of A, with no calculations, to write a basis for Col A and a basis for Nul A. [*Hint:* First write the columns of V.]

16. Repeat Exercise 15 for the following SVD of a 3×4 matrix A:

$$A = \begin{bmatrix} -.86 & -.11 & -.50 \\ .31 & .68 & -.67 \\ .41 & -.73 & -.55 \end{bmatrix}\begin{bmatrix} 12.48 & 0 & 0 & 0 \\ 0 & 6.34 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} .66 & -.03 & -.35 & .66 \\ -.13 & -.90 & -.39 & -.13 \\ .65 & .08 & -.16 & -.73 \\ -.34 & .42 & -.84 & -.08 \end{bmatrix}$$

In Exercises 17–24, A is an $m \times n$ matrix with a singular value decomposition $A = U\Sigma V^T$, where U is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ "diagonal" matrix with r positive entries and no negative entries, and V is an $n \times n$ orthogonal matrix. Justify each answer.

17. Show that if A is square, then $|\det A|$ is the product of the singular values of A.
18. Suppose A is square and invertible. Find a singular value decomposition of A^{-1}.
19. Show that the columns of V are eigenvectors of A^TA, the columns of U are eigenvectors of AA^T, and the diagonal

entries of Σ are the singular values of A. [*Hint:* Use the SVD to compute A^TA and AA^T.]

20. Show that if P is an orthogonal $m \times m$ matrix, then PA has the same singular values as A.

21. Justify the statement in Example 2 that the second singular value of a matrix A is the maximum of $\|A\mathbf{x}\|$ as $\mathbf{x}$ varies over all unit vectors orthogonal to $\mathbf{v}_1$, with $\mathbf{v}_1$ a right singular vector corresponding to the first singular value of A. [*Hint:* Use Theorem 7 in Section 7.3.]

22. Show that if A is an $n \times n$ positive definite matrix, then an orthogonal diagonalization $A = PDP^T$ is a singular value decomposition of A.

23. Let $U = [\,\mathbf{u}_1 \ \cdots \ \mathbf{u}_m\,]$ and $V = [\,\mathbf{v}_1 \ \cdots \ \mathbf{v}_n\,]$, where the $\mathbf{u}_i$ and $\mathbf{v}_i$ are as in Theorem 10. Show that

$$A = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T.$$

24. Using the notation of Exercise 23, show that $A^T\mathbf{u}_j = \sigma_j\mathbf{v}_j$ for $1 \le j \le r = \text{rank } A$.

25. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Describe how to find a basis $\mathcal{B}$ for $\mathbb{R}^n$ and a basis $\mathcal{C}$ for $\mathbb{R}^m$ such that the matrix for T relative to $\mathcal{B}$ and $\mathcal{C}$ is an $m \times n$ "diagonal" matrix.

[**M**] Compute an SVD of each matrix in Exercises 26 and 27. Report the final matrix entries accurate to two decimal places. Use the method of Examples 3 and 4.

26. $A = \begin{bmatrix} -18 & 13 & -4 & 4 \\ 2 & 19 & -4 & 12 \\ -14 & 11 & -12 & 8 \\ -2 & 21 & 4 & 8 \end{bmatrix}$

27. $A = \begin{bmatrix} 6 & -8 & -4 & 5 & -4 \\ 2 & 7 & -5 & -6 & 4 \\ 0 & -1 & -8 & 2 & 2 \\ -1 & -2 & 4 & 4 & -8 \end{bmatrix}$

28. [**M**] Compute the singular values of the 4×4 matrix in Exercise 9 in Section 2.3, and compute the condition number σ_1/σ_4.

29. [**M**] Compute the singular values of the 5×5 matrix in Exercise 10 in Section 2.3, and compute the condition number σ_1/σ_5.

SOLUTIONS TO PRACTICE PROBLEMS

1. If $A = U\Sigma V^T$, where Σ is $m \times n$, then $A^T = (V^T)^T\Sigma^TU^T = V\Sigma^TU^T$. This is an SVD of A^T because V and U are orthogonal matrices and Σ^T is an $n \times m$ "diagonal" matrix. Since Σ and Σ^T have the same nonzero diagonal entries, A and A^T have the same nonzero singular values. [*Note:* If A is $2 \times n$, then AA^T is only 2×2 and its eigenvalues may be easier to compute (by hand) than the eigenvalues of A^TA.]
2. Use the SVD to write $A = U\Sigma V^T$, where U and V are $n \times n$ orthogonal matrices and Σ is an $n \times n$ diagonal matrix. Notice that $U^TU = I = V^TV$ and $\Sigma^T = \Sigma$, since U and V are orthogonal matrices and Σ is a diagonal matrix. Substituting the SVD for A into AA^T and A^TA results in

$$AA^T = U\Sigma V^T(U\Sigma V^T)^T = U\Sigma V^TV\Sigma^TU^T = U\Sigma\Sigma^TU^T = U\Sigma^2U^T,$$

and

$$A^TA = (U\Sigma V^T)^TU\Sigma V^T = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T = V\Sigma^2V^T.$$

Let $Q = VU^T$. Then

$$Q^T(A^TA)Q = (VU^T)^T(V\Sigma^2V^T)(VU^T) = UV^TV\Sigma^2V^TVU^T = U\Sigma^2U^T = AA^T.$$

7.5 APPLICATIONS TO IMAGE PROCESSING AND STATISTICS

The satellite photographs in this chapter's introduction provide an example of multidimensional, or *multivariate*, data—information organized so that each datum in the data set is identified with a point (vector) in $\mathbb{R}^n$. The main goal of this section is to explain a technique, called *principal component analysis*, used to analyze such multivariate data. The calculations will illustrate the use of orthogonal diagonalization and the singular value decomposition.

Principal component analysis can be applied to any data that consist of lists of measurements made on a collection of objects or individuals. For instance, consider a chemical process that produces a plastic material. To monitor the process, 300 samples are taken of the material produced, and each sample is subjected to a battery of eight tests, such as melting point, density, ductility, tensile strength, and so on. The laboratory report for each sample is a vector in $\mathbb{R}^8$, and the set of such vectors forms an 8×300 matrix, called the **matrix of observations**.

Loosely speaking, we can say that the process control data are eight-dimensional. The next two examples describe data that can be visualized graphically.

EXAMPLE 1 An example of two-dimensional data is given by a set of weights and heights of N college students. Let $\mathbf{X}_j$ denote the **observation vector** in $\mathbb{R}^2$ that lists the weight and height of the jth student. If w denotes weight and h height, then the matrix of observations has the form

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_N \\ h_1 & h_2 & \cdots & h_N \end{bmatrix}$$
$$\uparrow \qquad \uparrow \qquad\qquad \uparrow$$
$$\mathbf{X}_1 \quad \mathbf{X}_2 \qquad\quad \mathbf{X}_N$$

The set of observation vectors can be visualized as a two-dimensional *scatter plot*. See Figure 1. ■

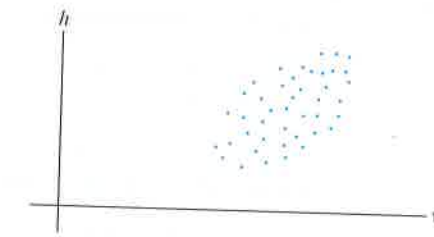


FIGURE 1 A scatter plot of observation vectors $\mathbf{X}_1, \ldots, \mathbf{X}_N$.

EXAMPLE 2 The first three photographs of Railroad Valley, Nevada, shown in the chapter introduction can be viewed as *one* image of the region, with *three spectral components*, because simultaneous measurements of the region were made at three separate wavelengths. Each photograph gives different information about the same physical region. For instance, the first pixel in the upper-left corner of each photograph corresponds to the same place on the ground (about 30 meters by 30 meters). To each pixel there corresponds an observation vector in $\mathbb{R}^3$ that lists the signal intensities for that pixel in the three spectral bands.

Typically, the image is 2000×2000 pixels, so there are 4 million pixels in the image. The data for the image form a matrix with 3 rows and 4 million columns (with columns arranged in any convenient order). In this case, the "multidimensional" character of the data refers to the three *spectral* dimensions rather than the two *spatial* dimensions that naturally belong to any photograph. The data can be visualized as a cluster of 4 million points in $\mathbb{R}^3$, perhaps as in Figure 2. ■

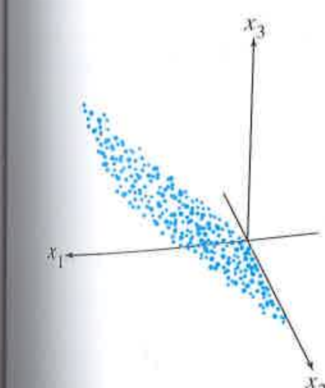


FIGURE 2 A scatter plot of spectral data for a satellite image.

Mean and Covariance

To prepare for principal component analysis, let $[\,\mathbf{X}_1 \ \cdots \ \mathbf{X}_N\,]$ be a $p \times N$ matrix of observations, such as described above. The **sample mean**, $\mathbf{M}$, of the observation vectors

$\mathbf{X}_1, \dots, \mathbf{X}_N$ is given by

$$\mathbf{M} = \frac{1}{N}(\mathbf{X}_1 + \cdots + \mathbf{X}_N)$$

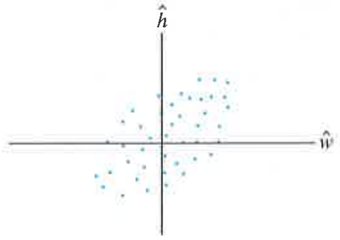


FIGURE 3
Weight–height data in mean-deviation form.

For the data in Figure 1, the sample mean is the point in the "center" of the scatter plot. For $k = 1, \dots, N$, let

$$\hat{\mathbf{X}}_k = \mathbf{X}_k - \mathbf{M}$$

The columns of the $p \times N$ matrix

$$B = [\,\hat{\mathbf{X}}_1 \;\; \hat{\mathbf{X}}_2 \;\; \cdots \;\; \hat{\mathbf{X}}_N\,]$$

have a zero sample mean, and B is said to be in **mean-deviation form**. When the sample mean is subtracted from the data in Figure 1, the resulting scatter plot has the form in Figure 3.

The **(sample) covariance matrix** is the $p \times p$ matrix S defined by

$$S = \frac{1}{N-1}BB^T$$

Since any matrix of the form BB^T is positive semidefinite, so is S. (See Exercise 25 in Section 7.2 with B and B^T interchanged.)

EXAMPLE 3 Three measurements are made on each of four individuals in a random sample from a population. The observation vectors are

$$\mathbf{X}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{X}_2 = \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix}, \quad \mathbf{X}_3 = \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix}, \quad \mathbf{X}_4 = \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix}$$

Compute the sample mean and the covariance matrix.

SOLUTION The sample mean is

$$\mathbf{M} = \frac{1}{4}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 13 \end{bmatrix} + \begin{bmatrix} 7 \\ 8 \\ 1 \end{bmatrix} + \begin{bmatrix} 8 \\ 4 \\ 5 \end{bmatrix} \right) = \frac{1}{4}\begin{bmatrix} 20 \\ 16 \\ 20 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \\ 5 \end{bmatrix}$$

Subtract the sample mean from $\mathbf{X}_1, \dots, \mathbf{X}_4$ to obtain

$$\hat{\mathbf{X}}_1 = \begin{bmatrix} -4 \\ -2 \\ -4 \end{bmatrix}, \quad \hat{\mathbf{X}}_2 = \begin{bmatrix} -1 \\ -2 \\ 8 \end{bmatrix}, \quad \hat{\mathbf{X}}_3 = \begin{bmatrix} 2 \\ 4 \\ -4 \end{bmatrix}, \quad \hat{\mathbf{X}}_4 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

and

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

The sample covariance matrix is

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

$$= \frac{1}{3}\begin{bmatrix} 30 & 18 & 0 \\ 18 & 24 & -24 \\ 0 & -24 & 96 \end{bmatrix} = \begin{bmatrix} 10 & 6 & 0 \\ 6 & 8 & -8 \\ 0 & -8 & 32 \end{bmatrix}$$

To discuss the entries in $S = [s_{ij}]$, let $\mathbf{X}$ represent a vector that varies over the set of observation vectors and denote the coordinates of $\mathbf{X}$ by $x_1, \dots, x_p$. Then x_1, for example, is a scalar that varies over the set of first coordinates of $\mathbf{X}_1, \dots, \mathbf{X}_N$. For $j = 1, \dots, p$, the diagonal entry s_{jj} in S is called the **variance** of x_j.

The variance of x_j measures the spread of the values of x_j. (See Exercise 13.) In Example 3, the variance of x_1 is 10 and the variance of x_3 is 32. The fact that 32 is more than 10 indicates that the set of third entries in the response vectors contains a wider spread of values than the set of first entries.

The **total variance** of the data is the sum of the variances on the diagonal of S. In general, the sum of the diagonal entries of a square matrix S is called the **trace** of the matrix, written $\text{tr}(S)$. Thus

$$\{\text{total variance}\} = \text{tr}(S)$$

The entry s_{ij} in S for $i \neq j$ is called the **covariance** of x_i and x_j. Observe that in Example 3, the covariance between x_1 and x_3 is 0 because the $(1, 3)$-entry in S is 0. Statisticians say that x_1 and x_3 are **uncorrelated**. Analysis of the multivariate data in $\mathbf{X}_1, \dots, \mathbf{X}_N$ is greatly simplified when most or all of the variables $x_1, \dots, x_p$ are uncorrelated, that is, when the covariance matrix of $\mathbf{X}_1, \dots, \mathbf{X}_N$ is diagonal or nearly diagonal.

Principal Component Analysis

For simplicity, assume that the matrix $[\,\mathbf{X}_1 \;\; \cdots \;\; \mathbf{X}_N\,]$ is already in mean-deviation form. The goal of principal component analysis is to find an orthogonal $p \times p$ matrix $P = [\,\mathbf{u}_1 \;\; \cdots \;\; \mathbf{u}_p\,]$ that determines a change of variable, $\mathbf{X} = P\mathbf{Y}$, or

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = [\,\mathbf{u}_1 \;\; \mathbf{u}_2 \;\; \cdots \;\; \mathbf{u}_p\,]\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix}$$

with the property that the new variables $y_1, \dots, y_p$ are uncorrelated and are arranged in order of decreasing variance.

The orthogonal change of variable $\mathbf{X} = P\mathbf{Y}$ means that each observation vector $\mathbf{X}_k$ receives a "new name," $\mathbf{Y}_k$, such that $\mathbf{X}_k = P\mathbf{Y}_k$. Notice that $\mathbf{Y}_k$ is the coordinate vector of $\mathbf{X}_k$ with respect to the columns of P, and $\mathbf{Y}_k = P^{-1}\mathbf{X}_k = P^T\mathbf{X}_k$ for $k = 1, \dots, N$.

It is not difficult to verify that for any orthogonal P, the covariance matrix of $\mathbf{Y}_1, \dots, \mathbf{Y}_N$ is P^TSP (Exercise 11). So the desired orthogonal matrix P is one that makes P^TSP diagonal. Let D be a diagonal matrix with the eigenvalues $\lambda_1, \dots, \lambda_p$ of S on the diagonal, arranged so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, and let P be an orthogonal matrix whose columns are the corresponding unit eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_p$. Then $S = PDP^T$ and $P^TSP = D$.

The unit eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_p$ of the covariance matrix S are called the **principal components** of the data (in the matrix of observations). The **first principal component** is the eigenvector corresponding to the largest eigenvalue of S, the **second principal component** is the eigenvector corresponding to the second largest eigenvalue, and so on.

The first principal component $\mathbf{u}_1$ determines the new variable y_1 in the following way. Let $c_1, \dots, c_p$ be the entries in $\mathbf{u}_1$. Since $\mathbf{u}_1^T$ is the first row of P^T, the equation $\mathbf{Y} = P^T\mathbf{X}$ shows that

$$y_1 = \mathbf{u}_1^T\mathbf{X} = c_1x_1 + c_2x_2 + \cdots + c_px_p$$

Thus y_1 is a linear combination of the original variables $x_1, \dots, x_p$, using the entries in the eigenvector $\mathbf{u}_1$ as weights. In a similar fashion, $\mathbf{u}_2$ determines the variable y_2, and so on.

EXAMPLE 4 The initial data for the multispectral image of Railroad Valley (Example 2) consisted of 4 million vectors in $\mathbb{R}^3$. The associated covariance matrix is[1]

$$S = \begin{bmatrix} 2382.78 & 2611.84 & 2136.20 \\ 2611.84 & 3106.47 & 2553.90 \\ 2136.20 & 2553.90 & 2650.71 \end{bmatrix}$$

Find the principal components of the data, and list the new variable determined by the first principal component.

SOLUTION The eigenvalues of S and the associated principal components (the unit eigenvectors) are

$$\lambda_1 = 7614.23 \qquad \lambda_2 = 427.63 \qquad \lambda_3 = 98.10$$

$$\mathbf{u}_1 = \begin{bmatrix} .5417 \\ .6295 \\ .5570 \end{bmatrix} \quad \mathbf{u}_2 = \begin{bmatrix} -.4894 \\ -.3026 \\ .8179 \end{bmatrix} \quad \mathbf{u}_3 = \begin{bmatrix} .6834 \\ -.7157 \\ .1441 \end{bmatrix}$$

Using two decimal places for simplicity, the variable for the first principal component is

$$y_1 = .54x_1 + .63x_2 + .56x_3$$

This equation was used to create photograph (d) in the chapter introduction. The variables x_1, x_2, and x_3 are the signal intensities in the three spectral bands. The values of x_1, converted to a gray scale between black and white, produced photograph (a). Similarly, the values of x_2 and x_3 produced photographs (b) and (c), respectively. At each pixel in photograph (d), the gray scale value is computed from y_1, a weighted linear combination of x_1, x_2, and x_3. In this sense, photograph (d) "displays" the first principal component of the data. ■

In Example 4, the covariance matrix for the transformed data, using variables y_1, y_2, and y_3, is

$$D = \begin{bmatrix} 7614.23 & 0 & 0 \\ 0 & 427.63 & 0 \\ 0 & 0 & 98.10 \end{bmatrix}$$

Although D is obviously simpler than the original covariance matrix S, the merit of constructing the new variables is not yet apparent. However, the variances of the variables y_1, y_2, and y_3 appear on the diagonal of D, and obviously the first variance in D is much larger than the other two. As we shall see, this fact will permit us to view the data as essentially one-dimensional rather than three-dimensional.

Reducing the Dimension of Multivariate Data

Principal component analysis is potentially valuable for applications in which most of the variation, or dynamic range, in the data is due to variations in *only a few* of the new variables, $y_1, \dots, y_p$.

It can be shown that an orthogonal change of variables, $\mathbf{X} = P\mathbf{Y}$, does not change the total variance of the data. (Roughly speaking, this is true because left-multiplication by P does not change the lengths of vectors or the angles between them. See Exercise 12.) This means that if $S = PDP^T$, then

$$\begin{Bmatrix} \text{total variance} \\ \text{of } x_1, \dots, x_p \end{Bmatrix} = \begin{Bmatrix} \text{total variance} \\ \text{of } y_1, \dots, y_p \end{Bmatrix} = \text{tr}(D) = \lambda_1 + \dots + \lambda_p$$

The variance of y_j is λ_j, and the quotient $\lambda_j / \text{tr}(S)$ measures the fraction of the total variance that is "explained" or "captured" by y_j.

[1] Data for Example 4 and Exercises 5 and 6 were provided by Earth Satellite Corporation, Rockville, Maryland.

EXAMPLE 5 Compute the various percentages of variance of the Railroad Valley multispectral data that are displayed in the principal component photographs, (d)–(f), shown in the chapter introduction.

SOLUTION The total variance of the data is

$$\text{tr}(D) = 7614.23 + 427.63 + 98.10 = 8139.96$$

[Verify that this number also equals tr(S).] The percentages of the total variance explained by the principal components are

First component	Second component	Third component
$\dfrac{7614.23}{8139.96} = 93.5\%$	$\dfrac{427.63}{8139.96} = 5.3\%$	$\dfrac{98.10}{8139.96} = 1.2\%$

In a sense, 93.5% of the information collected by Landsat for the Railroad Valley region is displayed in photograph (d), with 5.3% in (e) and only 1.2% remaining for (f). ■

The calculations in Example 5 show that the data have practically no variance in the third (new) coordinate. The values of y_3 are all close to zero. Geometrically, the data points lie nearly in the plane $y_3 = 0$, and their locations can be determined fairly accurately by knowing only the values of y_1 and y_2. In fact, y_2 also has relatively small variance, which means that the points lie approximately along a line, and the data are essentially one-dimensional. See Figure 2, in which the data resemble a popsicle stick.

Characterizations of Principal Component Variables

If $y_1, \dots, y_p$ arise from a principal component analysis of a $p \times N$ matrix of observations, then the variance of y_1 is as large as possible in the following sense: If $\mathbf{u}$ is any unit vector and if $y = \mathbf{u}^T\mathbf{X}$, then the variance of the values of y as $\mathbf{X}$ varies over the original data $\mathbf{X}_1, \dots, \mathbf{X}_N$ turns out to be $\mathbf{u}^T S\mathbf{u}$. By Theorem 8 in Section 7.3, the maximum value of $\mathbf{u}^T S\mathbf{u}$, over all unit vectors $\mathbf{u}$, is the largest eigenvalue λ_1 of S, and this variance is attained when $\mathbf{u}$ is the corresponding eigenvector $\mathbf{u}_1$. In the same way, Theorem 8 shows that y_2 has maximum possible variance among all variables $y = \mathbf{u}^T\mathbf{X}$ that are *uncorrelated* with y_1. Likewise, y_3 has maximum possible variance among all variables uncorrelated with both y_1 and y_2, and so on.

NUMERICAL NOTE

The singular value decomposition is the main tool for performing principal component analysis in practical applications. If B is a $p \times N$ matrix of observations in mean-deviation form, and if $A = (1/\sqrt{N-1})B^T$, then A^TA is the covariance matrix, S. The squares of the singular values of A are the p eigenvalues of S, and the right singular vectors of A are the principal components of the data.

As mentioned in Section 7.4, iterative calculation of the SVD of A is faster and more accurate than an eigenvalue decomposition of S. This is particularly true, for instance, in the hyperspectral image processing (with $p = 224$) mentioned in the chapter introduction. Principal component analysis is completed in seconds on specialized workstations.

Further Reading

Lillesand, Thomas M., and Ralph W. Kiefer, *Remote Sensing and Image Interpretation*, 4th ed. (New York: John Wiley, 2000).

PRACTICE PROBLEMS

The following table lists the weights and heights of five boys:

Boy	#1	#2	#3	#4	#5
Weight (lb)	120	125	125	135	145
Height (in.)	61	60	64	68	72

1. Find the covariance matrix for the data.
2. Make a principal component analysis of the data to find a single *size index* that explains most of the variation in the data.

7.5 EXERCISES

In Exercises 1 and 2, convert the matrix of observations to mean-deviation form, and construct the sample covariance matrix.

1. $\begin{bmatrix} 19 & 22 & 6 & 3 & 2 & 20 \\ 12 & 6 & 9 & 15 & 13 & 5 \end{bmatrix}$

2. $\begin{bmatrix} 1 & 5 & 2 & 6 & 7 & 3 \\ 3 & 11 & 6 & 8 & 15 & 11 \end{bmatrix}$

3. Find the principal components of the data for Exercise 1.

4. Find the principal components of the data for Exercise 2.

5. **[M]** A Landsat image with three spectral components was made of Homestead Air Force Base in Florida (after the base was hit by Hurricane Andrew in 1992). The covariance matrix of the data is shown below. Find the first principal component of the data, and compute the percentage of the total variance that is contained in this component.
$$S = \begin{bmatrix} 164.12 & 32.73 & 81.04 \\ 32.73 & 539.44 & 249.13 \\ 81.04 & 249.13 & 189.11 \end{bmatrix}$$

6. **[M]** The covariance matrix below was obtained from a Landsat image of the Columbia River in Washington, using data from three spectral bands. Let x_1, x_2, x_3 denote the spectral components of each pixel in the image. Find a new variable of the form $y_1 = c_1x_1 + c_2x_2 + c_3x_3$ that has maximum possible variance, subject to the constraint that $c_1^2 + c_2^2 + c_3^2 = 1$. What percentage of the total variance in the data is explained by y_1?
$$S = \begin{bmatrix} 29.64 & 18.38 & 5.00 \\ 18.38 & 20.82 & 14.06 \\ 5.00 & 14.06 & 29.21 \end{bmatrix}$$

7. Let x_1, x_2 denote the variables for the two-dimensional data in Exercise 1. Find a new variable y_1 of the form $y_1 = c_1x_1 + c_2x_2$, with $c_1^2 + c_2^2 = 1$, such that y_1 has maximum possible variance over the given data. How much of the variance in the data is explained by y_1?

8. Repeat Exercise 7 for the data in Exercise 2.

9. Suppose three tests are administered to a random sample of college students. Let $\mathbf{X}_1, \ldots, \mathbf{X}_N$ be observation vectors in $\mathbb{R}^3$ that list the three scores of each student, and for $j = 1, 2, 3$, let x_j denote a student's score on the jth exam. Suppose the covariance matrix of the data is
$$S = \begin{bmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 7 \end{bmatrix}$$
Let y be an "index" of student performance, with $y = c_1x_1 + c_2x_2 + c_3x_3$ and $c_1^2 + c_2^2 + c_3^2 = 1$. Choose c_1, c_2, c_3 so that the variance of y over the data set is as large as possible. [*Hint:* The eigenvalues of the sample covariance matrix are $\lambda = 3, 6$, and 9.]

10. **[M]** Repeat Exercise 9 with $S = \begin{bmatrix} 5 & 4 & 2 \\ 4 & 11 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

11. Given multivariate data $\mathbf{X}_1, \ldots, \mathbf{X}_N$ (in $\mathbb{R}^p$) in mean-deviation form, let P be a $p \times p$ matrix, and define $\mathbf{Y}_k = P^T\mathbf{X}_k$ for $k = 1, \ldots, N$.
    a. Show that $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ are in mean-deviation form. [*Hint:* Let $\mathbf{w}$ be the vector in $\mathbb{R}^N$ with a 1 in each entry. Then $[\,\mathbf{X}_1 \ \cdots \ \mathbf{X}_N\,]\,\mathbf{w} = \mathbf{0}$ (the zero vector in $\mathbb{R}^p$).]
    b. Show that if the covariance matrix of $\mathbf{X}_1, \ldots, \mathbf{X}_N$ is S, then the covariance matrix of $\mathbf{Y}_1, \ldots, \mathbf{Y}_N$ is P^TSP.

12. Let $\mathbf{X}$ denote a vector that varies over the columns of a $p \times N$ matrix of observations, and let P be a $p \times p$ orthogonal matrix. Show that the change of variable $\mathbf{X} = P\mathbf{Y}$ does not change the total variance of the data. [*Hint:* By Exercise 11, it suffices to show that $\operatorname{tr}(P^TSP) = \operatorname{tr}(S)$. Use a property of the trace mentioned in Exercise 25 in Section 5.4.]

13. The sample covariance matrix is a generalization of a formula for the variance of a sample of N scalar measurements, say, $t_1, \ldots, t_N$. If m is the average of $t_1, \ldots, t_N$, then the *sample variance* is given by
$$\frac{1}{N-1}\sum_{k=1}^{n}(t_k - m)^2 \tag{1}$$
Show how the sample covariance matrix, S, defined prior to Example 3, may be written in a form similar to (1). [*Hint:* Use partitioned matrix multiplication to write S as $1/(N-1)$ times the sum of N matrices of size $p \times p$. For $1 \le k \le N$, write $\mathbf{X}_k - \mathbf{M}$ in place of $\hat{\mathbf{X}}_k$.]

SOLUTIONS TO PRACTICE PROBLEMS

1. First arrange the data in mean-deviation form. The sample mean vector is easily seen to be $\mathbf{M} = \begin{bmatrix} 130 \\ 65 \end{bmatrix}$. Subtract $\mathbf{M}$ from the observation vectors (the columns in the table) and obtain
$$B = \begin{bmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix}$$
Then the sample covariance matrix is
$$S = \frac{1}{5-1}\begin{bmatrix} -10 & -5 & -5 & 5 & 15 \\ -4 & -5 & -1 & 3 & 7 \end{bmatrix}\begin{bmatrix} -10 & -4 \\ -5 & -5 \\ -5 & -1 \\ 5 & 3 \\ 15 & 7 \end{bmatrix}$$
$$= \frac{1}{4}\begin{bmatrix} 400 & 190 \\ 190 & 100 \end{bmatrix} = \begin{bmatrix} 100.0 & 47.5 \\ 47.5 & 25.0 \end{bmatrix}$$

2. The eigenvalues of S are (to two decimal places)
$$\lambda_1 = 123.02 \quad \text{and} \quad \lambda_2 = 1.98$$
The unit eigenvector corresponding to λ_1 is $\mathbf{u} = \begin{bmatrix} .900 \\ .436 \end{bmatrix}$. (Since S is 2×2, the computations can be done by hand if a matrix program is not available.) For the *size index*, set
$$y = .900\hat{w} + .436\hat{h}$$
where $\hat{w}$ and $\hat{h}$ are weight and height, respectively, in mean-deviation form. The variance of this index over the data set is 123.02. Because the total variance is $\operatorname{tr}(S) = 100 + 25 = 125$, the size index accounts for practically all (98.4%) of the variance of the data.

The original data for Practice Problem 1 and the line determined by the first principal component $\mathbf{u}$ are shown in Figure 4. (In parametric vector form, the line is $\mathbf{x} = \mathbf{M} + t\mathbf{u}$.) It can be shown that the line is the best approximation to the data,

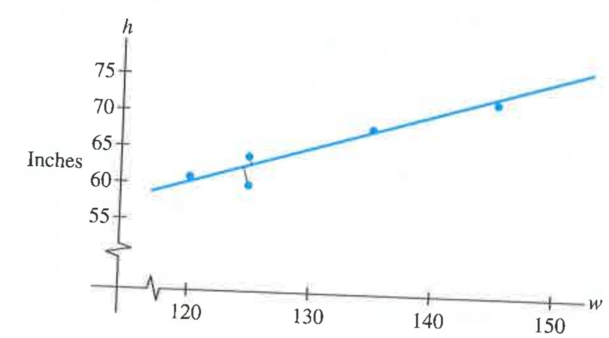


FIGURE 4 An orthogonal regression line determined by the first principal component of the data.