7.4 The Singular Value Decomposition

In Chapter 5, we saw that every symmetric matrix A can be factored as $A = PDP^T$, where P is an orthogonal matrix and D is a diagonal matrix displaying the eigenvalues of A. If A is not symmetric, such a factorization is not possible, but as we learned in Chapter 4, we may still be able to factor a square matrix A as $A = PDP^{-1}$, where D is as before but P is now simply an invertible matrix. However, not every matrix is diagonalizable, so it may surprise you that we will now show that *every* matrix (symmetric or not, square or not) has a factorization of the form $A = PDQ^T$, where P and Q are orthogonal and D is a diagonal matrix! This remarkable result is the *singular value decomposition* (SVD), and it is one of the most important of all matrix factorizations.

In this section, we will show how to compute the SVD of a matrix and then consider some of its many applications. Along the way, we will tie up some loose ends by answering a few questions that were left open in previous sections.

The Singular Values of a Matrix

For any $m \times n$ matrix A, the $n \times n$ matrix A^TA is symmetric and hence can be orthogonally diagonalized, by the Spectral Theorem. Not only are the eigenvalues of A^TA all real (Theorem 5.18), they are all *nonnegative*. To show this, let λ be an eigenvalue of A^TA with corresponding unit eigenvector $\mathbf{v}$. Then

$$\begin{aligned} 0 \le \|A\mathbf{v}\|^2 &= (A\mathbf{v})\cdot(A\mathbf{v}) = (A\mathbf{v})^TA\mathbf{v} = \mathbf{v}^TA^TA\mathbf{v} \\ &= \mathbf{v}^T\lambda\mathbf{v} = \lambda(\mathbf{v}\cdot\mathbf{v}) = \lambda\|\mathbf{v}\|^2 = \lambda \end{aligned}$$

It therefore makes sense to take (positive) square roots of these eigenvalues.

Definition If A is an $m \times n$ matrix, the ***singular values*** of A are the square roots of the eigenvalues of A^TA and are denoted by $\sigma_1, \ldots, \sigma_n$. It is conventional to arrange the singular values so that $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$.

Example 7.33 Find the singular values of

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

Solution The matrix

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

has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 1$. Consequently, the singular values of A are $\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}$ and $\sigma_2 = \sqrt{\lambda_2} = 1$.

To understand the significance of the singular values of an $m \times n$ matrix A, consider the eigenvectors of A^TA. Since A^TA is symmetric, we know that there is an *orthonormal* basis for $\mathbb{R}^n$ that consists of eigenvectors of A^TA. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be such a basis corresponding to the eigenvalues of A^TA, ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. From our calculations just before the definition,

$$\lambda_i = \|A\mathbf{v}_i\|^2$$

Therefore,

$$\sigma_i = \sqrt{\lambda_i} = \|A\mathbf{v}_i\|$$

In other words, the singular values of A are the lengths of the vectors $A\mathbf{v}_1, \ldots, A\mathbf{v}_n$.

Geometrically, this result has a nice interpretation. Consider Example 7.33 again. If $\mathbf{x}$ lies on the unit circle in $\mathbb{R}^2$ (i.e., $\|\mathbf{x}\| = 1$), then

$$\begin{aligned}\|A\mathbf{x}\|^2 &= (A\mathbf{x})\cdot(A\mathbf{x}) = (A\mathbf{x})^T(A\mathbf{x}) = \mathbf{x}^TA^TA\mathbf{x} \\ &= [x_1 \quad x_2]\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + 2x_1x_2 + 2x_2^2\end{aligned}$$

which we recognize is a quadratic form. By Theorem 5.25, the maximum and minimum values of this quadratic form, subject to the constraint $\|\mathbf{x}\| = 1$, are $\lambda_1 = 3$ and $\lambda_2 = 1$, respectively, and they occur at the corresponding eigenvectors of A^TA—that is, when $\mathbf{x} = \mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{x} = \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$, respectively. Since

$$\|A\mathbf{v}_i\|^2 = \mathbf{v}_i^TA^TA\mathbf{v}_i = \lambda_i$$

for $i = 1, 2$, we see that $\sigma_1 = \|A\mathbf{v}_1\| = \sqrt{3}$ and $\sigma_2 = \|A\mathbf{v}_2\| = 1$ are the maximum and minimum values of the lengths $\|A\mathbf{x}\|$ as $\mathbf{x}$ traverses the unit circle in $\mathbb{R}^2$.

Now, the linear transformation corresponding to A maps $\mathbb{R}^2$ onto the plane in $\mathbb{R}^3$ with equation $x - y - z = 0$ (verify this), and the image of the unit circle under this transformation is an ellipse that lies in this plane. (We will verify this fact in general shortly; see Figure 7.18.) So σ_1 and σ_2 are the lengths of half of the major and minor axes of this ellipse, as shown in Figure 7.19.

We can now describe the singular value decomposition of a matrix.

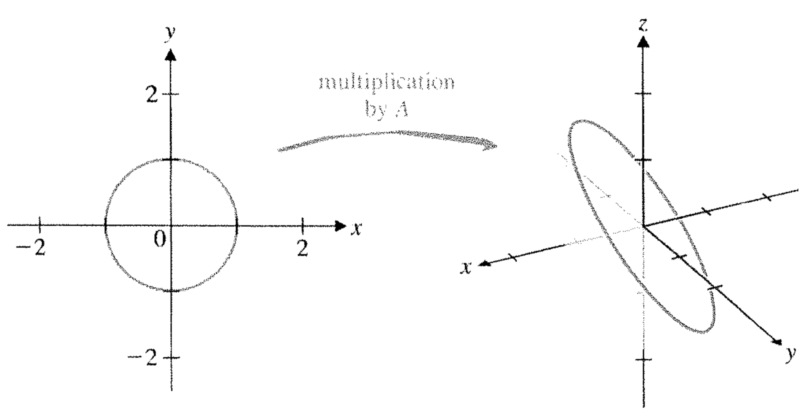

Figure 7.18
The matrix A transforms the unit circle in $\mathbb{R}^2$ into an ellipse in $\mathbb{R}^3$

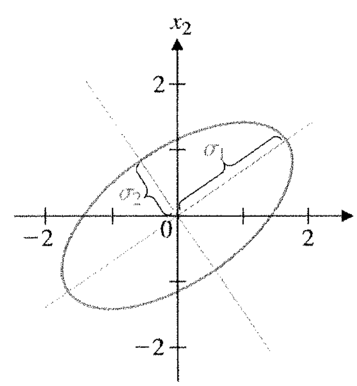

Figure 7.19

The Singular Value Decomposition

We want to show that an $m \times n$ matrix A can be factored as

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

where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and Σ is an $m \times n$ "diagonal" matrix. If the *nonzero* singular values of A are

$$\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$$

and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$, then Σ will have the block form

$$\Sigma = \begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \left[\begin{array}{c|c} D & O \\ \hline O & O \end{array}\right] \end{array} \begin{array}{l} \} \, r \\ \} \, m-r \end{array}, \quad \text{where} \quad D = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix} \tag{1}$$

and each matrix O is a zero matrix of the appropriate size. (If $r = m$ or $r = n$, some of these will not appear.) Some examples of such a matrix Σ with $r = 2$ are

$$\Sigma = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 8 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(What is D in each case?)

To construct the orthogonal matrix V, we first find an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors of the $n \times n$ symmetric matrix A^TA. Then

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

is an orthogonal $n \times n$ matrix.

For the orthogonal matrix U, we first note that $\{A\mathbf{v}_1, \ldots, A\mathbf{v}_n\}$ is an orthogonal set of vectors in $\mathbb{R}^m$. To see this, suppose that $\mathbf{v}_i$ is the eigenvector of A^TA corresponding to the eigenvalue λ_i. Then, for $i \ne j$, we have

$$\begin{aligned} (A\mathbf{v}_i) \cdot (A\mathbf{v}_j) &= (A\mathbf{v}_i)^T A\mathbf{v}_j \\ &= \mathbf{v}_i^T A^T A\mathbf{v}_j \\ &= \mathbf{v}_i^T \lambda_j \mathbf{v}_j \\ &= \lambda_j (\mathbf{v}_i \cdot \mathbf{v}_j) = 0 \end{aligned}$$

since the eigenvectors $\mathbf{v}_i$ are orthogonal. Now recall that the singular values satisfy $\sigma_i = \|A\mathbf{v}_i\|$ and that the first r of these are nonzero. Therefore, we can normalize $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ by setting

$$\mathbf{u}_i = \frac{1}{\sigma_i} A\mathbf{v}_i \quad \text{for } i = 1, \ldots, r$$

This guarantees that $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal set in $\mathbb{R}^m$, but if $r < m$ it will not be a basis for $\mathbb{R}^m$. In this case, we extend the set $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ for $\mathbb{R}^m$. (This is the only tricky part of the construction; we will describe techniques for carrying it out in the examples below and in the exercises.) Then we set

$$U = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m]$$

All that remains to be shown is that this works; that is, we need to verify that with U, V, and Σ as described, we have $A = U\Sigma V^T$. Since $V^T = V^{-1}$, this is equivalent to showing that

$$AV = U\Sigma$$

We know that $$A\mathbf{v}_i = \sigma_i\mathbf{u}_i \quad \text{for } i = 1, \dots, r$$

and $\|A\mathbf{v}_i\| = \sigma_i = 0$ for $i = r + 1, \dots, n$. Hence,

$$A\mathbf{v}_i = \mathbf{0} \quad \text{for } i = r + 1, \dots, n$$

Therefore,
$$\begin{aligned}
AV &= A[\mathbf{v}_1 \ \cdots \ \mathbf{v}_n] \\
&= [A\mathbf{v}_1 \ \cdots \ A\mathbf{v}_n] \\
&= [A\mathbf{v}_1 \ \cdots \ A\mathbf{v}_r \ \ \mathbf{0} \ \cdots \ \mathbf{0}] \\
&= [\sigma_1\mathbf{u}_1 \ \cdots \ \sigma_r\mathbf{u}_r \ \ \mathbf{0} \ \cdots \ \mathbf{0}] \\
&= [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right] \\
&= U\Sigma
\end{aligned}$$

as required.

We have just proved the following extremely important theorem.

Theorem 7.13 **The Singular Value Decomposition**

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Then there exist an $m \times m$ orthogonal matrix U, an $n \times n$ orthogonal matrix V, and an $m \times n$ matrix Σ of the form shown in Equation (1) such that

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

A factorization of A as in Theorem 7.13 is called a ***singular value decomposition (SVD)*** of A. The columns of U are called ***left singular vectors*** of A, and the columns of V are called ***right singular vectors*** of A. The matrices U and V are not uniquely determined by A, but Σ *must* contain the singular values of A, as in Equation (1). (See Exercise 25.)

Example 7.34

Find a singular value decomposition for the following matrices:

(a) $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

Solution (a) We compute

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

and find that its eigenvalues are $\lambda_1 = 2$, $\lambda_2 = 1$, and $\lambda_3 = 0$, with corresponding eigenvectors

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

(Verify this.) These vectors are orthogonal, so we normalize them to obtain

$$\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

The singular values of A are $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sqrt{1} = 1$, and $\sigma_3 = \sqrt{0} = 0$. Thus,

$$V = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

To find U, we compute

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

and

$$\mathbf{u}_2 = \frac{1}{\sigma_2}A\mathbf{v}_2 = \frac{1}{1}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

These vectors already form an orthonormal basis (the standard basis) for $\mathbb{R}^2$, so we have

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

This yields the SVD

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

which can be easily checked. (Note that V had to be transposed. Also note that the singular value σ_3 does not appear in Σ.)

(b) This is the matrix in Example 7.33, so we already know that the singular values are $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$, corresponding to $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. So

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

For U, we compute

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

and

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

This time, we need to extend $\{\mathbf{u}_1, \mathbf{u}_2\}$ to an orthonormal basis for $\mathbb{R}^3$. There are several ways to proceed; one method is to use the Gram-Schmidt Process, as in Example 5.14. We first need to find a linearly independent set of three vectors that contains $\mathbf{u}_1$ and $\mathbf{u}_2$. If $\mathbf{e}_3$ is the third standard basis vector in $\mathbb{R}^3$, it is clear that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{e}_3\}$ is linearly independent. (Here, you should be able to determine this by inspection, but a reliable method to use in general is to row reduce the matrix with these vectors as its columns and use the Fundamental Theorem.) Applying Gram-Schmidt (with normalization) to $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{e}_3\}$ (only the last step is needed), we find

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

so

$$U = \begin{bmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$$

and we have the SVD

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

There is another form of the singular value decomposition, analogous to the spectral decomposition of a symmetric matrix. It is obtained from the SVD by an outer product expansion and is very useful in applications. We can obtain this version of the SVD by imitating what we did to obtain the spectral decomposition.

Accordingly, we have

$$A = U\Sigma V^T = [\mathbf{u}_1 \ \cdots \ \mathbf{u}_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right]\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r \,|\, \mathbf{u}_{r+1} \ \cdots \ \mathbf{u}_m]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O \\ 0 & \cdots & \sigma_r & \\ \hline & O & & O \end{array}\right]\left[\begin{array}{c} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \\ \hline \mathbf{v}_{r+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array}\right]$$

$$= [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r]\begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} + [\mathbf{u}_{r+1} \ \cdots \ \mathbf{u}_m]\,[O]\begin{bmatrix} \mathbf{v}_{r+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

$$= [\mathbf{u}_1 \ \cdots \ \mathbf{u}_r]\begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}$$

$$= [\sigma_1\mathbf{u}_1 \ \cdots \ \sigma_r\mathbf{u}_r]\begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}$$

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

using block multiplication and the column-row representation of the product. The following theorem summarizes the process for obtaining this ***outer product form of the SVD.***

Theorem 7.14 **The Outer Product Form of the SVD**

Let A be an $m \times n$ matrix with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$. Let $\mathbf{u}_1, \ldots, \mathbf{u}_r$ be left singular vectors and let $\mathbf{v}_1, \ldots, \mathbf{v}_r$ be right singular vectors of A corresponding to these singular values. Then

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

Remark If A is a positive definite, symmetric matrix, then Theorems 7.13 and 7.14 both reduce to results that we already know. In this case, it is not hard to show that the SVD generalizes the Spectral Theorem and that Theorem 7.14 generalizes the spectral decomposition. (See Exercise 27.)

The SVD of a matrix A contains much important information about A, as outlined in the crucial Theorem 7.15.

Theorem 7.15 Let $A = U\Sigma V^T$ be a singular value decomposition of an $m \times n$ matrix A. Let $\sigma_1, \ldots, \sigma_r$ be all the nonzero singular values of A. Then:

a. The rank of A is r.
b. $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal basis for col(A).
c. $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis for null(A^T).
d. $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for row(A).
e. $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for null(A).

Proof (a) By Exercise 61 in Section 3.5, we have

$$\begin{aligned}\operatorname{rank}(A) &= \operatorname{rank}(U\Sigma V^T) \\ &= \operatorname{rank}(\Sigma V^T) \\ &= \operatorname{rank}(\Sigma) = r\end{aligned}$$

(b) We already know that $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal set. Therefore, it is linearly independent, by Theorem 5.1. Since $\mathbf{u}_i = (1/\sigma_i)A\mathbf{v}_i$ for $i = 1, \ldots, r$, each $\mathbf{u}_i$ is in the column space of A. (Why?) Furthermore,

$$r = \operatorname{rank}(A) = \dim(\operatorname{col}(A))$$

Therefore, $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is an orthonormal basis for col(A), by Theorem 6.10(c).

(c) Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ is an orthonormal basis for $\mathbb{R}^m$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ is a basis for col(A), by property (b), it follows that $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis for the orthogonal complement of col(A). But $(\operatorname{col}(A)) = \operatorname{null}(A^T)$, by Theorem 5.10.

(e) Since

$$A\mathbf{v}_{r+1} = \cdots = A\mathbf{v}_n = \mathbf{0}$$

the set $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal set contained in the null space of A. Therefore, $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is a linearly independent set of $n - r$ vectors in null(A). But

$$\dim(\operatorname{null}(A)) = n - r$$

by the Rank Theorem, so $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for null(A), by Theorem 6.10(c).

(d) Property (d) follows from property (e) and Theorem 5.10. (You are asked to prove this in Exercise 32.)

The SVD provides new geometric insight into the effect of matrix transformations. We have noted several times (without proof) that an $m \times n$ matrix transforms the unit sphere in $\mathbb{R}^n$ into an ellipsoid in $\mathbb{R}^m$. This point arose, for example, in our discussions of Perron's Theorem and of operator norms, as well as in the introduction to singular values in this section. We now prove this result.

Theorem 7.16 Let A be an $m \times n$ matrix with rank r. Then the image of the unit sphere in $\mathbb{R}^n$ under the matrix transformation that maps $\mathbf{x}$ to $A\mathbf{x}$ is

a. the surface of an ellipsoid in $\mathbb{R}^m$ if $r = n$.
b. a solid ellipsoid in $\mathbb{R}^m$ if $r < n$.

Proof Let $A = U\Sigma V^T$ be a singular value decomposition of the $m \times n$ matrix A. Let the left and right singular vectors of A be $\mathbf{u}_1, \dots, \mathbf{u}_m$ and $\mathbf{v}_1, \dots, \mathbf{v}_n$, respectively. Since $\operatorname{rank}(A) = r$, the singular values of A satisfy

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \quad \text{and} \quad \sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$$

by Theorem 7.15(a). Let $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be a unit vector in $\mathbb{R}^n$. Now, since V is an orthogonal matrix, so is V^T, and hence $V^T\mathbf{x}$ is a unit vector, by Theorem 5.6. Now

$$V^T\mathbf{x} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{v}_1^T\mathbf{x} \\ \vdots \\ \mathbf{v}_n^T\mathbf{x} \end{bmatrix}$$

so $(\mathbf{v}_1^T\mathbf{x})^2 + \cdots + (\mathbf{v}_n^T\mathbf{x})^2 = 1$.

By the outer product form of the SVD, we have $A = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T$. Therefore,

$$\begin{aligned} A\mathbf{x} &= \sigma_1\mathbf{u}_1\mathbf{v}_1^T\mathbf{x} + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T\mathbf{x} \\ &= (\sigma_1\mathbf{v}_1^T\mathbf{x})\mathbf{u}_1 + \cdots + (\sigma_r\mathbf{v}_r^T\mathbf{x})\mathbf{u}_r \\ &= y_1\mathbf{u}_1 + \cdots + y_r\mathbf{u}_r \end{aligned}$$

where we are letting y_i denote the scalar $\sigma_i\mathbf{v}_i^T\mathbf{x}$.

(a) If $r = n$, then we must have $n \leq m$ and

$$\begin{aligned} A\mathbf{x} &= y_1\mathbf{u}_1 + \cdots + y_n\mathbf{u}_n \\ &= Uy \end{aligned}$$

where $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Therefore, again by Theorem 5.6, $\|A\mathbf{x}\| = \|U\mathbf{y}\| = \|\mathbf{y}\|$, since U is orthogonal. But

$$\left(\frac{y_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{y_n}{\sigma_n}\right)^2 = (\mathbf{v}_1^T\mathbf{x})^2 + \cdots + (\mathbf{v}_n^T\mathbf{x})^2 = 1$$

which shows that the vectors $A\mathbf{x}$ form the surface of an ellipsoid in $\mathbb{R}^m$. (Why?)

(b) If $r < n$, the only difference in the above steps is that the equation becomes

$$\left(\frac{y_1}{\sigma_1}\right)^2 + \cdots + \left(\frac{y_r}{\sigma_r}\right)^2 \leq 1$$

since we are missing some terms. This inequality corresponds to a solid ellipsoid in $\mathbb{R}^m$.

Example 7.35

Describe the image of the unit sphere in $\mathbb{R}^3$ under the action of the matrix

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

Solution In Example 7.34(a), we found the following SVD of A:

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

Since $r = \text{rank}(A) = 2 < 3 = n$, the second part of Theorem 7.16 applies. The image of the unit sphere will satisfy the inequality

$$\left(\frac{y_1}{\sqrt{2}}\right)^2 + \left(\frac{y_2}{1}\right)^2 \leq 1 \quad \text{or} \quad \frac{y_1^2}{2} + y_2^2 \leq 1$$

relative to $y_1 y_2$ coordinate axes in $\mathbb{R}^2$ (corresponding to the left singular vectors $\mathbf{u}_1$ and $\mathbf{u}_2$). Since $\mathbf{u}_1 = \mathbf{e}_1$ and $\mathbf{u}_2 = \mathbf{e}_2$, the image is as shown in Figure 7.20.

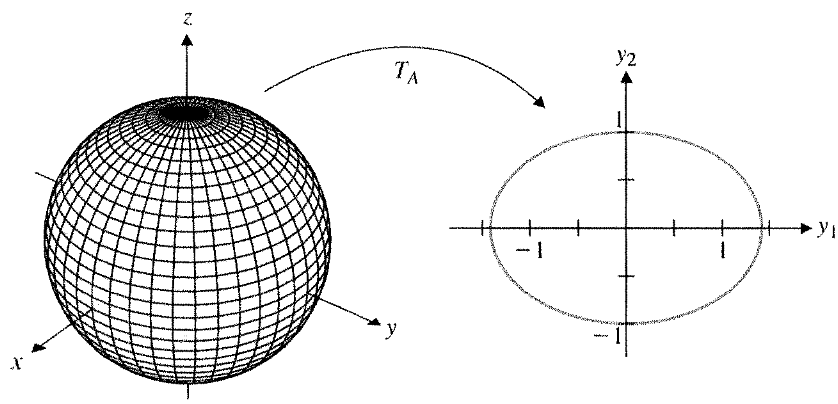

Figure 7.20

In general, we can describe the effect of an $m \times n$ matrix A on the unit sphere in $\mathbb{R}^n$ in terms of the effect of each factor in its SVD, $A = U\Sigma V^T$, from right to left. Since V^T is an orthogonal matrix, it maps the unit sphere to itself. The $m \times n$ matrix Σ does two things: The diagonal entries $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$ collapse $n - r$ of the dimensions of the unit sphere, leaving an r-dimensional unit sphere, which the nonzero diagonal entries $\sigma_1, \ldots, \sigma_r$ then distort into an ellipsoid. The orthogonal matrix U then aligns the axes of this ellipsoid with the orthonormal basis vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ in $\mathbb{R}^m$. (See Figure 7.21.)

Applications of the SVD

The singular value decomposition is an extremely useful tool, both practically and theoretically. We will look at just a few of its many applications.

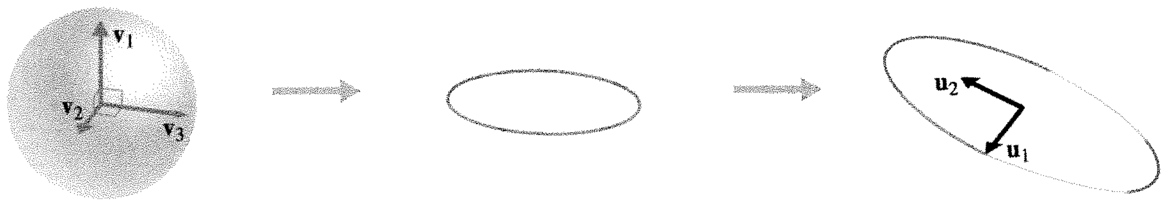

Figure 7.21

Rank Until now, we have not worried about calculating the rank of a matrix from a computational point of view. We compute the rank of a matrix by row reducing it to echelon form and counting the number of nonzero rows. However, as we have seen, roundoff errors can affect this process, especially if the matrix is ill-conditioned. Entries that should be zero may end up as very small nonzero numbers, affecting our ability to accurately determine the rank and other quantities associated with the matrix. In practice, the SVD is often used to find the rank of a matrix, since it is much more reliable when roundoff errors are present. The basic idea behind this approach is that the orthogonal matrices U and V in the SVD preserve lengths and thus do not introduce additional errors; any errors that occur will tend to show up in the matrix Σ.

CAS **Example 7.36** Let

$$A = \begin{bmatrix} 8.1650 & -0.0041 & -0.0041 \\ 4.0825 & -3.9960 & 4.0042 \\ 4.0825 & 4.0042 & -3.9960 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 8.17 & 0 & 0 \\ 4.08 & -4 & 4 \\ 4.08 & 4 & -4 \end{bmatrix}$$

The matrix B has been obtained by rounding off the entries in A to two decimal places. If we compute the ranks of these two approximately equal matrices, we find that $\operatorname{rank}(A) = 3$ but $\operatorname{rank}(B) = 2$. By the Fundamental Theorem, this implies, among other things, that A is invertible but B is not.

The explanation for this critical difference between two matrices that are approximately equal lies in their SVDs. The singular values of A are 10, 8, and 0.01, so A has rank 3. The singular values of B are 10, 8, and 0, so B has rank 2.

In practical applications, it is often assumed that if a singular value is computed to be close to zero, then roundoff error has crept in and the actual value should be zero. In this way, "noise" can be filtered out. In this example, if we compute $A = U\Sigma V^T$ and replace

$$\Sigma = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 0.01 \end{bmatrix} \quad \text{by} \quad \Sigma' = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

then $U\Sigma' V^T = B$. (Try it!)

Matrix Norms and the Condition Number The SVD can provide simple formulas for certain expressions involving matrix norms. Consider, for example, the Frobenius norm of a matrix. The following theorem shows that it is completely determined by the singular values of the matrix.

equal to the nonzero singular values of A. It follows immediately that the largest of these is σ_1, so

$$\|A\|_2 = \sigma_1$$

This provides us with a neat way to express the condition number of a (square) matrix with respect to the operator 2-norm. Recall that the condition number (with respect to the operator 2-norm) of an invertible matrix A is defined as

$$\text{cond}_2(A) = \|A^{-1}\|_2\|A\|_2$$

As you will be asked to show in Exercise 28, if $A = U\Sigma V^T$, then $A^{-1} = V\Sigma^{-1}U^T$. Therefore, the singular values of A^{-1} are $1/\sigma_1, \ldots, 1/\sigma_n$ (why?), and

$$1/\sigma_n \geq \cdots \geq 1/\sigma_1$$

It follows that $\|A^{-1}\|_2 = 1/\sigma_n$, so

$$\text{cond}_2(A) = \frac{\sigma_1}{\sigma_n}$$

Example 7.38 Find the 2-condition number of the matrix A in Example 7.36.

Solution Since $\sigma_1 = 10$ and $\sigma_3 = 0.01$,

$$\text{cond}_2(A) = \frac{\sigma_1}{\sigma_3} = \frac{10}{0.01} = 1000$$

This value is large enough to suggest that A may be ill-conditioned and we should be wary of the effect of roundoff errors.

The Pseudoinverse and Least Squares Approximation In Section 7.3, we produced the formula $A^+ = (A^TA)^{-1}A^T$ for the pseudoinverse of a matrix A. Clearly, this formula is valid only if A^TA is invertible, as we noted at the time. Equipped with the SVD, we can now define the pseudoinverse of *any* matrix, generalizing our previous formula.

E. H. Moore (1862–1932) was an American mathematician who worked in group theory, number theory, and geometry. He was the first head of the mathematics department at the University of Chicago when it opened in 1892. In 1920, he introduced a generalized matrix inverse that included rectangular matrices. His work did not receive much attention because of his obscure writing style.

Definition Let $A = U\Sigma V^T$ be an SVD for an $m \times n$ matrix A, where $\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$ and D is an $r \times r$ diagonal matrix containing the nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ of A. The ***pseudoinverse*** (or ***Moore-Penrose inverse***) of A is the $n \times m$ matrix A^+ defined by

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

where Σ^+ is the $n \times m$ matrix

$$\Sigma^+ = \begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}$$

Example 7.39 Find the pseudoinverses of the matrices in Example 7.34.

Solution (a) From the SVD

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

we form

$$\Sigma^+ = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Then

$$A^+ = V\Sigma^+U^T = \begin{bmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \\ 0 & 1 \end{bmatrix}$$

(b) We have the SVD

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

so

$$\Sigma^+ = \begin{bmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

and

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

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

It is straightforward to check that this new definition of the pseudoinverse generalizes the old one, for if the $m \times n$ matrix $A = U\Sigma V^T$ has linearly independent columns, then direct substitution shows that $(A^TA)^{-1}A^T = V\Sigma^+U^T$. (You are asked to verify this in Exercise 50.) Other properties of the pseudoinverse are explored in the exercises.

We have seen that when A has linearly independent columns, there is a unique least squares solution $\bar{\mathbf{x}}$ to $A\mathbf{x} = \mathbf{b}$; that is, the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$ have the unique solution

$$\bar{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = A^+\mathbf{b}$$

When the columns of A are linearly dependent, then A^TA is not invertible, so the normal equations have infinitely many solutions. In this case, we will ask for the solution $\bar{\mathbf{x}}$ of *minimum length* (i.e., the one closest to the origin). It turns out that this time we simply use the general version of the pseudoinverse.

Jerry Bauer

One of those who was unaware of Moore's work on matrix inverses was Roger Penrose (b.1931), who introduced his own notion of a generalized matrix inverse in 1955. Penrose has made many contributions to geometry and theoretical physics. He is also the inventor of a type of *nonperiodic tiling* that covers the plane with only two different shapes of tile, yet has no repeating pattern. He has received many awards, including the 1988 Wolf Prize in Physics, which he shared with Stephen Hawking. In 1994, he was knighted for services to science. Sir Roger Penrose is currently the Emeritus Rouse Ball Professor of Mathematics at the University of Oxford.

Theorem 7.18 The least squares problem $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution $\overline{\mathbf{x}}$ of minimal length that is given by

$$\overline{\mathbf{x}} = A^{+}\mathbf{b}$$

Proof Let A be an $m \times n$ matrix of rank r with SVD $A = U\Sigma V^T$ (so that $A^+ = V\Sigma^+U^T$). Let $\mathbf{y} = V^T\mathbf{x}$ and let $\mathbf{c} = U^T\mathbf{b}$. Write $\mathbf{y}$ and $\mathbf{c}$ in block form as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix}$$

where $\mathbf{y}_1$ and $\mathbf{c}_1$ are in $\mathbb{R}^r$.

We wish to minimize $\|\mathbf{b} - A\mathbf{x}\|$ or, equivalently, $\|\mathbf{b} - A\mathbf{x}\|^2$. Using Theorem 5.6 and the fact that U^T is orthogonal (because U is), we have

$$\begin{aligned}\|\mathbf{b} - A\mathbf{x}\|^2 &= \|U^T(\mathbf{b} - A\mathbf{x})\|^2 = \|U^T(\mathbf{b} - U\Sigma V^T\mathbf{x})\|^2 = \|U^T\mathbf{b} - U^TU\Sigma V^T\mathbf{x}\|^2 \\ &= \|\mathbf{c} - \Sigma\mathbf{y}\|^2 = \left\| \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} - \begin{bmatrix} D & O \\ O & O \end{bmatrix}\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \mathbf{c}_1 - D\mathbf{y}_1 \\ \mathbf{c}_2 \end{bmatrix} \right\|^2\end{aligned}$$

The only part of this expression that we have any control over is $\mathbf{y}_1$, so the minimum value occurs when $\mathbf{c}_1 - D\mathbf{y}_1 = \mathbf{0}$ or, equivalently, when $\mathbf{y}_1 = D^{-1}\mathbf{c}_1$. So all least squares solutions $\mathbf{x}$ are of the form

$$\mathbf{x} = V\mathbf{y} = V\begin{bmatrix} D^{-1}\mathbf{c}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

Set

$$\overline{\mathbf{x}} = V\overline{\mathbf{y}} = V\begin{bmatrix} D^{-1}\mathbf{c}_1 \\ \mathbf{0} \end{bmatrix}$$

We claim that this $\overline{\mathbf{x}}$ is the least squares solution of minimal length. To show this, let's suppose that

$$\mathbf{x}' = V\mathbf{y}' = V\begin{bmatrix} D^{-1}\mathbf{c}_1 \\ \mathbf{y}_2 \end{bmatrix}$$

is a different least squares solution (hence, $\mathbf{y}_2 \neq \mathbf{0}$). Then

$$\|\overline{\mathbf{x}}\| = \|V\overline{\mathbf{y}}\| = \|\overline{\mathbf{y}}\| < \|\mathbf{y}'\| = \|V\mathbf{y}'\| = \|\mathbf{x}'\|$$

as claimed.

We still must show that $\overline{\mathbf{x}}$ is equal to $A^+\mathbf{b}$. To do so, we simply compute

$$\begin{aligned}\overline{\mathbf{x}} = V\overline{\mathbf{y}} &= V\begin{bmatrix} D^{-1}\mathbf{c}_1 \\ \mathbf{0} \end{bmatrix} = V\begin{bmatrix} D^{-1} & O \\ O & O \end{bmatrix}\begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{bmatrix} \\ &= V\Sigma^+\mathbf{c} = V\Sigma^+U^T\mathbf{b} = A^+\mathbf{b}\end{aligned}$$

Example 7.40 Find the minimum length least squares solution of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Solution The corresponding equations

$$x + y = 0$$
$$x + y = 1$$

are clearly inconsistent, so a least squares solution is our only hope. Moreover, the columns of A are linearly dependent, so there will be infinitely many least squares solutions—among which we want the one with minimal length.

An SVD of A is given by

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

(Verify this.) It follows that

$$A^+ = V\Sigma^+U^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

so
$$\bar{\mathbf{x}} = A^+\mathbf{b} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

You can see that the minimum least squares solution in Example 7.40 satisfies $x + y = \frac{1}{2}$. In a sense, this is a compromise between the two equations we started with. In Exercise 49, you are asked to solve the normal equations for this problem directly and to verify that this solution really is the one closest to the origin.

The Fundamental Theorem of Invertible Matrices It is appropriate to conclude by revisiting the Fundamental Theorem of Invertible Matrices one more time. Not surprisingly, the singular values of a square matrix tell us when the matrix is invertible.

Theorem 7.19 **The Fundamental Theorem of Invertible Matrices: Final Version**

Let A be an $n \times n$ matrix and let $T : V \to W$ be a linear transformation whose matrix $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ with respect to bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively, is A. The following statements are equivalent:

a. A is invertible.
b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b}$ in $\mathbb{R}^n$.
c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
d. The reduced row echelon form of A is I_n.
e. A is a product of elementary matrices.
f. $\operatorname{rank}(A) = n$
g. $\operatorname{nullity}(A) = 0$
h. The column vectors of A are linearly independent.
i. The column vectors of A span $\mathbb{R}^n$.
j. The column vectors of A form a basis for $\mathbb{R}^n$.
k. The row vectors of A are linearly independent.
l. The row vectors of A span $\mathbb{R}^n$.