REDUCED SINGULAR VALUE DECOMPOSITION

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

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

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

Algebraically, the zero rows and columns of the matrix Σ in Formula (12) are superfluous and can be eliminated by multiplying out the expression $U\Sigma V^T$ using block multiplication and the partitioning shown in that formula.

The products that involve zero blocks as factors drop out, leaving

$$A = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix} \tag{13}$$

which is called a ***reduced singular value decomposition*** of A. In this text we will denote the matrices on the right side of (13) by U_1, Σ_1, and V_1^T, respectively, and we will write this equation as

$$A = U_1 \Sigma_1 V_1^T \tag{14}$$

Note that the sizes of U_1, Σ_1, and V_1^T are $m \times k$, $k \times k$, and $k \times n$, respectively, and that the matrix Σ_1 is invertible, since its diagonal entries are positive.

It is now a square matrix.

If we multiply out on the right side of (13) using the column-row rule of Theorem 3.8.1, then we obtain

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \tag{15}$$

which is called a ***reduced singular value expansion*** of A. This result applies to *all* matrices, whereas the spectral decomposition [Formula (5) of Section 8.3] applies only to symmetric matrices.

You should also compare (15) to the column-row expansion of a general matrix A given in Theorem 7.6.5. In the singular value expansion the $\mathbf{u}$'s and $\mathbf{v}$'s are orthonormal, whereas the $\mathbf{c}$'s and $\mathbf{r}$'s in Theorem 7.6.5 need not be so.

Theorem 7.6.5 (***Column-Row Expansion***) *If A is a nonzero matrix of rank k, then A can be expressed as*

$$A = \mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_k\mathbf{r}_k \tag{4}$$

where $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k$ are the successive pivot columns of A and $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$ are the successive nonzero row vectors in the reduced row echelon form of A.

EXAMPLE 5 Reduced Singular Value Decomposition

Find a reduced singular value decomposition and a reduced singular value expansion of the matrix

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

Solution In Example 4 we found the singular value decomposition

$$\underset{A}{\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}} = \underset{U}{\begin{bmatrix} \frac{\sqrt{6}}{3} & 0 & -\frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} & \frac{1}{\sqrt{3}} \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}} \tag{16}$$

Since A has rank 2 (verify), it follows from (13) with $k = 2$ that the reduced singular value decomposition of A corresponding to (16) is

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

This yields the reduced singular value expansion

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

$$= \sqrt{3} \begin{bmatrix} \frac{\sqrt{3}}{3} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} \end{bmatrix} + (1) \begin{bmatrix} 0 & 0 \\ -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Note that the matrices in the expansion have rank 1, as expected. ■

DATA COMPRESSION AND IMAGE PROCESSING

Singular value decompositions can be used to "compress" visual information for the purpose of reducing its required storage space and speeding up its electronic transmission. The first step in compressing a visual image is to represent it as a numerical matrix from which the visual image can be recovered when needed.

For example, a black and white photograph might be scanned as a rectangular array of pixels (points) and then stored as a matrix A by assigning each pixel a numerical value in accordance with its gray level. If 256 different gray levels are used ($0 =$ white to $255 =$ black), then the entries in the matrix would be integers between 0 and 255. The image can be recovered from the matrix A by printing or displaying the pixels with their assigned gray levels.

If the matrix A has size $m \times n$, then one might store each of its mn entries individually. An alternative procedure is to compute the reduced singular value decomposition

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \tag{17}$$

in which $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, and store the σ's, the $\mathbf{u}$'s, and the $\mathbf{v}$'s. When needed, the matrix A (and hence the image it represents) can be reconstructed from (17). Since each $\mathbf{u}_j$ has m entries and each $\mathbf{v}_j$ has n entries, this method requires storage space for

$$km + kn + k = k(m + n + 1)$$

numbers. Suppose, however, that the singular values $\sigma_{r+1}, \ldots, \sigma_k$ are sufficiently small that dropping the corresponding terms in (17) produces an acceptable approximation

$$A_r = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \tag{18}$$

to A and the image that it represents. We call (18) the ***rank r approximation of A***.

This matrix requires storage space for only

$$rm + rn + r = r(m + n + 1)$$

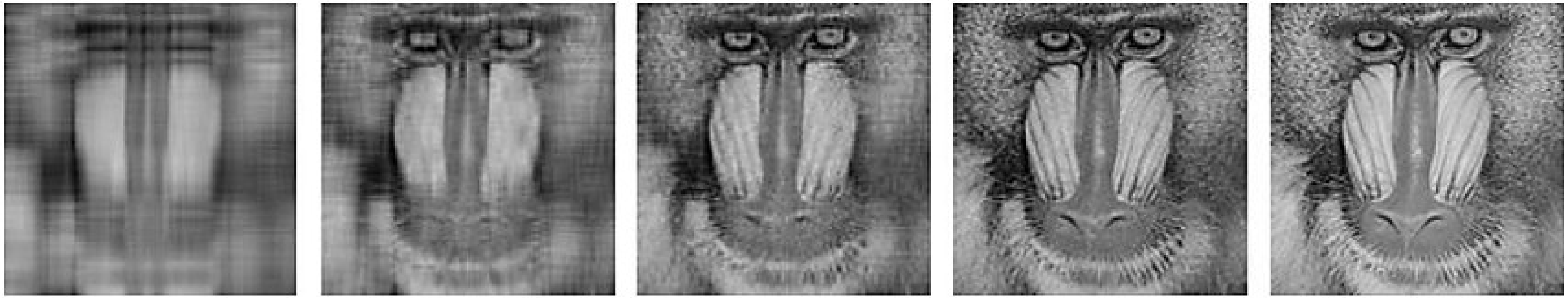
numbers, compared to mn numbers required for entry-by-entry storage of A.

For example, the rank 100 approximation of a 1000×1000 matrix A requires storage for only

$$100(1000 + 1000 + 1) = 200,100$$

numbers, compared to the 1,000,000 numbers required for entry-by-entry storage of A—a compression of almost 80%.

Figure 8.6.3 shows some approximations of a digitized mandrill image obtained using (18).



Rank 4 Rank 10 Rank 20 Rank 50 Rank 128

Figure 8.6.3

REMARK It can be proved that A_r has rank r, that A_r does not depend on the basis vectors used in Formula (18), and that A_r is the best possible approximation to A by $m \times n$ matrices of rank r in the sense that the sum of the squares of the differences between the entries of A and A_r is as small as possible.

SINGULAR VALUE DECOMPOSITION FROM THE TRANSFORMATION POINT OF VIEW

If A is an $m \times n$ matrix and $T_A : R^n \to R^m$ is multiplication by A, then the matrix

$$\Sigma = \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & 0_{k\times(n-k)} \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & & 0_{(m-k)\times k} & & 0_{(m-k)\times(n-k)} \end{array}\right]$$

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

in (12) is the matrix for T_A with respect to the bases $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ for R^n and R^m, respectively (verify).

Thus, when vectors are expressed in terms of these bases, we see that the effect of multiplying a vector by A is

to scale the first k coordinates of the vector by the factors $\sigma_1, \sigma_2, \ldots, \sigma_k$,

map the rest of the coordinates to zero,

and possibly to discard coordinates or append zeros, if needed, to account for a decrease or increase in dimension.

This idea is illustrated in Figure 8.6.4 for a 2×3 matrix A of rank 2.

The effect of multiplication by A on the unit sphere in R^3 is to collapse the three dimensions of the domain into the two dimensions of the range and then stretch or compress components in the directions of the left singular vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ in accordance with the magnitudes of the factors σ_1 and σ_2 to produce an ellipse in R^2.

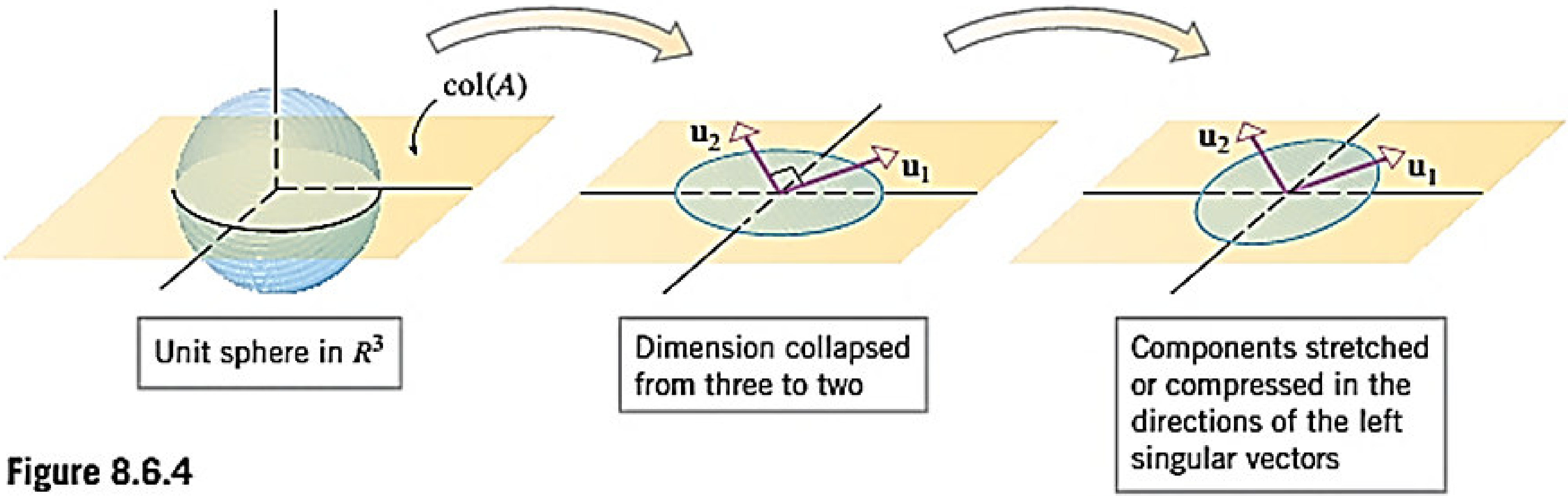


Figure 8.6.4

Some further insight into the singular value decomposition and reduced singular value decomposition of a matrix A can be obtained by focusing on the algebraic properties of the linear transformation $T_A(\mathbf{x}) = A\mathbf{x}$.

Since $\text{row}(A)^{\perp} = \text{null}(A)$, it follows from Theorem 7.7.4 that every vector $\mathbf{x}$ in R^n can be expressed uniquely as

$$\mathbf{x} = \mathbf{x}_{\text{row}(A)} + \mathbf{x}_{\text{null}(A)}$$

where $\mathbf{x}_{\text{row}(A)}$ is the orthogonal projection of $\mathbf{x}$ on the row space of A and $\mathbf{x}_{\text{null}(A)}$ is its orthogonal projection on the null space of A.

Since $A\mathbf{x}_{\text{null}(A)} = \mathbf{0}$, it follows that

$$T_A(\mathbf{x}) = A\mathbf{x} = A\mathbf{x}_{\text{row}(A)} + A\mathbf{x}_{\text{null}(A)} = A\mathbf{x}_{\text{row}(A)}$$

Theorem 7.7.4 (*Projection Theorem for Subspaces*) *If W is a subspace of R^n, then every vector $\mathbf{x}$ in R^n can be expressed in exactly one way as*

$$\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 \tag{20}$$

where $\mathbf{x}_1$ is in W and $\mathbf{x}_2$ is in $W^{\perp}$.

$$T_A(\mathbf{x}) = A\mathbf{x} = A\mathbf{x}_{\text{row}(A)} + A\mathbf{x}_{\text{null}(A)} = A\mathbf{x}_{\text{row}(A)}$$

This tells us three things:

1. The image of any vector in R^n under multiplication by A is the same as the image of the orthogonal projection of that vector on row(A).
2. The range of the transformation T_A, namely col(A), is the image of row(A).
3. T_A maps distinct vectors in row(A) into distinct vectors in R^m (why?). Thus, even though T_A may not be one-to-one when considered as a transformation with domain R^n, it is one-to-one if its domain is restricted to row(A).

If $A.\mathbf{l}_i = A.\mathbf{l}_j$, $A(\mathbf{l}_i - \mathbf{l}_j) = \mathbf{0}$. Then, $\mathbf{l}_i - \mathbf{l}_j$ would be at row(A) and at null(A) = row(A)$^{inv(T)}$. So, $\mathbf{l}_1 - \mathbf{l}_j$ would be $\mathbf{0}$ (theorem 7.3.4), which is not the case.

Since the behavior of a matrix transformation T_A is completely determined by its action on row(A), it makes sense, in the interest of efficiency, to eliminate the superfluous part of the domain and consider T_A as a transformation with domain row(A).

The matrix for this restricted transformation with respect to the bases $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ for row(A) and $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ for col(A) is the matrix

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_k \end{bmatrix}$$

that occurs in the reduced singular value decomposition of A.

REMARK Loosely phrased, the preceding discussion tells us that

"hiding" inside of every nonzero matrix transformation T_A there is a one-to-one matrix transformation that maps the row space of A onto the column space of A.

Moreover, that hidden transformation is represented by the reduced singular value decomposition of A with respect to appropriate bases.

Section 8.7 The Pseudoinverse

THE PSEUDOINVERSE

If A is an invertible $n \times n$ matrix with reduced singular value decomposition

$$A = U_1 \Sigma_1 V_1^T$$

then U_1, Σ_1, and V_1 are all $n \times n$ invertible matrices (why?), so the orthogonality of U_1 and V_1 implies that

$$A^{-1} = V_1 \Sigma_1^{-1} U_1^T \tag{1}$$

If A is not square or if it is square but not invertible, then this formula does not apply.

$$A^{-1} = V_1 \Sigma_1^{-1} U_1^T \tag{1}$$

However, we noted earlier that the matrix Σ_1 is always invertible, so the product on the right side of (1) is defined for every matrix A, though it is only for invertible A that it represents A^{-1}.

If A is a nonzero $m \times n$ matrix, then we call the $n \times m$ matrix

$$A^{+} = V_1 \Sigma_1^{-1} U_1^T \tag{2}$$

the ***pseudoinverse**** of A. If $A = 0$, then we define $A^{+} = 0$.

Also called the Moore-Penrose inverse.

The pseudoinverse is the same as the ordinary inverse for invertible matrices, but it is more general in that it applies to *all* matrices.

EXAMPLE 1

Finding the Pseudoinverse from the Reduced SVD

Find the pseudoinverse of the matrix

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

using the reduced singular value decomposition that was obtained in Example 5 of Section 8.6.

Solution In Example 5 of Section 8.6 we obtained the reduced singular value decomposition

$$A = [\mathbf{u}_1 \quad \mathbf{u}_2] \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{6}}{3} & 0 \\ \frac{\sqrt{6}}{6} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

Thus, it follows from (2) that

$$A^{+} = [\mathbf{v}_1 \quad \mathbf{v}_2] \begin{bmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{6}}{3} & \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} \\ 0 & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} \qquad \blacksquare$$

Theorem 8.7.1 *If A is an $m \times n$ matrix with full column rank, then*

$$A^{+} = (A^TA)^{-1}A^T \tag{3}$$

Proof Let $A = U_1\Sigma_1V_1^T$ be a reduced singular value decomposition of A. Then

$$A^TA = (V_1\Sigma_1^TU_1^T)(U_1\Sigma_1V_1^T) = V_1\Sigma_1^2V_1^T$$

Since A has full column rank, the matrix A^TA is invertible (Theorem 7.5.10) and V_1 is an $n \times n$ orthogonal matrix. Thus,

$$(A^TA)^{-1} = V_1\Sigma_1^{-2}V_1^T$$

from which it follows that

$$(A^TA)^{-1}A^T = (V_1\Sigma_1^{-2}V_1^T)(U_1\Sigma_1V_1^T)^T = (V_1\Sigma_1^{-2}V_1^T)(V_1\Sigma_1U_1^T) = V_1\Sigma_1^{-1}U_1^T = A^{+} \qquad \blacksquare$$

Theorem 7.5.10 *If A is an $m \times n$ matrix, then the following statements are equivalent.*

(*a*) $A\mathbf{x} = \mathbf{0}$ *has only the trivial solution.*

(*b*) $A\mathbf{x} = \mathbf{b}$ *has at most one solution for every* $\mathbf{b}$ *in* R^m.

(*c*) *A has full column rank.*

(*d*) *A^TA is invertible.*

EXAMPLE 2 Pseudoinverse in the Case of Full Column Rank

We computed the pseudoinverse of

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

in Example 1 using singular value decomposition. However, A has full column rank so its pseuodoinverse can also be computed from Formula (3). To do this we first compute

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

from which it follows that

$$A^+ = (A^T A)^{-1} A^T = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix}$$

This agrees with the result obtained in Example 1. ■

PROPERTIES OF THE PSEUDOINVERSE

The following theorem states some algebraic facts about the pseudoinverse, the proofs of which are left as exercises.

Theorem 8.7.2 *If A^+ is the pseudoinverse of an $m \times n$ matrix A, then:*

(a) $AA^+A = A$

(b) $A^+AA^+ = A^+$

(c) $(AA^+)^T = AA^+$

(d) $(A^+A)^T = A^+A$

(e) $(A^T)^+ = (A^+)^T$

(f) $A^{++} = A$

Formula (2) should be used in the proofs.

$$A^+ = V_1 \Sigma_1^{-1} U_1^T \qquad (2)$$

The next theorem states some properties of the pseudoinverse from the transformation point of view. We will prove the first three parts, and leave the last two as exercises.

Theorem 8.7.3 *If $A^+ = V_1\Sigma_1^{-1}U_1^T$ is the pseudoinverse of an $m \times n$ matrix A of rank k, and if the column vectors of U_1 and V_1 are $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, respectively, then:*

(a) $A^+\mathbf{y}$ *is in* $\text{row}(A)$ *for every vector* $\mathbf{y}$ *in* R^m.

(b) $A^+\mathbf{u}_i = \dfrac{1}{\sigma_i}\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

(c) $A^+\mathbf{y} = \mathbf{0}$ *for every vector* $\mathbf{y}$ *in* $\text{null}(A^T)$.

(d) AA^+ *is the orthogonal projection of* R^m *onto* $\text{col}(A)$.

(e) A^+A *is the orthogonal projection of* R^n *onto* $\text{row}(A)$.

(*a*) $A^{+}\mathbf{y}$ *is in* row(A) *for every vector* $\mathbf{y}$ *in* R^m.

Proof **(*a*)** If $\mathbf{y}$ is a vector in R^m, then it follows from (2) that

$$A^{+}\mathbf{y} = V_1\Sigma_1^{-1}U_1^T\mathbf{y} = V_1(\Sigma_1^{-1}U_1^T\mathbf{y})$$

so $A^{+}\mathbf{y}$ must be a linear combination of the column vectors of V_1. Since Theorem 8.6.5 states that these vectors are in row(A), it follows that $A^{+}\mathbf{y}$ is in row(A).

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

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

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

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

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

(b) $A^{+}\mathbf{u}_i = \dfrac{1}{\sigma_i}\mathbf{v}_i \qquad (i = 1, 2, \ldots, k)$

Proof (b) Multiplying A^{+} on the right by U_1 yields

$$A^{+}U_1 = V_1\Sigma_1^{-1}U_1^{T}U_1 = V_1\Sigma_1^{-1}$$

The result now follows by comparing corresponding column vectors on the two sides of this equation.

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

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

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

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

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

(c) $A^+\mathbf{y} = \mathbf{0}$ *for every vector* $\mathbf{y}$ *in* $\text{null}(A^T)$.

Proof (c) If $\mathbf{y}$ is a vector in $\text{null}(A^T)$, then $\mathbf{y}$ is orthogonal to each vector in $\text{col}(A)$, and, in particular, it is orthogonal to each column vector of $U_1 = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_k]$. This implies that $U_1^T\mathbf{y} = \mathbf{0}$ (why?), and hence that

$$A^+\mathbf{y} = V_1\Sigma_1^{-1}U_1^T\mathbf{y} = (V_1\Sigma_1^{-1})U_1^T\mathbf{y} = \mathbf{0}$$ ■

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

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

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

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

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

EXAMPLE 3 Orthogonal Projection Using the Pseudoinverse

Use the pseudoinverse of

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

to find the standard matrix for the orthogonal projection of R^3 onto the column space of A.

Solution The pseudoinverse of A was computed in Example 2. Using that result we see that the orthogonal projection of R^3 onto col(A) is

$$AA^+ = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{bmatrix} \qquad \blacksquare$$

PSEUDOINVERSE AND LEAST SQUARES

The pseudoinverse is important because it provides a way of using singular value decompositions to solve least squares problems.

Recall that the least squares solutions of a linear system $A\mathbf{x} = \mathbf{b}$ are the exact solutions of the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$. In the case where A has full column rank the matrix A^TA is invertible and there is a unique least squares solution

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b} = A^+\mathbf{b} \tag{4}$$

Thus, in the case of full column rank the least squares solution can be obtained by multiplying $\mathbf{b}$ by the pseudoinverse of A.

In the case where A does not have full column rank the matrix A^TA is not invertible and there are infinitely many solutions of the normal equation, each of which is a least squares solution of $A\mathbf{x} = \mathbf{b}$.

However, we know that among these least squares solutions there is a unique least squares solution in the row space of A (Theorem 7.8.3), and we also know that it is the least squares solution of minimum norm. The following theorem generalizes (4).

Theorem 8.7.4 *If A is an $m \times n$ matrix, and $\mathbf{b}$ is any vector in R^m, then*

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

is the least squares solution of $A\mathbf{x} = \mathbf{b}$ that has minimum norm.

Proof We will show first that $\mathbf{x} = A^{+}\mathbf{b}$ satisfies the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$ and hence is a least squares solution. For this purpose, let $A = U_1\Sigma_1V_1^T$ be a reduced singular value decomposition of A, so

$$A^{+}\mathbf{b} = V_1\Sigma_1^{-1}U_1^T\mathbf{b}$$

Thus,

$$(A^TA)A^{+}\mathbf{b} = V_1\Sigma_1^2V_1^TV_1\Sigma_1^{-1}U_1^T\mathbf{b} = V_1\Sigma_1^2\Sigma_1^{-1}U_1^T\mathbf{b} = V_1\Sigma_1U_1^T\mathbf{b} = A^T\mathbf{b}$$

which shows that $\mathbf{x} = A^{+}\mathbf{b}$ satisfies the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$.

To show that $\mathbf{x} = A^{+}\mathbf{b}$ is the least squares solution of minimum norm, it suffices to show that this vector lies in the row space of A (Theorem 7.8.3). But we know this to be true by part (a) of Theorem 8.7.3. ■

Some of the ideas we have been discussing are illustrated by the Strang diagram in Figure 8.7.1. The linear system $A\mathbf{x} = \mathbf{b}$ represented in that diagram is inconsistent, since $\mathbf{b}$ is not in $\text{col}(A)$. We have split $\mathbf{x}$ and $\mathbf{b}$ into orthogonal terms as

$$\mathbf{x} = \mathbf{x}_{\text{row}(A)} + \mathbf{x}_{\text{null}(A)} \quad \text{and} \quad \mathbf{b} = \mathbf{b}_{\text{col}(A)} + \mathbf{b}_{\text{null}(A^T)}$$

and have denoted $\mathbf{x}_{\text{row}(A)}$ by $\mathbf{x}^+$ for brevity. This vector is the least squares solution of minimum norm and is an exact solution of the equation $A\mathbf{x} = \mathbf{b}_{\text{col}(A)}$; that is,

$$A\mathbf{x}^+ = \mathbf{b}_{\text{col}(A)}$$

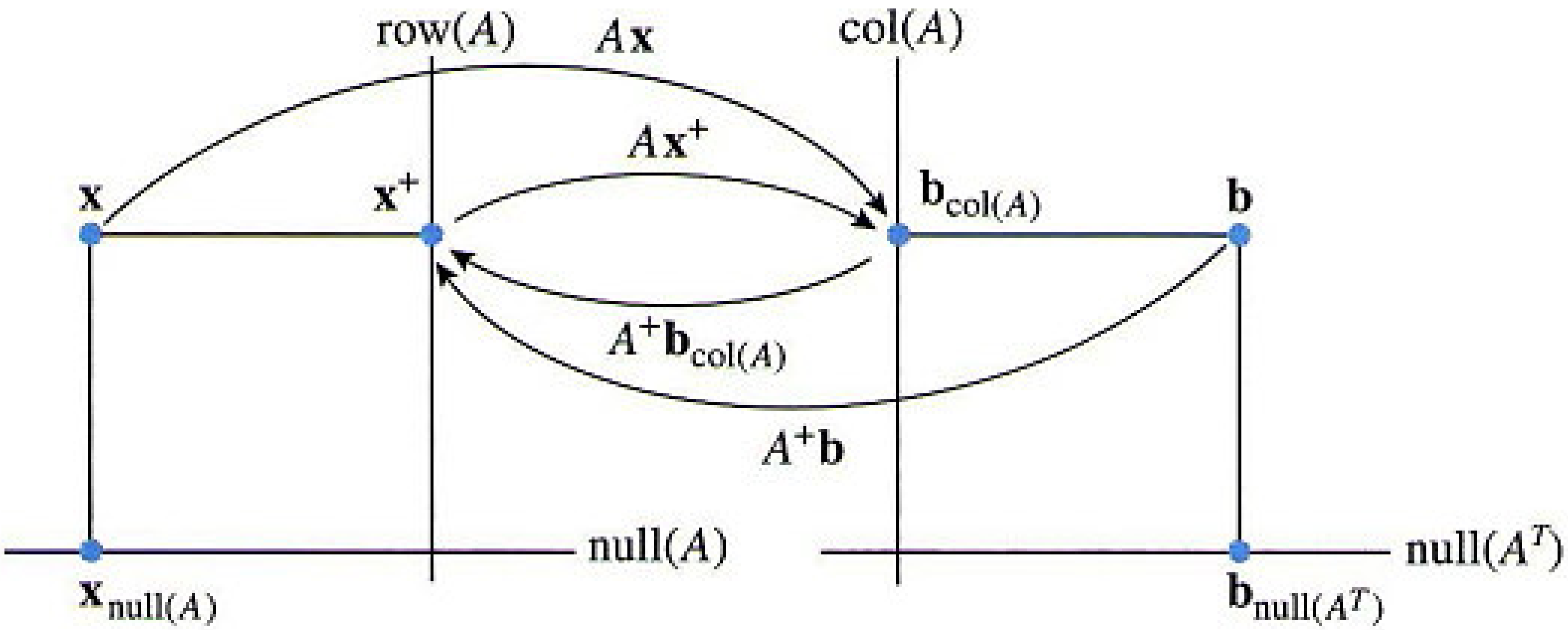


Figure 8.7.1

To solve this equation for $\mathbf{x}^+$, we can first multiply through by the pseudoinverse A^+ to obtain

$$A^+A\mathbf{x}^+ = A^+\mathbf{b}_{\mathrm{col}(A)}$$

and then use Theorem 8.7.3(e) and the fact that $\mathbf{x}^+ = \mathbf{x}_{\mathrm{row}(A)}$ is in the row space of A to obtain

$$\mathbf{x}^+ = A^+\mathbf{b}_{\mathrm{col}(A)}$$

Thus, A maps $\mathbf{x}^+$ into $\mathbf{b}_{\mathrm{col}(A)}$ and A^+ recovers $\mathbf{x}^+$ from $\mathbf{b}_{\mathrm{col}(A)}$.

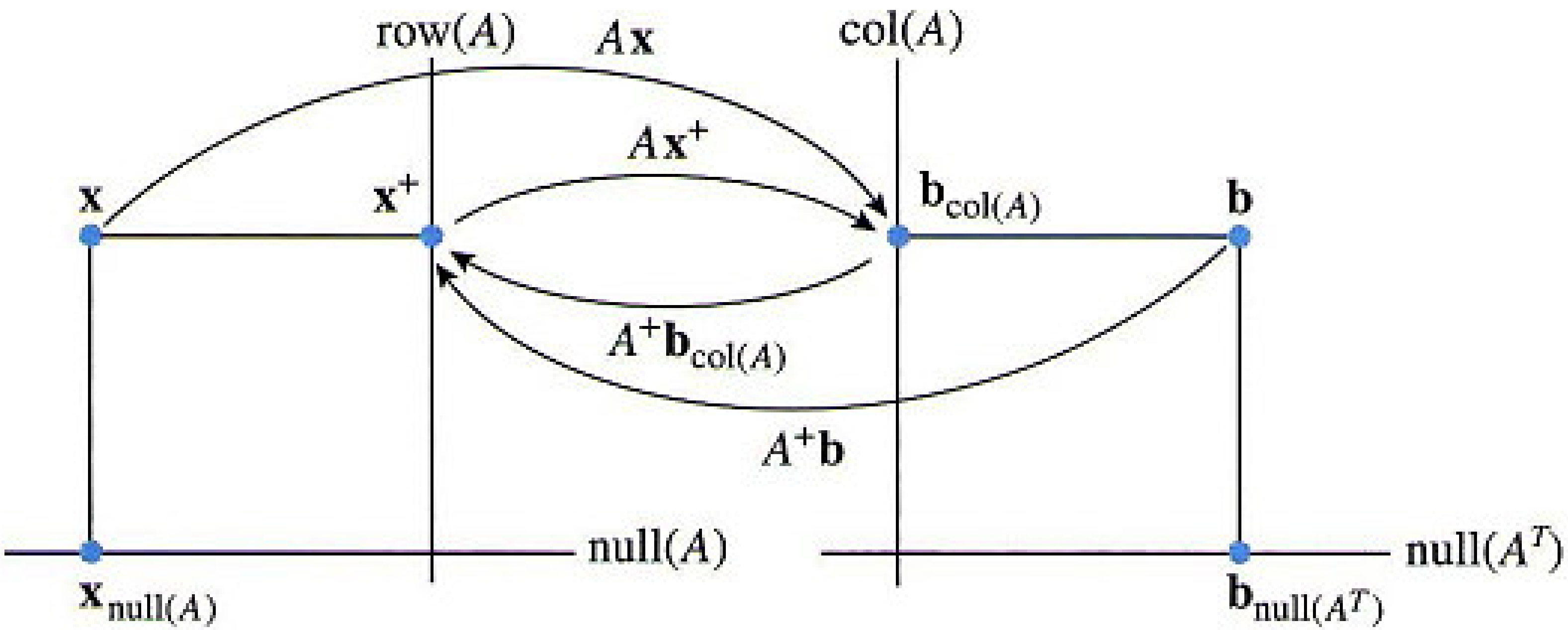


Figure 8.7.1

CONDITION NUMBER AND NUMERICAL CONSIDERATIONS

Singular value decomposition plays an important role in the analysis and solution of linear systems that are difficult to solve accurately because of their sensitivity to roundoff error.

In the case of a consistent linear system $A\mathbf{x} = \mathbf{b}$ this typically occurs when the coefficient matrix is "nearly singular" in the sense that one or more of its singular values is close to zero. Such linear systems are said to be ***ill conditioned***.

A good measure of how roundoff error will affect the accuracy of a computed solution is given by the ratio of the largest singular value of A to the smallest singular values of A. This ratio, called the ***condition number*** of A, is denoted by

$$\operatorname{cond}(A) = \frac{\sigma_1}{\sigma_k} \tag{5}$$

The larger the condition number, the more sensitive the system to small roundoff errors.

The basic method for finding least squares solutions of a linear system $A\mathbf{x} = \mathbf{b}$ is to solve the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$ exactly. However, the singular values of A^TA are the squares of the singular values of A (Exercise 21 of Section 8.6), so $\mathrm{cond}(A^TA)$ is the square of the condition number of A. Thus, if $A\mathbf{x} = \mathbf{b}$ is ill conditioned, then the normal equations are even worse!

In theory, one could determine the condition number of A by finding the singular value decomposition and then use that decomposition to compute the pseudoinverse and the least squares solution $\mathbf{x} = A^+\mathbf{b}$ if the system is not ill conditioned.

While all of this sounds reasonable, the difficulty is that the singular values of A are the square roots of the eigenvalues of A^TA, and calculating those singular values directly from the problematical A^TA may produce an inaccurate estimate of the condition number as well as an inaccurate least squares solution.

Fortunately, there are methods for finding singular value decompositions that do not involve computing with A^TA. These produce some of the best algorithms known for finding least squares solutions of linear systems and are discussed in books on numerical methods of linear algebra.

Two standard books on the subject are *Matrix Computations*, by G. H. Golub and C. F. Van Loan, Johns Hopkins University Press, Baltimore, 1996; and *Numerical Recipes in C, The Art of Scientific Computing*, by William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery, Cambridge University Press, New York, 1999.