Exercises for 8.5

Exercise 8.5.1 In each case, find the exact eigenvalues and determine corresponding eigenvectors. Then start with $\mathbf{x}_0 = \left[\begin{array}{c} 1 \\ 1 \end{array}\right]$ and compute $\mathbf{x}_4$ and r_3 using the power method.

a. $A = \left[\begin{array}{rr} 2 & -4 \\ -3 & 3 \end{array}\right]$

b. $A = \left[\begin{array}{rr} 5 & 2 \\ -3 & -2 \end{array}\right]$

c. $A = \left[\begin{array}{rr} 1 & 2 \\ 2 & 1 \end{array}\right]$

d. $A = \left[\begin{array}{rr} 3 & 1 \\ 1 & 0 \end{array}\right]$

Exercise 8.5.2 In each case, find the exact eigenvalues and then approximate them using the QR-algorithm.

a. $A = \left[\begin{array}{rr} 1 & 1 \\ 1 & 0 \end{array}\right]$

b. $A = \left[\begin{array}{rr} 3 & 1 \\ 1 & 0 \end{array}\right]$

Exercise 8.5.3 Apply the power method to $A = \left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array}\right]$, starting at $\mathbf{x}_0 = \left[\begin{array}{c} 1 \\ 1 \end{array}\right]$. Does it converge? Explain.

Exercise 8.5.4 If A is symmetric, show that each matrix A_k in the QR-algorithm is also symmetric. Deduce that they converge to a diagonal matrix.

Exercise 8.5.5 Apply the QR-algorithm to $A = \left[\begin{array}{rr} 2 & -3 \\ 1 & -2 \end{array}\right]$. Explain.

Exercise 8.5.6 Given a matrix A, let A_k, Q_k, and R_k, $k \geq 1$, be the matrices constructed in the QR-algorithm. Show that $A_k = (Q_1Q_2\cdots Q_k)(R_k\cdots R_2R_1)$ for each $k \geq 1$ and hence that this is a QR-factorization of A_k.
[*Hint*: Show that $Q_kR_k = R_{k-1}Q_{k-1}$ for each $k \geq 2$, and use this equality to compute $(Q_1Q_2\cdots Q_k)(R_k\cdots R_2R_1)$ "from the centre out." Use the fact that $(AB)^{n+1} = A(BA)^nB$ for any square matrices A and B.]

8.6 The Singular Value Decomposition

When working with a square matrix A it is clearly useful to be able to "diagonalize" A, that is to find a factorization $A = Q^{-1}DQ$ where Q is invertible and D is diagonal. Unfortunately such a factorization may not exist for A. However, even if A is not square gaussian elimination provides a factorization of the form $A = PDQ$ where P and Q are invertible and D is diagonal—the Smith Normal form (Theorem 2.5.3). However, if A is real we can choose P and Q to be *orthogonal* real matrices and D to be real. Such a factorization is called a **singular value decomposition (SVD)** for A, one of the most useful tools in applied linear algebra. In this Section we show how to explicitly compute an SVD for any real matrix A, and illustrate some of its many applications.

We need a fact about two subspaces associated with an $m \times n$ matrix A:

$$\text{im } A = \{A\mathbf{x} \mid \mathbf{x} \text{ in } \mathbb{R}^n\} \qquad \text{and} \qquad \text{col } A = \text{span }\{\mathbf{a} \mid \mathbf{a} \text{ is a column of } A\}$$

Then im A is called the **image** of A (so named because of the linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ with $\mathbf{x} \mapsto A\mathbf{x}$); and col A is called the **column space** of A (Definition 5.10). Surprisingly, these spaces are equal:

Lemma 8.6.1

For any $m \times n$ matrix A, im A = col A.

Proof. Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ in terms of its columns. Let $\mathbf{x} \in \text{im}\,A$, say $\mathbf{x} = A\mathbf{y}$, $\mathbf{y}$ in $\mathbb{R}^n$. If $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$, then $A\mathbf{y} = y_1\mathbf{a}_1 + y_2\mathbf{a}_2 + \cdots + y_n\mathbf{a}_n \in \text{col}\,A$ by Definition 2.5. This shows that $\text{im}\,A \subseteq \text{col}\,A$. For the other inclusion, each $\mathbf{a}_k = A\mathbf{e}_k$ where $\mathbf{e}_k$ is column k of I_n. □

8.6.1. Singular Value Decompositions

We know a lot about any real symmetric matrix: Its eigenvalues are real (Theorem 5.5.7), and it is orthogonally diagonalizable by the Principal Axes Theorem (Theorem 8.2.2). So for any real matrix A (square or not), the fact that both A^TA and AA^T are real and symmetric suggests that we can learn a lot about A by studying them. This section shows just how true this is.

The following Lemma reveals some similarities between A^TA and AA^T which simplify the statement and the proof of the SVD we are constructing.

Lemma 8.6.2

Let A be a real $m \times n$ matrix. Then:

1. *The eigenvalues of A^TA and AA^T are real and non-negative.*
2. *A^TA and AA^T have the same set of positive eigenvalues.*

Proof.

1. Let λ be an eigenvalue of A^TA, with eigenvector $\mathbf{0} \neq \mathbf{q} \in \mathbb{R}^n$. Then:

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

   Then (1.) follows for A^TA, and the case AA^T follows by replacing A by A^T.

2. Write $N(B)$ for the set of positive eigenvalues of a matrix B. We must show that $N(A^TA) = N(AA^T)$. If $\lambda \in N(A^TA)$ with eigenvector $\mathbf{0} \neq \mathbf{q} \in \mathbb{R}^n$, then $A\mathbf{q} \in \mathbb{R}^m$ and

$$AA^T(A\mathbf{q}) = A[(A^TA)\mathbf{q}] = A(\lambda\mathbf{q}) = \lambda(A\mathbf{q})$$

   Moreover, $A\mathbf{q} \neq \mathbf{0}$ since $A^TA\mathbf{q} = \lambda\mathbf{q} \neq \mathbf{0}$ and both $\lambda \neq 0$ and $\mathbf{q} \neq \mathbf{0}$. Hence λ is an eigenvalue of AA^T, proving $N(A^TA) \subseteq N(AA^T)$. For the other inclusion replace A by A^T.

□

To analyze an $m \times n$ matrix A we have two symmetric matrices to work with: A^TA and AA^T. In view of Lemma 8.6.2, we choose A^TA (sometimes called the **Gram** matrix of A), and derive a series of facts which we will need. This narrative is a bit long, but trust that it will be worth the effort. We parse it out in several steps:

1. The $n \times n$ matrix A^TA is real and symmetric so, by the Principal Axes Theorem 8.2.2, let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\} \subseteq \mathbb{R}^n$ be an orthonormal basis of eigenvectors of A^TA, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. By Lemma 8.6.2(1), λ_i is real for each i and $\lambda_i \geq 0$. By re-ordering the $\mathbf{q}_i$ we may (and do) assume that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0 \quad \text{and}^{8} \quad \lambda_i = 0 \text{ if } i > r \tag{i}$$

By Theorems 8.2.1 and 3.3.4, the matrix

$$Q = \left[\begin{array}{cccc} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{array} \right] \text{ is orthogonal} \quad \text{and} \quad \text{orthogonally diagonalizes } A^T A \tag{ii}$$

2. Even though the λ_i are the eigenvalues of $A^T A$, the number r in (**i**) turns out to be $\operatorname{rank} A$. To understand why, consider the vectors $A\mathbf{q}_i \in \operatorname{im} A$. For all i, j:

$$A\mathbf{q}_i \cdot A\mathbf{q}_j = (A\mathbf{q}_i)^T A\mathbf{q}_j = \mathbf{q}_i^T (A^T A)\mathbf{q}_j = \mathbf{q}_i^T (\lambda_j \mathbf{q}_j) = \lambda_j (\mathbf{q}_i^T \mathbf{q}_j) = \lambda_j (\mathbf{q}_i \cdot \mathbf{q}_j)$$

Because $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n\}$ is an orthonormal set, this gives

$$A\mathbf{q}_i \cdot A\mathbf{q}_j = 0 \text{ if } i \neq j \qquad \text{and} \qquad \|A\mathbf{q}_i\|^2 = \lambda_i \|\mathbf{q}_i\|^2 = \lambda_i \text{ for each } i \tag{iii}$$

We can extract two conclusions from (**iii**) and (**i**):

$$\{A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_r\} \subseteq \operatorname{im} A \text{ is an orthogonal set} \quad \text{and} \quad A\mathbf{q}_i = \mathbf{0} \text{ if } i > r \tag{iv}$$

With this write $U = \operatorname{span}\{A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_r\} \subseteq \operatorname{im} A$; we claim that $U = \operatorname{im} A$, that is $\operatorname{im} A \subseteq U$. For this we must show that $A\mathbf{x} \in U$ for each $\mathbf{x} \in \mathbb{R}^n$. Since $\{\mathbf{q}_1, \dots, \mathbf{q}_r, \dots, \mathbf{q}_n\}$ is a basis of $\mathbb{R}^n$ (it is orthonormal), we can write $\mathbf{x}_k = t_1\mathbf{q}_1 + \cdots + t_r\mathbf{q}_r + \cdots + t_n\mathbf{q}_n$ where each $t_j \in \mathbb{R}$. Then, using (**iv**) we obtain

$$A\mathbf{x} = t_1 A\mathbf{q}_1 + \cdots + t_r A\mathbf{q}_r + \cdots + t_n A\mathbf{q}_n = t_1 A\mathbf{q}_1 + \cdots + t_r A\mathbf{q}_r \in U$$

This shows that $U = \operatorname{im} A$, and so

$$\{A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_r\} \text{ is an } \textit{orthogonal} \text{ basis of } \operatorname{im}(A) \tag{v}$$

But $\operatorname{col} A = \operatorname{im} A$ by Lemma 8.6.1, and $\operatorname{rank} A = \dim(\operatorname{col} A)$ by Theorem 5.4.1, so

$$\operatorname{rank} A = \dim(\operatorname{col} A) = \dim(\operatorname{im} A) \overset{(\mathbf{v})}{=} r \tag{vi}$$

3. Before proceeding, some definitions are in order:

Definition 8.7

The real numbers $\sigma_i = \sqrt{\lambda_i} \overset{(\mathbf{iii})}{=} \|A\bar{\mathbf{q}}_i\|$ *for* $i = 1, 2, \dots, n$, *are called the* ***singular values*** *of the matrix* A.

Clearly $\sigma_1, \sigma_2, \dots, \sigma_r$ are the *positive* singular values of A. By (**i**) we have

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \qquad \text{and} \qquad \sigma_i = 0 \text{ if } i > r \tag{vii}$$

With (**vi**) this makes the following definitions depend only upon A.

[8] Of course they could *all* be positive ($r = n$) or *all* zero (so $A^T A = 0$, and hence $A = 0$ by Exercise 5.3.9).

Definition 8.8

Let A be a real, $m \times n$ matrix of rank r, with positive singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_i = 0$ if $i > r$. Define:

$$D_A = \operatorname{diag}(\sigma_1,\ \dots,\ \sigma_r) \qquad \text{and} \qquad \Sigma_A = \left[\begin{array}{cc} D_A & 0 \\ 0 & 0 \end{array}\right]_{m\times n}$$

*Here Σ_A is in block form and is called the **singular matrix** of A.*

The singular values σ_i and the matrices D_A and Σ_A will be referred to frequently below.

4. Returning to our narrative, normalize the vectors $A\mathbf{q}_1, A\mathbf{q}_2, \dots, A\mathbf{q}_r$, by defining

$$\mathbf{p}_i = \frac{1}{\|A\mathbf{q}_i\|}A\mathbf{q}_i \in \mathbb{R}^m \quad \text{for each } i = 1,\ 2,\ \dots,\ r \tag{viii}$$

By (**v**) and Lemma 8.6.1, we conclude that

$$\{\mathbf{p}_1,\ \mathbf{p}_2,\ \dots,\ \mathbf{p}_r\} \text{ is an \textit{orthonormal} basis of } \operatorname{col} A \subseteq \mathbb{R}^m \tag{ix}$$

Employing the Gram-Schmidt algorithm (or otherwise), construct $\mathbf{p}_{r+1},\ \dots,\ \mathbf{p}_m$ so that

$$\{\mathbf{p}_1,\ \dots,\ \mathbf{p}_r,\ \dots,\ \mathbf{p}_m\} \text{ is an orthonormal basis of } \mathbb{R}^m \tag{x}$$

5. By (**x**) and (**ii**) we have *two* orthogonal matrices

$$P = \left[\begin{array}{ccccc} \mathbf{p}_1 & \cdots & \mathbf{p}_r & \cdots & \mathbf{p}_m \end{array}\right] \text{ of size } m\times m \quad \text{and} \quad Q = \left[\begin{array}{ccccc} \mathbf{q}_1 & \cdots & \mathbf{q}_r & \cdots & \mathbf{q}_n \end{array}\right] \text{ of size } n\times n$$

These matrices are related. In fact we have:

$$\sigma_i\mathbf{p}_i = \sqrt{\lambda_i}\mathbf{p}_i \overset{\text{(iii)}}{=} \|A\mathbf{q}_i\|\mathbf{p}_i \overset{\text{(viii)}}{=} A\mathbf{q}_i \quad \text{for each } i = 1,\ 2,\ \dots,\ r \tag{xi}$$

This yields the following expression for AQ in terms of its columns:

$$AQ = \left[\begin{array}{cccccc} A\mathbf{q}_1 & \cdots & A\mathbf{q}_r & A\mathbf{q}_{r+1} & \cdots & A\mathbf{q}_n \end{array}\right] \overset{\text{(iv)}}{=} \left[\begin{array}{cccccc} \sigma_1\mathbf{p}_1 & \cdots & \sigma_r\mathbf{p}_r & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right] \tag{xii}$$

Then we compute:

$$\begin{aligned} P\Sigma_A &= \left[\begin{array}{cccccc} \mathbf{p}_1 & \cdots & \mathbf{p}_r & \mathbf{p}_{r+1} & \cdots & \mathbf{p}_m \end{array}\right] \left[\begin{array}{cccccc} \sigma_1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right] \\ &= \left[\begin{array}{cccccc} \sigma_1\mathbf{p}_1 & \cdots & \sigma_r\mathbf{p}_r & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right] \\ &\overset{\text{(xii)}}{=} AQ \end{aligned}$$

Finally, as $Q^{-1} = Q^T$ it follows that $A = P\Sigma_A Q^T$.

With this we can state the main theorem of this Section.

Theorem 8.6.1

Let A be a real $m \times n$ matrix, and let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the positive singular values of A. Then r is the rank of A and we have the factorization

$$A = P\Sigma_A Q^T \qquad \text{where } P \text{ and } Q \text{ are orthogonal matrices}$$

The factorization $A = P\Sigma_A Q^T$ in Theorem 8.6.1, where P and Q are orthogonal matrices, is called a *Singular Value Decomposition* (*SVD*) of A. This decomposition is not unique. For example if $r < m$ then the vectors $\mathbf{p}_{r+1}, \dots, \mathbf{p}_m$ can be *any* extension of $\{\mathbf{p}_1, \dots, \mathbf{p}_r\}$ to an orthonormal basis of $\mathbb{R}^m$, and each will lead to a different matrix P in the decomposition. For a more dramatic example, if $A = I_n$ then $\Sigma_A = I_n$, and $A = P\Sigma_A P^T$ is a SVD of A for *any* orthogonal $n \times n$ matrix P.

Example 8.6.1

Find a singular value decomposition for $A = \left[\begin{array}{rrr} 1 & 0 & 1 \\ -1 & 1 & 0 \end{array} \right]$.

Solution. We have $A^TA = \left[\begin{array}{rrr} 2 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{array} \right]$, so the characteristic polynomial is

$$c_{A^TA}(x) = \det \left[\begin{array}{ccc} x-2 & 1 & -1 \\ 1 & x-1 & 0 \\ -1 & 0 & x-1 \end{array} \right] = (x-3)(x-1)x$$

Hence the eigenvalues of A^TA (in descending order) are $\lambda_1 = 3$, $\lambda_2 = 1$ and $\lambda_3 = 0$ with, respectively, unit eigenvectors

$$\mathbf{q}_1 = \frac{1}{\sqrt{6}} \left[\begin{array}{r} 2 \\ -1 \\ 1 \end{array} \right], \quad \mathbf{q}_2 = \frac{1}{\sqrt{2}} \left[\begin{array}{r} 0 \\ 1 \\ 1 \end{array} \right], \quad \text{and} \quad \mathbf{q}_3 = \frac{1}{\sqrt{3}} \left[\begin{array}{r} -1 \\ -1 \\ 1 \end{array} \right]$$

It follows that the orthogonal matrix Q in Theorem 8.6.1 is

$$Q = \left[\begin{array}{ccc} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{array} \right] = \frac{1}{\sqrt{6}} \left[\begin{array}{rrr} 2 & 0 & -\sqrt{2} \\ -1 & \sqrt{3} & -\sqrt{2} \\ 1 & \sqrt{3} & \sqrt{2} \end{array} \right]$$

The singular values here are $\sigma_1 = \sqrt{3}$, $\sigma_2 = 1$ and $\sigma_3 = 0$, so $\operatorname{rank}(A) = 2$—clear in this case—and the singular matrix is

$$\Sigma_A = \left[\begin{array}{ccc} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{array} \right] = \left[\begin{array}{ccc} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{array} \right]$$

So it remains to find the 2×2 orthogonal matrix P in Theorem 8.6.1. This involves the vectors

$$A\mathbf{q}_1 = \frac{\sqrt{6}}{2} \left[\begin{array}{r} 1 \\ -1 \end{array} \right], \quad A\mathbf{q}_2 = \frac{\sqrt{2}}{2} \left[\begin{array}{r} 1 \\ 1 \end{array} \right], \quad \text{and} \quad A\mathbf{q}_3 = \left[\begin{array}{r} 0 \\ 0 \end{array} \right]$$

Normalize $A\mathbf{q}_1$ and $A\mathbf{q}_2$ to get

$$\mathbf{p}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{p}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

In this case, $\{\mathbf{p}_1, \mathbf{p}_2\}$ is *already* a basis of $\mathbb{R}^2$ (so the Gram-Schmidt algorithm is not needed), and we have the 2×2 orthogonal matrix

$$P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Finally (by Theorem 8.6.1) the singular value decomposition for A is

$$A = P\Sigma_A Q^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\frac{1}{\sqrt{6}}\begin{bmatrix} 2 & -1 & 1 \\ 0 & \sqrt{3} & \sqrt{3} \\ -\sqrt{2} & -\sqrt{2} & \sqrt{2} \end{bmatrix}$$

Of course this can be confirmed by direct matrix multiplication.

Thus, computing an SVD for a real matrix A is a routine matter, and we now describe a systematic procedure for doing so.

SVD Algorithm

Given a real $m \times n$ matrix A, find an SVD $A = P\Sigma_A Q^T$ as follows:

1. *Use the Diagonalization Algorithm (see page 181) to find the (real and non-negative) eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ of A^TA with corresponding (orthonormal) eigenvectors $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$. Reorder the $\mathbf{q}_i$ (if necessary) to ensure that the nonzero eigenvalues are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$ and $\lambda_i = 0$ if $i > r$.*

2. *The integer r is the rank of the matrix A.*

3. *The $n \times n$ orthogonal matrix Q in the SVD is $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$.*

4. *Define $\mathbf{p}_i = \frac{1}{\|A\mathbf{q}_i\|}A\mathbf{q}_i$ for $i = 1, 2, \dots, r$ (where r is as in step 1). Then $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_r\}$ is orthonormal in $\mathbb{R}^m$ so (using Gram-Schmidt or otherwise) extend it to an orthonormal basis $\{\mathbf{p}_1, \dots, \mathbf{p}_r, \dots, \mathbf{p}_m\}$ in $\mathbb{R}^m$.*

5. *The $m \times m$ orthogonal matrix P in the SVD is $P = \begin{bmatrix} \mathbf{p}_1 & \cdots & \mathbf{p}_r & \cdots & \mathbf{p}_m \end{bmatrix}$.*

6. *The singular values for A are $\sigma_1, \sigma_2, \dots, \sigma_n$ where $\sigma_i = \sqrt{\lambda_i}$ for each i. Hence the nonzero singular values are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and so the singular matrix of A in the SVD is $\Sigma_A = \begin{bmatrix} \operatorname{diag}(\sigma_1, \dots, \sigma_r) & 0 \\ 0 & 0 \end{bmatrix}_{m \times n}$.*

7. *Thus $A = P\Sigma Q^T$ is a SVD for A.*

In practise the singular values σ_i, the matrices P and Q, and even the rank of an $m \times n$ matrix are not calculated this way. There are sophisticated numerical algorithms for calculating them to a high degree of accuracy. The reader is referred to books on numerical linear algebra.

So the main virtue of Theorem 8.6.1 is that it provides a way of *constructing* an SVD for every real matrix A. In particular it shows that every real matrix A *has* a singular value decomposition[9] in the following, more general, sense:

Definition 8.9

*A **Singular Value Decomposition (SVD)** of an $m \times n$ matrix A of rank r is a factorization $A = U\Sigma V^T$ where U and V are orthogonal and $\Sigma = \left[\begin{array}{cc} D & 0 \\ 0 & 0 \end{array}\right]_{m \times n}$ in block form where $D = \operatorname{diag}(d_1, d_2, \dots, d_r)$ where each $d_i > 0$, and $r \leq m$ and $r \leq n$.*

Note that for *any* SVD $A = U\Sigma V^T$ we immediately obtain some information about A:

Lemma 8.6.3

If $A = U\Sigma V^T$ is any SVD for A as in Definition 8.9, then:

1. $r = \operatorname{rank} A$.
2. *The numbers $d_1, d_2, \dots, d_r$ are the singular values of A^TA in some order.*

Proof. Use the notation of Definition 8.9. We have

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

so $\Sigma^T\Sigma$ and A^TA are similar $n \times n$ matrices (Definition 5.11). Hence $r = \operatorname{rank} A$ by Corollary 5.4.3, proving (1.). Furthermore, $\Sigma^T\Sigma$ and A^TA have the same eigenvalues by Theorem 5.5.1; that is (using (1.)):

$$\{d_1^2, d_2^2, \dots, d_r^2\} = \{\lambda_1, \lambda_2, \dots, \lambda_r\} \quad \text{are equal as sets}$$

where $\lambda_1, \lambda_2, \dots, \lambda_r$ are the positive eigenvalues of A^TA. Hence there is a permutation τ of $\{1, 2, \cdots, r\}$ such that $d_i^2 = \lambda_{i\tau}$ for each $i = 1, 2, \dots, r$. Hence $d_i = \sqrt{\lambda_{i\tau}} = \sigma_{i\tau}$ for each i by Definition 8.7. This proves (2.). □

We note in passing that more is true. Let A be $m \times n$ of rank r, and let $A = U\Sigma V^T$ be any SVD for A. Using the proof of Lemma 8.6.3 we have $d_i = \sigma_{i\tau}$ for some permutation τ of $\{1, 2, \dots, r\}$. In fact, it can be shown that there exist orthogonal matrices U_1 and V_1 obtained from U and V by τ-permuting columns and rows respectively, such that $A = U_1\Sigma_AV_1^T$ is an SVD of A.

[9] In fact every complex matrix has an SVD [J.T. Scheick, Linear Algebra with Applications, McGraw-Hill, 1997]

8.6.2. Fundamental Subspaces

It turns out that any singular value decomposition contains a great deal of information about an $m \times n$ matrix A and the subspaces associated with A. For example, in addition to Lemma 8.6.3, the set $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_r\}$ of vectors constructed in the proof of Theorem 8.6.1 is an orthonormal basis of col A (by (**v**) and (**viii**) in the proof). There are more such examples, which is the thrust of this subsection. In particular, there are four subspaces associated to a real $m \times n$ matrix A that have come to be called fundamental:

Definition 8.10

*The **fundamental subspaces** of an $m \times n$ matrix A are:*

$$\text{row } A = \text{span}\,\{\mathbf{x} \mid \mathbf{x} \textit{ is a row of } A\}$$

$$\text{col } A = \text{span}\,\{\mathbf{x} \mid \mathbf{x} \textit{ is a column of } A\}$$

$$\text{null } A = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

$$\text{null } A^T = \{\mathbf{x} \in \mathbb{R}^n \mid A^T\mathbf{x} = \mathbf{0}\}$$

If $A = U\Sigma V^T$ is *any* SVD for the real $m \times n$ matrix A, any orthonormal bases of U and V provide orthonormal bases for each of these fundamental subspaces. We are going to prove this, but first we need three properties related to the *orthogonal complement* $U^\perp$ of a subspace U of $\mathbb{R}^n$, where (Definition 8.1):

$$U^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{u} \cdot \mathbf{x} = 0 \text{ for all } \mathbf{u} \in U\}$$

The orthogonal complement plays an important role in the Projection Theorem (Theorem 8.1.3), and we return to it in Section 10.2. For now we need:

Lemma 8.6.4

If A is any matrix then:

1. $(\text{row } A)^\perp = \text{null } A$ *and* $(\text{col } A)^\perp = \text{null } A^T$.
2. *If U is any subspace of $\mathbb{R}^n$ then $U^{\perp\perp} = U$.*
3. *Let $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ be an orthonormal basis of $\mathbb{R}^m$. If $U = \text{span}\,\{\mathbf{f}_1, \dots, \mathbf{f}_k\}$, then*

$$U^\perp = \text{span}\,\{\mathbf{f}_{k+1}, \dots, \mathbf{f}_m\}$$

Proof.

1. Assume A is $m \times n$, and let $\mathbf{b}_1, \dots, \mathbf{b}_m$ be the rows of A. If $\mathbf{x}$ is a column in $\mathbb{R}^n$, then entry i of $A\mathbf{x}$ is $\mathbf{b}_i \cdot \mathbf{x}$, so $A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{b}_i \cdot \mathbf{x} = 0$ for each i. Thus:

$$\mathbf{x} \in \text{null } A \quad \Leftrightarrow \quad \mathbf{b}_i \cdot \mathbf{x} = 0 \text{ for each } i \quad \Leftrightarrow \quad \mathbf{x} \in (\text{span}\,\{\mathbf{b}_1, \dots, \mathbf{b}_m\})^\perp = (\text{row } A)^\perp$$

Hence $\text{null } A = (\text{row } A)^{\perp}$. Now replace A by A^T to get $\text{null } A^T = (\text{row } A^T)^{\perp} = (\text{col } A)^{\perp}$, which is the other identity in (1).

2. If $\mathbf{x} \in U$ then $\mathbf{y} \cdot \mathbf{x} = 0$ for all $\mathbf{y} \in U^{\perp}$, that is $\mathbf{x} \in U^{\perp\perp}$. This proves that $U \subseteq U^{\perp\perp}$, so it is enough to show that $\dim U = \dim U^{\perp\perp}$. By Theorem 8.1.4 we see that $\dim V^{\perp} = n - \dim V$ for any subspace $V \subseteq \mathbb{R}^n$. Hence

$$\dim U^{\perp\perp} = n - \dim U^{\perp} = n - (n - \dim U) = \dim U, \quad \text{as required}$$

3. We have $\text{span}\{\mathbf{f}_{k+1}, \dots, \mathbf{f}_m\} \subseteq U^{\perp}$ because $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ is orthogonal. For the other inclusion, let $\mathbf{x} \in U^{\perp}$ so $\mathbf{f}_i \cdot \mathbf{x} = 0$ for $i = 1, 2, \dots, k$. By the Expansion Theorem 5.3.6:

$$\begin{array}{rcccccccccccc} \mathbf{x} & = & (\mathbf{f}_1 \cdot \mathbf{x})\mathbf{f}_1 & + & \cdots & + & (\mathbf{f}_k \cdot \mathbf{x})\mathbf{f}_k & + & (\mathbf{f}_{k+1} \cdot \mathbf{x})\mathbf{f}_{k+1} & + & \cdots & + & (\mathbf{f}_m \cdot \mathbf{x})\mathbf{f}_m \\ & = & \mathbf{0} & + & \cdots & + & \mathbf{0} & + & (\mathbf{f}_{k+1} \cdot \mathbf{x})\mathbf{f}_{k+1} & + & \cdots & + & (\mathbf{f}_m \cdot \mathbf{x})\mathbf{f}_m \end{array}$$

Hence $U^{\perp} \subseteq \text{span}\{\mathbf{f}_{k+1}, \dots, \mathbf{f}_m\}$.

□

With this we can see how *any* SVD for a matrix A provides orthonormal bases for each of the four fundamental subspaces of A.

Theorem 8.6.2

Let A be an $m \times n$ real matrix, let $A = U\Sigma V^T$ be any SVD for A where U and V are orthogonal of size $m \times m$ and $n \times n$ respectively, and let

$$\Sigma = \left[\begin{array}{cc} D & 0 \\ 0 & 0 \end{array}\right]_{m \times n} \qquad \text{where} \qquad D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r), \quad \text{with each } \lambda_i > 0$$

Write $U = \left[\begin{array}{ccccc} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \end{array}\right]$ and $V = \left[\begin{array}{ccccc} \mathbf{v}_1 & \cdots & \mathbf{v}_r & \cdots & \mathbf{v}_n \end{array}\right]$, so $\{\mathbf{u}_1, \dots, \mathbf{u}_r, \dots, \mathbf{u}_m\}$ and $\{\mathbf{v}_1, \dots, \mathbf{v}_r, \dots, \mathbf{v}_n\}$ are orthonormal bases of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively. Then

1. *$r = \text{rank } A$, and the singular values of A are $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_r}$.*

2. *The fundamental spaces are described as follows:*

   a. *$\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ is an orthonormal basis of $\text{col } A$.*
   b. *$\{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}$ is an orthonormal basis of $\text{null } A^T$.*
   c. *$\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\}$ is an orthonormal basis of $\text{null } A$.*
   d. *$\{\mathbf{v}_1, \dots, \mathbf{v}_r\}$ is an orthonormal basis of $\text{row } A$.*

Proof.

1. This is Lemma 8.6.3.

2. a. As $\text{col}\, A = \text{col}\,(AV)$ by Lemma 5.4.3 and $AV = U\Sigma$, (a.) follows from

$$U\Sigma = \left[\begin{array}{ccccc} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \end{array}\right] \left[\begin{array}{cc} \text{diag}\,(\lambda_1,\, \lambda_2,\, \dots,\, \lambda_r) & 0 \\ 0 & 0 \end{array}\right] = \left[\begin{array}{cccccc} \lambda_1\mathbf{u}_1 & \cdots & \lambda_r\mathbf{u}_r & \mathbf{0} & \cdots & \mathbf{0} \end{array}\right]$$

b. We have $(\text{col}\, A)^\perp \overset{\text{(a.)}}{=} (\text{span}\,\{\mathbf{u}_1,\, \dots,\, \mathbf{u}_r\})^\perp = \text{span}\,\{\mathbf{u}_{r+1},\, \dots,\, \mathbf{u}_m\}$ by Lemma 8.6.4(3). This proves (b.) because $(\text{col}\, A)^\perp = \text{null}\, A^T$ by Lemma 8.6.4(1).

c. We have $\dim\,(\text{null}\, A) + \dim\,(\text{im}\, A) = n$ by the Dimension Theorem 7.2.4, applied to $T : \mathbb{R}^n \to \mathbb{R}^m$ where $T(\mathbf{x}) = A\mathbf{x}$. Since also $\text{im}\, A = \text{col}\, A$ by Lemma 8.6.1, we obtain

$$\dim\,(\text{null}\, A) = n - \dim\,(\text{col}\, A) = n - r = \dim\,(\text{span}\,\{\mathbf{v}_{r+1},\, \dots,\, \mathbf{v}_n\})$$

So to prove (c.) it is enough to show that $\mathbf{v}_j \in \text{null}\, A$ whenever $j > r$. To this end write

$$\lambda_{r+1} = \cdots = \lambda_n = 0, \quad \text{so} \quad E^T E = \text{diag}\,(\lambda_1^2,\, \dots,\, \lambda_r^2,\, \lambda_{r+1}^2,\, \dots,\, \lambda_n^2)$$

Observe that each λ_j is an eigenvalue of $\Sigma^T\Sigma$ with eigenvector $\mathbf{e}_j$ = column j of I_n. Thus $\mathbf{v}_j = V\mathbf{e}_j$ for each j. As $A^TA = V\Sigma^T\Sigma V^T$ (proof of Lemma 8.6.3), we obtain

$$(A^TA)\mathbf{v}_j = (V\Sigma^T\Sigma V^T)(V\mathbf{e}_j) = V(\Sigma^T\Sigma\mathbf{e}_j) = V\left(\lambda_j^2\mathbf{e}_j\right) = \lambda_j^2 V\mathbf{e}_j = \lambda_j^2\mathbf{v}_j$$

for $1 \le j \le n$. Thus each $\mathbf{v}_j$ is an eigenvector of A^TA corresponding to λ_j^2. But then

$$\|A\mathbf{v}_j\|^2 = (A\mathbf{v}_j)^T A\mathbf{v}_j = \mathbf{v}_j^T(A^TA\mathbf{v}_j) = \mathbf{v}_j^T(\lambda_j^2\mathbf{v}_j) = \lambda_j^2\|\mathbf{v}_j\|^2 = \lambda_j^2 \quad \text{for } i = 1,\, \dots,\, n$$

In particular, $A\mathbf{v}_j = \mathbf{0}$ whenever $j > r$, so $\mathbf{v}_j \in \text{null}\, A$ if $j > r$, as desired. This proves (c).

d. Observe that $\text{span}\,\{\mathbf{v}_{r+1},\, \dots,\, \mathbf{v}_n\} \overset{\text{(c.)}}{=} \text{null}\, A = (\text{row}\, A)^\perp$ by Lemma 8.6.4(1). But then parts (2) and (3) of Lemma 8.6.4 show

$$\text{row}\, A = \left((\text{row}\ A)^\perp\right)^\perp = (\text{span}\,\{\mathbf{v}_{r+1},\, \dots,\, \mathbf{v}_n\})^\perp = \text{span}\,\{\mathbf{v}_1,\, \dots,\, \mathbf{v}_r\}$$

This proves (d.), and hence Theorem 8.6.2.

□

Example 8.6.2

Consider the homogeneous linear system

$$A\mathbf{x} = \mathbf{0} \text{ of } m \text{ equations in } n \text{ variables}$$

Then the set of all solutions is $\text{null}\, A$. Hence if $A = U\Sigma V^T$ is any SVD for A then (in the notation of Theorem 8.6.2) $\{\mathbf{v}_{r+1},\, \dots,\, \mathbf{v}_n\}$ is an orthonormal basis of the set of solutions for the system. As such they are a set of **basic solutions** for the system, the most basic notion in Chapter 1.

8.6.3. The Polar Decomposition of a Real Square Matrix

If A is real and $n \times n$ the factorization in the title is related to the polar decomposition A. Unlike the SVD, in this case the decomposition is *uniquely* determined by A.

Recall (Section 8.3) that a symmetric matrix A is called positive definite if and only if $\mathbf{x}^T A\mathbf{x} > 0$ for every column $\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n$. Before proceeding, we must explore the following weaker notion:

Definition 8.11

A real $n \times n$ matrix G is called ***positive***[10] *if it is symmetric and*

$$\mathbf{x}^T G\mathbf{x} \geq 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n$$

Clearly every positive definite matrix is positive, but the converse fails. Indeed, $A = \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right]$ is positive because, if $\mathbf{x} = \left[\begin{array}{cc} a & b \end{array} \right]^T$ in $\mathbb{R}^2$, then $\mathbf{x}^T A\mathbf{x} = (a+b)^2 \geq 0$. But $\mathbf{y}^T A\mathbf{y} = 0$ if $\mathbf{y} = \left[\begin{array}{cc} 1 & -1 \end{array} \right]^T$, so A is not positive definite.

Lemma 8.6.5

Let G denote an $n \times n$ positive matrix.

1. *If A is any $m \times n$ matrix and G is positive, then $A^T GA$ is positive (and $m \times m$).*
2. *If $G = \operatorname{diag}(d_1, d_2, \cdots, d_n)$ and each $d_i \geq 0$ then G is positive.*

Proof.

1. $\mathbf{x}^T (A^T GA)\mathbf{x} = (A\mathbf{x})^T G(A\mathbf{x}) \geq 0$ because G is positive.
2. If $\mathbf{x} = \left[\begin{array}{cccc} x_1 & x_2 & \cdots & x_n \end{array} \right]^T$, then

$$\mathbf{x}^T G\mathbf{x} = d_1 x_1^2 + d_2 x_2^2 + \cdots + d_n x_n^2 \geq 0$$

because $d_i \geq 0$ for each i.

□

Definition 8.12

If A is a real $n \times n$ matrix, a factorization

$$A = GQ \text{ where } G \text{ is positive and } Q \text{ is orthogonal}$$

is called a ***polar decomposition*** *for A.*

[10] Also called **positive semi-definite.**

Any SVD for a real square matrix A yields a polar form for A.

Theorem 8.6.3

Every square real matrix has a polar form.

Proof. Let $A = U\Sigma V^T$ be a SVD for A with Σ as in Definition 8.9 and $m = n$. Since $U^TU = I_n$ here we have

$$A = U\Sigma V^T = (U\Sigma)(U^TU)V^T = (U\Sigma U^T)(UV^T)$$

So if we write $G = U\Sigma U^T$ and $Q = UV^T$, then Q is orthogonal, and it remains to show that G is positive. But this follows from Lemma 8.6.5. □

The SVD for a square matrix A is not unique ($I_n = PI_nP^T$ for any orthogonal matrix P). But given the proof of Theorem 8.6.3 it is surprising that the polar decomposition *is* unique.[11] We omit the proof.

The name "polar form" is reminiscent of the same form for complex numbers (see Appendix A). This is no coincidence. To see why, we represent the complex numbers as real 2×2 matrices. Write $\mathbf{M}_2(\mathbb{R})$ for the set of all real 2×2 matrices, and define

$$\sigma : \mathbb{C} \to \mathbf{M}_2(\mathbb{R}) \quad \text{by} \quad \sigma(a+bi) = \left[\begin{array}{rr} a & -b \\ b & a \end{array}\right] \text{ for all } a+bi \text{ in } \mathbb{C}$$

One verifies that σ preserves addition and multiplication in the sense that

$$\sigma(zw) = \sigma(z)\sigma(w) \qquad \text{and} \qquad \sigma(z+w) = \sigma(z) + \sigma(w)$$

for all complex numbers z and w. Since θ is one-to-one we may *identify* each complex number $a+bi$ with the matrix $\theta(a+bi)$, that is we write

$$a+bi = \left[\begin{array}{rr} a & -b \\ b & a \end{array}\right] \quad \text{for all } a+bi \text{ in } \mathbb{C}$$

Thus $0 = \left[\begin{array}{rr} 0 & 0 \\ 0 & 0 \end{array}\right]$, $1 = \left[\begin{array}{rr} 1 & 0 \\ 0 & 1 \end{array}\right] = I_2$, $i = \left[\begin{array}{rr} 0 & -1 \\ 1 & 0 \end{array}\right]$, and $r = \left[\begin{array}{rr} r & 0 \\ 0 & r \end{array}\right]$ if r is real.

If $z = a+bi$ is nonzero then the *absolute value* $r = |z| = \sqrt{a^2+b^2} \neq 0$. If θ is the *angle* of z in standard position, then $\cos\theta = a/r$ and $\sin\theta = b/r$. Observe:

$$\left[\begin{array}{rr} a & -b \\ b & a \end{array}\right] = \left[\begin{array}{rr} r & 0 \\ 0 & r \end{array}\right]\left[\begin{array}{rr} a/r & -b/r \\ b/r & a/r \end{array}\right] = \left[\begin{array}{rr} r & 0 \\ 0 & r \end{array}\right]\left[\begin{array}{rr} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right] = GQ \qquad \textbf{(xiii)}$$

where $G = \left[\begin{array}{rr} r & 0 \\ 0 & r \end{array}\right]$ is positive and $Q = \left[\begin{array}{rr} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array}\right]$ is orthogonal. But in $\mathbb{C}$ we have $G = r$ and $Q = \cos\theta + i\sin\theta$ so (**xiii**) reads $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$ which is the *classical polar form* for the complex number $a+bi$. This is why (**xiii**) is called the polar form of the matrix $\left[\begin{array}{rr} a & -b \\ b & a \end{array}\right]$; Definition 8.12 simply adopts the terminology for $n \times n$ matrices.

[11] See J.T. Scheick, Linear Algebra with Applications, McGraw-Hill, 1997, page 379.

8.6.4. The Pseudoinverse of a Matrix

It is impossible for a non-square matrix A to have an inverse (see the footnote to Definition 2.11). Nonetheless, one candidate for an "inverse" of A is an $m \times n$ matrix B such that

$$ABA = A \qquad \text{and} \qquad BAB = B$$

Such a matrix B is called a *middle inverse* for A. If A is invertible then A^{-1} is the unique middle inverse for A, but a middle inverse is not unique in general, even for square matrices. For example, if $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ then $B = \begin{bmatrix} 1 & 0 & 0 \\ b & 0 & 0 \end{bmatrix}$ is a middle inverse for A for any b.

If $ABA = A$ and $BAB = B$ it is easy to see that AB and BA are both idempotent matrices. In 1955 Roger Penrose observed that the middle inverse is unique if both AB and BA are symmetric. We omit the proof.

Theorem 8.6.4: Penrose' Theorem[12]

Given any real $m \times n$ matrix A, there is exactly one $n \times m$ matrix B such that A and B satisfy the following conditions:

P1 *$ABA = A$ and $BAB = B$.*

P2 *Both AB and BA are symmetric.*

Definition 8.13

Let A be a real $m \times n$ matrix. The **pseudoinverse** *of A is the unique $n \times m$ matrix A^+ such that A and A^+ satisfy* **P1** *and* **P2***, that is:*

$$AA^+A = A, \qquad A^+AA^+ = A^+, \qquad \text{and both } AA^+ \text{ and } A^+A \text{ are symmetric}$$[13]

If A is invertible then $A^+ = A^{-1}$ as expected. In general, the symmetry in conditions P1 and P2 shows that A is the pseudoinverse of A^+, that is $A^{++} = A$.

---

[12] R. Penrose, *A generalized inverse for matrices*, Proceedings of the Cambridge Philosophical Society **51** (1955), 406-413. In fact Penrose proved this for any complex matrix, where AB and BA are both required to be hermitian (see Definition 8.18 in the following section).

[13] Penrose called the matrix A^+ the generalized inverse of A, but the term pseudoinverse is now commonly used. The matrix A^+ is also called the **Moore-Penrose** inverse after E.H. Moore who had the idea in 1935 as part of a larger work on "General Analysis". Penrose independently re-discovered it 20 years later.

Theorem 8.6.5

Let A be an $m \times n$ matrix.

1. *If* rank $A = m$ *then* AA^T *is invertible and* $A^+ = A^T(AA^T)^{-1}$.
2. *If* rank $A = n$ *then* A^TA *is invertible and* $A^+ = (A^TA)^{-1}A^T$.

Proof. Here AA^T (respectively A^TA) is invertible by Theorem 5.4.4 (respectively Theorem 5.4.3). The rest is a routine verification. □

In general, given an $m \times n$ matrix A, the pseudoinverse A^+ can be computed from any SVD for A. To see how, we need some notation. Let $A = U\Sigma V^T$ be an SVD for A (as in Definition 8.9) where U and V are orthogonal and $\Sigma = \left[\begin{array}{cc} D & 0 \\ 0 & 0 \end{array}\right]_{m \times n}$ in block form where $D = \text{diag}\,(d_1,\ d_2,\ \dots,\ d_r)$ where each $d_i > 0$. Hence D is invertible, so we make:

Definition 8.14

$$\Sigma' = \left[\begin{array}{cc} D^{-1} & 0 \\ 0 & 0 \end{array}\right]_{n \times m}.$$

A routine calculation gives:

Lemma 8.6.6

- $\Sigma\Sigma'\Sigma = \Sigma$
- $\Sigma\Sigma' = \left[\begin{array}{cc} I_r & 0 \\ 0 & 0 \end{array}\right]_{m \times m}$
- $\Sigma'\Sigma\Sigma' = \Sigma'$
- $\Sigma'\Sigma = \left[\begin{array}{cc} I_r & 0 \\ 0 & 0 \end{array}\right]_{n \times n}$

That is, Σ' is the pseudoinverse of Σ.

Now given $A = U\Sigma V^T$, define $B = V\Sigma' U^T$. Then

$$ABA = (U\Sigma V^T)(V\Sigma' U^T)(U\Sigma V^T) = U(\Sigma\Sigma'\Sigma)V^T = U\Sigma V^T = A$$

by Lemma 8.6.6. Similarly $BAB = B$. Moreover $AB = U(\Sigma\Sigma')U^T$ and $BA = V(\Sigma'\Sigma)V^T$ are both symmetric again by Lemma 8.6.6. This proves

Theorem 8.6.6

Let A be real and $m \times n$, and let $A = U\Sigma V^T$ is any SVD for A as in Definition 8.9. Then $A^+ = V\Sigma' U^T$.

Of course we can always use the SVD constructed in Theorem 8.6.1 to find the pseudoinverse. If $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$, we observed above that $B = \begin{bmatrix} 1 & 0 & 0 \\ b & 0 & 0 \end{bmatrix}$ is a middle inverse for A for any b. Furthermore AB is symmetric but BA is not, so $B \neq A^+$.

Example 8.6.3

Find A^+ if $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Solution. $A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ with eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 0$ and corresponding eigenvectors $\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{q}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Hence $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{bmatrix} = I_2$. Also A has rank 1 with singular values $\sigma_1 = 1$ and $\sigma_2 = 0$, so $\Sigma_A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = A$ and $\Sigma_A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = A^T$ in this case.

Since $A\mathbf{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and $A\mathbf{q}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, we have $\mathbf{p}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ which extends to an orthonormal basis $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ of $\mathbb{R}^3$ where (say) $\mathbf{p}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{p}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Hence $P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} = I$, so the SVD for A is $A = P\Sigma_A Q^T$. Finally, the pseudoinverse of A is $A^+ = Q\Sigma_A' P^T = \Sigma_A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Note that $A^+ = A^T$ in this case.

The following Lemma collects some properties of the pseudoinverse that mimic those of the inverse. The verifications are left as exercises.

Lemma 8.6.7

Let A be an $m \times n$ matrix of rank r.

1. $A^{++} = A$.
2. *If A is invertible then* $A^+ = A^{-1}$.
3. $(A^T)^+ = (A^+)^T$.
4. $(kA)^+ = kA^+$ *for any real k.*
5. $(UAV)^+ = U^T(A^+)V^T$ *whenever U and V are orthogonal.*

Exercises for 8.6

Exercise 8.6.1 If $ACA = A$ show that $B = CAC$ is a middle inverse for A.

Exercise 8.6.2 For any matrix A show that

$$\Sigma_{A^T} = (\Sigma_A)^T$$

Exercise 8.6.3 If A is $m \times n$ with all singular values positive, what is $\operatorname{rank} A$?

Exercise 8.6.4 If A has singular values $\sigma_1, \dots, \sigma_r$, what are the singular values of:

a. A^T

b. tA where $t > 0$ is real

c. A^{-1} assuming A is invertible.

Exercise 8.6.5 If A is square show that $|\det A|$ is the product of the singular values of A.

Exercise 8.6.6 If A is square and real, show that $A = 0$ if and only if every eigenvalue of A is 0.

Exercise 8.6.7 Given a SVD for an invertible matrix A, find one for A^{-1}. How are Σ_A and $\Sigma_{A^{-1}}$ related?

Exercise 8.6.8 Let $A^{-1} = A = A^T$ where A is $n \times n$. Given any orthogonal $n \times n$ matrix U, find an orthogonal matrix V such that $A = U\Sigma_A V^T$ is an SVD for A.

If $A = \left[\begin{array}{rr} 0 & 1 \\ 1 & 0 \end{array}\right]$ do this for:

a. $U = \frac{1}{5}\left[\begin{array}{rr} 3 & -4 \\ 4 & 3 \end{array}\right]$

b. $U = \frac{1}{\sqrt{2}}\left[\begin{array}{rr} 1 & -1 \\ 1 & 1 \end{array}\right]$

Exercise 8.6.9 Find a SVD for the following matrices:

a. $A = \left[\begin{array}{rr} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{array}\right]$

b. $\left[\begin{array}{rrr} 1 & 1 & 1 \\ -1 & 0 & -2 \\ 1 & 2 & 0 \end{array}\right]$

Exercise 8.6.10 Find an SVD for $A = \left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array}\right]$.

Exercise 8.6.11 If $A = U\Sigma V^T$ is an SVD for A, find an SVD for A^T.

Exercise 8.6.12 Let A be a real, $m \times n$ matrix with positive singular values $\sigma_1, \sigma_2, \dots, \sigma_r$, and write

$$s(x) = (x - \sigma_1)(x - \sigma_2)\cdots(x - \sigma_r)$$

a. Show that $c_{A^TA}(x) = s(x)x^{n-r}$ and $c_{A^TA}(c) = s(x)x^{m-r}$.

b. If $m \leq n$ conclude that $c_{A^TA}(x) = s(x)x^{n-m}$.

Exercise 8.6.13 If G is positive show that:

a. rG is positive if $r \geq 0$

b. $G + H$ is positive for any positive H.

Exercise 8.6.14 If G is positive and λ is an eigenvalue, show that $\lambda \geq 0$.

Exercise 8.6.15 If G is positive show that $G = H^2$ for some positive matrix H. [Hint: Preceding exercise and Lemma 8.6.5]

Exercise 8.6.16 If A is $n \times n$ show that AA^T and A^TA are similar. [Hint: Start with an SVD for A.]

Exercise 8.6.17 Find A^+ if:

a. $A = \left[\begin{array}{rr} 1 & 2 \\ -1 & -2 \end{array}\right]$

b. $A = \left[\begin{array}{rr} 1 & -1 \\ 0 & 0 \\ 1 & -1 \end{array}\right]$

Exercise 8.6.18 Show that $(A^+)^T = (A^T)^+$.

8.7 Complex Matrices

If A is an $n \times n$ matrix, the characteristic polynomial $c_A(x)$ is a polynomial of degree n and the eigenvalues of A are just the roots of $c_A(x)$. In most of our examples these roots have been *real* numbers (in fact, the examples have been carefully chosen so this will be the case!); but it need not happen, even when the characteristic polynomial has real coefficients. For example, if $A = \left[\begin{array}{rr} 0 & 1 \\ -1 & 0 \end{array} \right]$ then $c_A(x) = x^2 + 1$ has roots i and $-i$, where i is a complex number satisfying $i^2 = -1$. Therefore, we have to deal with the possibility that the eigenvalues of a (real) square matrix might be complex numbers.

In fact, nearly everything in this book would remain true if the phrase *real number* were replaced by *complex number* wherever it occurs. Then we would deal with matrices with complex entries, systems of linear equations with complex coefficients (and complex solutions), determinants of complex matrices, and vector spaces with scalar multiplication by any complex number allowed. Moreover, the proofs of most theorems about (the real version of) these concepts extend easily to the complex case. It is not our intention here to give a full treatment of complex linear algebra. However, we will carry the theory far enough to give another proof that the eigenvalues of a real symmetric matrix A are real (Theorem 5.5.7) and to prove the spectral theorem, an extension of the principal axes theorem (Theorem 8.2.2).

The set of complex numbers is denoted $\mathbb{C}$. We will use only the most basic properties of these numbers (mainly conjugation and absolute values), and the reader can find this material in Appendix A.

If $n \geq 1$, we denote the set of all n-tuples of complex numbers by $\mathbb{C}^n$. As with $\mathbb{R}^n$, these n-tuples will be written either as row or column matrices and will be referred to as **vectors**. We define vector operations on $\mathbb{C}^n$ as follows:

$$\begin{aligned} (v_1,\, v_2,\, \dots,\, v_n) + (w_1,\, w_2,\, \dots,\, w_n) &= (v_1 + w_1,\, v_2 + w_2,\, \dots,\, v_n + w_n) \\ u(v_1,\, v_2,\, \dots,\, v_n) &= (uv_1,\, uv_2,\, \dots,\, uv_n) \quad \text{for } u \text{ in } \mathbb{C} \end{aligned}$$

With these definitions, $\mathbb{C}^n$ satisfies the axioms for a vector space (with complex scalars) given in Chapter 6. Thus we can speak of spanning sets for $\mathbb{C}^n$, of linearly independent subsets, and of bases. In all cases, the definitions are identical to the real case, except that the scalars are allowed to be complex numbers. In particular, the standard basis of $\mathbb{R}^n$ remains a basis of $\mathbb{C}^n$, called the **standard basis** of $\mathbb{C}^n$.

A matrix $A = \left[a_{ij}\right]$ is called a **complex matrix** if every entry a_{ij} is a complex number. The notion of conjugation for complex numbers extends to matrices as follows: Define the **conjugate** of $A = \left[a_{ij}\right]$ to be the matrix

$$\overline{A} = \left[\begin{array}{c} \overline{a}_{ij} \end{array} \right]$$

obtained from A by conjugating every entry. Then (using Appendix A)

$$\overline{A+B} = \overline{A} + \overline{B} \quad \text{ and } \quad \overline{AB} = \overline{A}\,\overline{B}$$

holds for all (complex) matrices of appropriate size.