3 · Singular Value Decomposition

Thus far we have focused on matrix factorizations that reveal the eigenvalues of a square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$, such as the SCHUR factorization and the JORDAN canonical form. Eigenvalue-based decompositions are ideal for analyzing the behavior of dynamical systems likes $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ or $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$. When it comes to solving linear systems of equations or tackling more general problems in data science, eigenvalue-based factorizations are often not so illuminating. In this chapter we develop another decomposition that provides deep insight into the rank structure of a matrix, showing the way to solving all variety of linear equations and exposing optimal low-rank approximations.

3.1 Singular Value Decomposition

The *singular value decomposition (SVD)* is remarkable factorization that writes a general rectangular matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ in the form

$$\mathbf{A} = (\text{unitary matrix}) \times (\text{diagonal matrix}) \times (\text{unitary matrix})^*.$$

From the unitary matrices we can extract bases for the four fundamental subspaces $\mathcal{R}(\mathbf{A})$, $\mathcal{N}(\mathbf{A})$, $\mathcal{R}(\mathbf{A}^*)$, and $\mathcal{N}(\mathbf{A}^*)$, and the diagonal matrix will reveal much about the rank structure of $\mathbf{A}$.

We will build up the SVD in a four-step process. For simplicity suppose that $\mathbf{A} \in \mathbb{C}^{m\times n}$ with $m \geq n$. (If $m < n$, apply the arguments below to $\mathbf{A}^* \in \mathbb{C}^{n\times m}$.) Note that $\mathbf{A}^*\mathbf{A} \in \mathbb{C}^{n\times n}$ is always Hermitian positive semidefinite. (Clearly $(\mathbf{A}^*\mathbf{A})^* = \mathbf{A}^*(\mathbf{A}^*)^* = \mathbf{A}^*\mathbf{A}$, so $\mathbf{A}^*\mathbf{A}$ is Hermitian. For any $\mathbf{x} \in \mathbb{C}^n$, note that $\mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^*(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$, so $\mathbf{A}^*\mathbf{A}$ is positive semidefinite.)

Step 1. Using the spectral decomposition of a Hermitian matrix discussed in Section 1.5, $\mathbf{A}^*\mathbf{A}$ has n eigenpairs $\{(\lambda_j, \mathbf{v}_j)\}_{j=1}^n$ with orthonormal unit eigenvectors ($\mathbf{v}_j^*\mathbf{v}_j = 1$, $\mathbf{v}_j^*\mathbf{v}_k = 0$ when $j \neq k$). We are free to pick any convenient indexing for these eigenpairs; label the eigenvalues in decreasing magnitude, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$.

Step 2. Define $s_j := \|\mathbf{A}\mathbf{v}_j\|$.
Note that $s_j^2 = \|\mathbf{A}\mathbf{v}_j\|^2 = \mathbf{v}_j^*\mathbf{A}^*\mathbf{A}\mathbf{v}_j = \lambda_j$. Since the eigenvalues $\lambda_1, \ldots, \lambda_n$ are decreasing in magnitude, so are the s_j values: $s_1 \geq s_2 \geq \cdots \geq s_n \geq 0$.

Step 3. Next, we will build a set of related orthonormal vectors in $\mathbb{C}^m$. Suppose we have already constructed such vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}$.
If $s_j \neq 0$, then define $\mathbf{u}_j = s_j^{-1}\mathbf{A}\mathbf{v}_j$, so that $\|\mathbf{u}_j\| = s_j^{-1}\|\mathbf{A}\mathbf{v}_j\| = 1$.
If $s_j = 0$, then pick $\mathbf{u}_j$ to be any unit vector such that

$$\mathbf{u}_j \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}\}^{\perp};$$

i.e., ensure $\mathbf{u}_j^*\mathbf{u}_k = 0$ for all $k < j$.[1]

By construction, $\mathbf{u}_j^*\mathbf{u}_k = 0$ for $j \neq k$ if s_j or s_k is zero. If both s_j and s_k are nonzero, then

$$\mathbf{u}_j^*\mathbf{u}_k = \frac{1}{s_j s_k}(\mathbf{A}\mathbf{v}_j)^*(\mathbf{A}\mathbf{v}_k) = \frac{1}{s_j s_k}\mathbf{v}_j^*\mathbf{A}^*\mathbf{A}\mathbf{v}_k = \frac{\lambda_k}{s_j s_k}\mathbf{v}_j^*\mathbf{v}_k,$$

where we used the fact that $\mathbf{v}_j$ is an eigenvector of $\mathbf{A}^*\mathbf{A}$. Now if $j \neq k$, then $\mathbf{v}_j^*\mathbf{v}_k = 0$, and hence $\mathbf{u}_j^*\mathbf{u}_k = 0$. On the other hand, $j = k$ implies that $\mathbf{v}_j^*\mathbf{v}_k = 1$, so $\mathbf{u}_j^*\mathbf{u}_k = \lambda_j / s_j^2 = 1$.

In conclusion, we have constructed a set of orthonormal vectors $\{\mathbf{u}_j\}_{j=1}^n$ with $\mathbf{u}_j \in \mathbb{C}^m$.
Step 4. For all $j = 1, \ldots, n$,

$$\mathbf{A}\mathbf{v}_j = s_j \mathbf{u}_j,$$

regardless of whether $s_j = 0$ or not. We can stack these n vector equations as columns of a single matrix equation,

$$\left[\begin{array}{cccc} | & | & & | \\ \mathbf{A}\mathbf{v}_1 & \mathbf{A}\mathbf{v}_2 & \cdots & \mathbf{A}\mathbf{v}_n \\ | & | & & | \end{array}\right] = \left[\begin{array}{cccc} | & | & & | \\ s_1\mathbf{u}_1 & s_2\mathbf{u}_2 & \cdots & s_n\mathbf{u}_n \\ | & | & & | \end{array}\right].$$

[1] If $s_j = 0$, then $\lambda_j = 0$, and so $\mathbf{A}^*\mathbf{A}$ has a zero eigenvalue; i.e., this matrix is singular.

Note that both matrices in this equation can be factored into the product of simpler matrices:

$$\mathbf{A}\begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_n \end{bmatrix}.$$

Denote these matrices as $\mathbf{AV} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}$, where $\mathbf{A} \in \mathbb{C}^{m\times n}$, $\mathbf{V} \in \mathbb{C}^{n\times n}$, $\widehat{\mathbf{U}} \in \mathbb{C}^{m\times n}$, and $\widehat{\mathbf{\Sigma}} \in \mathbb{C}^{n\times n}$.

The (j,k) entry of $\mathbf{V}^*\mathbf{V}$ is simply $\mathbf{v}_j^*\mathbf{v}_k$, and so $\mathbf{V}^*\mathbf{V} = \mathbf{I}$. Since $\mathbf{V}$ is a square matrix, we have just proved that it is unitary, and hence, $\mathbf{VV}^* = \mathbf{I}$ as well. We conclude that

$$\mathbf{A} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\mathbf{V}^*.$$

This matrix factorization is known as the *reduced singular value decomposition* or the *economy-sized singular value decomposition* (or, informally, the *skinny SVD*). It can be obtained via the MATLAB command

```
[Uhat, Sighat, V] = svd(A,'econ');
```

The Reduced Singular Value Decomposition

Theorem 3.1. *Any matrix* $\mathbf{A} \in \mathbb{C}^{m\times n}$ *with* $m \geq n$ *can be written as*

$$\mathbf{A} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\mathbf{V}^*,$$

where $\widehat{\mathbf{U}} \in \mathbb{C}^{m\times n}$ *has orthonormal columns,* $\mathbf{V} \in \mathbb{C}^{n\times n}$ *is unitary, and* $\widehat{\mathbf{\Sigma}} = \operatorname{diag}(s_1,\ldots,s_n) \in \mathbb{C}^{n\times n}$ *has real nonnegative decreasing entries. The columns of* $\mathbf{U}$ *are left singular vectors, the columns of* $\mathbf{V}$ *are right singular vectors, and the values* $s_1,\ldots,s_n$ *are the singular values.*

While the matrix $\widehat{\mathbf{U}}$ has orthonormal columns, *it is not a unitary matrix* when $m > n$. In particular, we have $\widehat{\mathbf{U}}^*\widehat{\mathbf{U}} = \mathbf{I} \in \mathbb{C}^{n\times n}$, but

$$\widehat{\mathbf{U}}\widehat{\mathbf{U}}^* \in \mathbb{C}^{m\times m}$$

cannot be the identity unless $m = n$. (To see this, note that $\widehat{\mathbf{U}}\widehat{\mathbf{U}}^*$ is an orthogonal projection onto $\mathcal{R}(\widehat{\mathbf{U}}) = \operatorname{span}\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$. Since $\dim(\mathcal{R}(\widehat{\mathbf{U}})) = n$, this projection cannot equal the m-by-m identity matrix when $m > n$.)

Though $\widehat{\mathbf{U}}$ is not unitary, it is *subunitary*. We can construct $m - n$ additional column vectors to append to $\widehat{\mathbf{U}}$ to make it unitary. Here is the recipe: For $j = n+1, \ldots, m$, pick

$$\mathbf{u}_j \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{j-1}\}^{\perp}$$

with $\mathbf{u}_j^*\mathbf{u}_j = 1$. Then define

$$\mathbf{U} = \left[\begin{array}{cccc} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ | & | & & | \end{array}\right].$$

Confirm that $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I} \in \mathbb{C}^{m\times m}$, showing that $\mathbf{U}$ is unitary.

We wish to replace the $\widehat{\mathbf{U}}$ in the reduced SVD with the unitary matrix $\mathbf{U}$. To do so, we also need to replace $\widehat{\mathbf{\Sigma}}$ by some $\mathbf{\Sigma}$ in such a way that $\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}} = \mathbf{U}\mathbf{\Sigma}$. The simplest approach constructs $\mathbf{\Sigma}$ by appending zeros to the end of $\widehat{\mathbf{\Sigma}}$, thus ensuring there is no contribution when the new entries of $\mathbf{U}$ multiply against the new entries of $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} = \left[\begin{array}{c} \widehat{\mathbf{\Sigma}} \\ \mathbf{0} \end{array}\right] \in \mathbb{C}^{m\times n}.$$

The factorization $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ is called the *full singular value decomposition*.

The Full Singular Value Decomposition

Theorem 3.2. *Any matrix* $\mathbf{A} \in \mathbb{C}^{m\times n}$ *can be written in the form*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*,$$

where $\mathbf{U} \in \mathbb{C}^{m\times m}$ *and* $\mathbf{V} \in \mathbb{C}^{n\times n}$ *are unitary matrices and* $\mathbf{\Sigma} \in \mathbb{C}^{m\times n}$ *is zero except for the main diagonal;*
The columns of $\mathbf{U}$ *are left singular vectors, the columns of* $\mathbf{V}$ *are right singular vectors, and the values* $s_1, \ldots, s_{\min\{m,n\}}$ *are the singular values.*

A third version of the singular value decomposition is often very helpful. Start with the reduced SVD $\mathbf{A} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\mathbf{V}^*$ in Theorem 3.1 Multiply $\widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}$ together to get

$$\mathbf{A} = \widehat{\mathbf{U}}\widehat{\mathbf{\Sigma}}\mathbf{V}^* = \left[\begin{array}{ccc} | & & | \\ s_1\mathbf{u}_1 & \cdots & s_n\mathbf{u}_n \\ | & & | \end{array}\right] \left[\begin{array}{ccc} — & \mathbf{v}_1^* & — \\ & \vdots & \\ — & \mathbf{v}_n^* & — \end{array}\right] = \sum_{j=1}^{n} s_j\mathbf{u}_j\mathbf{v}_j^*,$$

which renders $\mathbf{A}$ as the *sum of outer products* $\mathbf{u}_j\mathbf{v}_j^* \in \mathbb{C}^{m\times n}$, *weighted by nonnegative numbers* s_j. Let r denote the number of nonzero singular values, so that if $r < n$, then

$$s_{r+1} = \cdots = s_n = 0.$$

Thus $\mathbf{A}$ can be written as

$$\mathbf{A} = \sum_{j=1}^{r} s_j\mathbf{u}_j\mathbf{v}_j^*, \tag{3.1}$$

known as the *dyadic form of the SVD.*

The Dyadic Form of the Singular Value Decomposition

Theorem 3.3. *For any matrix* $\mathbf{A} \in \mathbb{C}^{m\times n}$, *there exists some* $r \in \{1,\ldots,n\}$ *such that*

$$\mathbf{A} = \sum_{j=1}^{r} s_j\mathbf{u}_j\mathbf{v}_j^*,$$

where $s_1 \geq s_2 \geq \cdots \geq s_r > 0$ *and* $\mathbf{u}_1,\ldots,\mathbf{u}_r \in \mathbb{C}^m$ *are orthonormal, and* $\mathbf{v}_1,\ldots,\mathbf{v}_r \in \mathbb{C}^n$ *are orthonormal.*

Corollary 3.4. *The rank of a matrix equals its number of nonzero singular values.*

The singular value decomposition also gives an immediate formula for the 2-norm of a matrix.

Theorem 3.5. *Let* $\mathbf{A} \in \mathbb{C}^{m\times n}$ *have singular value decomposition* (3.1). *Then*

$$\|\mathbf{A}\| = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = s_1.$$

Proof. The proof follows from the construction of the SVD at the start of this chapter. Note that

$$\|\mathbf{A}\|^2 = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|^2}{\|\mathbf{x}\|^2} = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^*\mathbf{A}^*\mathbf{A}\mathbf{x}}{\mathbf{x}^*\mathbf{x}},$$

and so $\|\mathbf{A}\|^2$ is the maximum Rayleigh quotient of the Hermitian matrix $\mathbf{A}^*\mathbf{A}$. By Theorem 2.2, this maximum value is the largest eigenvalue of

$\mathbf{A}^*\mathbf{A}$, i.e., $\lambda_1 = s_1^2$ in "Step 2" of the construction on page 92. Thus $\|\mathbf{A}\| = \sqrt{\lambda_1} = s_1$. ■

3.1.1 Inductive Proof of the SVD

3.1.2 The SVD and the Four Fundamental Subspaces

3.2 The SVD: Undoer of Many Knots

In the introduction to this chapter, we claimed that eigenvalue-based decompositions were the right tool for handling systems that involved *dynamics*. The SVD, in turn, is the perfect tool for handling *static* systems, i.e., systems that do not change with time. We identify three canonical static problems:

1. Find the unique solution $\mathbf{x}$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is an invertible square matrix.

2. Find the solution $\mathbf{x}$ of minimum norm that solves the *underdetermined* system $\mathbf{A}\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is a matrix with a nontrivial null space and $\mathbf{b} \in \mathcal{R}(\mathbf{A})$.

3. Find the minimum-norm vector $\mathbf{x}$ that minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$, for any given $\mathbf{b}$.

Notice that this last problem subsumes the first two. (If $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution $\mathbf{x}$, then $\|\mathbf{A}\mathbf{x} - \mathbf{b}\| = 0$ is minimal; if there are multiple solutions, one then finds the one having smallest norm.) Thus, our discussion will focus on problem 3.

Let $\mathbf{A} \in \mathbb{C}^{m\times n}$ have rank r and let $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{C}^m$ be a full set of *left* singular vectors, giving an orthonormal basis for $\mathbb{C}^m$ with

$$\begin{aligned} \mathcal{R}(\mathbf{A}) &= \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_r\} \\ \mathcal{N}(\mathbf{A}^*) &= \operatorname{span}\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}. \end{aligned}$$

Expand $\mathbf{b} \in \mathbb{C}^m$ as a linear combination of the left singular vectors:

$$\mathbf{b} = \beta_1\mathbf{u}_1 + \cdots + \beta_m\mathbf{u}_m.$$

We can find the coefficients β_j very easily, since the orthonormality of the singular vectors gives

$$\begin{aligned} \mathbf{u}_j^*\mathbf{b} &= \beta_1\mathbf{u}_j^*\mathbf{u}_1 + \cdots + \beta_m\mathbf{u}_j^*\mathbf{u}_m \\ &= \beta_j. \end{aligned}$$

Now for any $\mathbf{x} \in \mathbb{C}^n$, notice that $\mathbf{A}\mathbf{x} \in \mathcal{R}(\mathbf{A})$, and so

$$\begin{aligned}\mathbf{A}\mathbf{x} - \mathbf{b} &= \Big(\mathbf{A}\mathbf{x} - (\beta_1\mathbf{u}_1 + \cdots + \beta_r\mathbf{u}_r)\Big) - \Big(\beta_{r+1}\mathbf{u}_{r+1} + \cdots \beta_m\mathbf{u}_m\Big)\\ &= (\mathbf{A}\mathbf{x} - \mathbf{b}_R) - \mathbf{b}_N,\end{aligned}$$

where

$$\begin{aligned}\mathbf{b}_R &:= (\beta_1\mathbf{u}_1 + \cdots + \beta_r\mathbf{u}_r) \in \mathcal{R}(\mathbf{A})\\ \mathbf{b}_N &:= (\beta_{r+1}\mathbf{u}_{r+1} + \cdots \beta_m\mathbf{u}_m) \in \mathcal{N}(\mathbf{A}^*).\end{aligned}$$

Since the $\mathcal{R}(\mathbf{A}) \perp \mathcal{N}(\mathbf{A}^*)$ (by the Fundamental Theorem of Linear Algebra), $(\mathbf{A}\mathbf{x} - \mathbf{b}_R) \perp \mathbf{b}_N$, and hence, by the Pythagorean Theorem,

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \|(\mathbf{A}\mathbf{x} - \mathbf{b}_R) - \mathbf{b}_N\|^2 = \|\mathbf{A}\mathbf{x} - \mathbf{b}_R\|^2 + \|\mathbf{b}_N\|^2. \qquad (3.2)$$

Inspect this expression. The choice of $\mathbf{x}$ does not affect $\|\mathbf{b}_N\|^2$: $\mathbf{b}_N$ is the piece of $\mathbf{b}$ that is beyond the reach of $\mathbf{A}\mathbf{x}$. (If $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a solution, then $\mathbf{b}_N = \mathbf{0}$.) To minimize (3.2), the best we can do is find $\mathbf{x}$ such that $\|\mathbf{A}\mathbf{x} - \mathbf{b}_R\| = 0$. The dyadic form of the SVD,

$$\mathbf{A} = \sum_{j=1}^{r} s_j\mathbf{u}_j\mathbf{v}_j^*, \qquad (3.3)$$

makes quick work of this problem. Expand any $\mathbf{x} \in \mathbb{C}^n$ in the *right* singular vectors,

$$\mathbf{x} = \xi_1\mathbf{v}_1 + \cdots + \xi_n\mathbf{v}_n,$$

so that (via orthonormality of the singular vectors),

$$\mathbf{A}\mathbf{x} = \sum_{j=1}^{r} s_j\mathbf{u}_j\mathbf{v}_j^*\mathbf{x} = \sum_{j=1}^{r} s_j\xi_j\mathbf{u}_j.$$

We can equate this expression with

$$\begin{aligned}\mathbf{b}_R &= \beta_1\mathbf{u}_1 + \cdots + \beta_r\mathbf{u}_r\\ &= (\mathbf{u}_1^*\mathbf{b})\mathbf{u}_1 + \cdots + (\mathbf{u}_r^*\mathbf{b})\mathbf{u}_r\end{aligned} \qquad (3.4)$$

by simply matching the coefficients in the $\mathbf{u}_1, \ldots, \mathbf{u}_r$ directions:

$$s_k\xi_k = \mathbf{u}_k^*\mathbf{b}, \qquad k = 1, \ldots, r$$

giving

$$\xi_k = \frac{\mathbf{u}_k^*\mathbf{b}}{s_k}, \qquad k = 1, \ldots, r.$$

What about $\xi_{r+1}, \ldots, \xi_n$? *They can take any value!* To confirm this fact, define

$$\mathbf{x} = \sum_{k=1}^{r} \frac{\mathbf{u}_k^*\mathbf{b}}{s_k}\mathbf{v}_k + \sum_{k=r+1}^{n} \xi_k \mathbf{v}_k \tag{3.5}$$

for any $\xi_{r+1}, \ldots, \xi_n$, and verify that $\mathbf{Ax} = \mathbf{b}_R$:

$$\begin{aligned}
\mathbf{Ax} &= \left(\sum_{j=1}^{r} s_j \mathbf{u}_j \mathbf{v}_j^*\right)\left(\sum_{k=1}^{r} \frac{\mathbf{u}_k^*\mathbf{b}}{s_k}\mathbf{v}_k + \sum_{k=r+1}^{n} \xi_k \mathbf{v}_k\right)\\
&= \sum_{j=1}^{r}\sum_{k=1}^{r} s_j \frac{\mathbf{u}_k^*\mathbf{b}}{s_k}\mathbf{u}_j \mathbf{v}_j^* \mathbf{v}_k + \sum_{j=1}^{r}\sum_{k=r+1}^{n} s_j \xi_k \mathbf{u}_j \mathbf{v}_j^* \mathbf{v}_k\\
&= \sum_{j=1}^{r} (\mathbf{u}_j^*\mathbf{b})\mathbf{u}_j + 0\\
&= \mathbf{b}_R,
\end{aligned}$$

according to the expansion (3.4) for $\mathbf{b}_R$. Thus when $r < n$, equation (3.5) thus expresses the *infinitely many solutions* that minimize $\|\mathbf{Ab} - \mathbf{x}\|$. From all these solutions, we might naturally select the one that minimizes $\|\mathbf{x}\|$, the one that contains nothing extra. You might well suspect that this is the $\mathbf{x}$ we get from setting

$$\xi_{r+1} = \cdots = \xi_n = 0.$$

To confirm this fact, apply the Pythagorean theorem to (3.5) to get

$$\|\mathbf{x}\|^2 = \sum_{k=1}^{r} \frac{|\mathbf{u}_k^*\mathbf{b}|^2}{s_k^2} + \sum_{k=r+1}^{n} |\xi_k|^2,$$

making it obvious that the *unique* norm-minimizing solution is

$$\mathbf{x} = \sum_{j=1}^{r} \frac{\mathbf{u}_j^*\mathbf{b}}{s_j}\mathbf{v}_j = \left(\sum_{j=1}^{r} \frac{1}{s_j}\mathbf{v}_j \mathbf{u}_j^*\right)\mathbf{b}. \tag{3.6}$$

Take a moment to savor this beautiful formula, arguably one of the most important in matrix theory! In fact, the formula is so useful that the matrix on the right-hand side deserves some special nomenclature.

Pseudoinverse

Definition 3.6. *Let* $\mathbf{A} \in \mathbb{C}^{m\times n}$ *be a matrix of rank* r *with singular value decomposition (3.3). Then the pseudoinverse (or* MOORE–PENROSE *pseudoinverse) of* $\mathbf{A}$ *is given by*

$$\mathbf{A}^+ = \sum_{j=1}^{r} \frac{1}{s_j}\mathbf{v}_j\mathbf{u}_j^*. \tag{3.7}$$

We emphasize that $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ in (3.6) solves all three of the problems formulated at the beginning of this section.

1. If $\mathbf{A}$ is invertible, then $m = n = r$ and

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \left(\sum_{j=1}^{n} \frac{1}{s_j}\mathbf{v}_j\mathbf{u}_j^*\right)\mathbf{b}.$$

   In particular, whenever $\mathbf{A}$ is invertible, $\mathbf{A}^+ = \mathbf{A}^{-1}$.

2. If $\mathbf{b} \in \mathcal{R}(\mathbf{A})$, then $\mathbf{b}_R = \mathbf{b}$, and

$$\mathbf{x} = \mathbf{A}^{+}\mathbf{b} = \left(\sum_{j=1}^{r} \frac{1}{s_j}\mathbf{v}_j\mathbf{u}_j^*\right)\mathbf{b}$$

   solves $\mathbf{Ax} = \mathbf{b}$. If $r < n$, then there exist infinitely many solutions to $\mathbf{Ax} = \mathbf{b}$ that all have the form $\mathbf{x} = \mathbf{A}^+\mathbf{b} + \mathbf{n}$ for $\mathbf{n} \in \mathcal{N}(\mathbf{A}) = \text{span}\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$. Among all these solutions, $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ is the unique solution of smallest norm.

3. When $\mathbf{b} \notin \mathcal{R}(\mathbf{A})$, no $\mathbf{x}$ will solve $\mathbf{Ax} = \mathbf{b}$, but $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ will minimize $\|\mathbf{Ax} - \mathbf{b}\|$. If $r < n$, there will be infinitely many $\mathbf{x}$ that minimize $\|\mathbf{Ax} - \mathbf{b}\|$, each having the form $\mathbf{x} = \mathbf{A}^+\mathbf{b} + \mathbf{n}$ for $\mathbf{n} \in \mathcal{N}(\mathbf{A})$. Among all these solutions $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ is the unique minimizer of $\|\mathbf{Ax} - \mathbf{b}\|$ having smallest norm.

3.3 Optimal Low-Rank Approximations

Many applications call for use to approximate $\mathbf{A}$ with some low-rank matrix. Suppose we have the SVD

$$\mathbf{A} = \sum_{j=1}^{r} s_j\mathbf{u}_j\mathbf{v}_j^*, \tag{3.8}$$