3.5 SINGULAR VALUE DECOMPOSITION (OR SVD)

Eigenvalues and eigenvectors are defined only for square matrices, so it is natural to ask if extensions to rectangular matrices exist. The key to dealing with a rectangular $\mathbf{A}_{m\times n}$ is to revert to the analysis for square matrices by thinking in terms of $(\mathbf{A}^*\mathbf{A})_{n\times n}$ and $(\mathbf{A}\mathbf{A}^*)_{m\times m}$. Both are hermitian (or real and symmetric), and hence they are normal. And they have the following properties.

- $rank\,(\mathbf{A}^*\mathbf{A}) = rank\,(\mathbf{A}\mathbf{A}^*) = rank\,(\mathbf{A}) = r$ (Theorem 2.4.20, page 192).
- $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ have nonnegative eigenvalues (Exercise 3.1.12, page 297).
- $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ have the same positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$.
  - — The same holds for zero eigenvalues when $m = n$.
  - — If $m > n$, then $\mathbf{A}\mathbf{A}^*$ has $m - n$ more zero eigenvalues than $\mathbf{A}^*\mathbf{A}$.
  - — If $m < n$, then $\mathbf{A}^*\mathbf{A}$ has $n - m$ more zero eigenvalues than $\mathbf{A}\mathbf{A}^*$.

  (See Exercises 3.1.27 and 10.3.17, pages 299 and 824.)

These properties set the stage for the following definition of *singular values.*

3.5.1. Definition. For $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $rank\,(\mathbf{A}) = r$, let the positive eigenvalues of $\mathbf{A}^*\mathbf{A}$ (and $\mathbf{A}\mathbf{A}^*$) be $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. The r nonzero ***singular values*** of $\mathbf{A}$ are defined to be $\sigma_i = \sqrt{\lambda_i}$. When $r < p = \min\{m, n\}$, $\mathbf{A}$ is said to have $p - r$ zero singular values.

The goal is to use the singular values of a rectangular matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ in place of the eigenvalues of a square matrix to develop a decomposition similar (but not identical) to that in (3.4.1) on page 336. To do so, let $\mathbf{L}$ be the $r \times r$ diagonal matrix $\mathbf{L} = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_r)$ containing the positive eigenvalues of $\mathbf{A}^*\mathbf{A}$ (or $\mathbf{A}\mathbf{A}^*$). Theorem 3.4.2 (page 336) ensures that there are unitary matrices $\mathbf{V}_{n\times n} = [\underbrace{\mathbf{V}_1}_{n\times r} \mid \underbrace{\mathbf{V}_2}_{n\times(n-r)}]$ and $\mathbf{X}_{m\times m} = [\underbrace{\mathbf{X}_1}_{m\times r} \mid \underbrace{\mathbf{X}_2}_{m\times(m-r)}]$ such that

$$\begin{pmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n\times n} = \mathbf{V}^*\mathbf{A}^*\mathbf{A}\mathbf{V} = \left(\frac{\mathbf{V}_1^*}{\mathbf{V}_2^*}\right)\mathbf{A}^*\mathbf{A}\begin{bmatrix}\mathbf{V}_1 \mid \mathbf{V}_2\end{bmatrix} = \begin{pmatrix} \mathbf{V}_1^*\mathbf{A}^*\mathbf{A}\mathbf{V}_1 & \mathbf{V}_1^*\mathbf{A}^*\mathbf{A}\mathbf{V}_2 \\ \mathbf{V}_2^*\mathbf{A}^*\mathbf{A}\mathbf{V}_1 & \mathbf{V}_2^*\mathbf{A}^*\mathbf{A}\mathbf{V}_2 \end{pmatrix}$$

and (3.5.1)

$$\begin{pmatrix} \mathbf{L} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{m\times m} = \mathbf{X}^*\mathbf{A}\mathbf{A}^*\mathbf{X} = \left(\frac{\mathbf{X}_1^*}{\mathbf{X}_2^*}\right)\mathbf{A}\mathbf{A}^*\begin{bmatrix}\mathbf{X}_1 \mid \mathbf{X}_2\end{bmatrix} = \begin{pmatrix} \mathbf{X}_1^*\mathbf{A}\mathbf{A}^*\mathbf{X}_1 & \mathbf{X}_1^*\mathbf{A}\mathbf{A}^*\mathbf{X}_2 \\ \mathbf{X}_2^*\mathbf{A}\mathbf{A}^*\mathbf{X}_1 & \mathbf{X}_2^*\mathbf{A}\mathbf{A}^*\mathbf{X}_2 \end{pmatrix}.$$

Looking that the lower right-hand blocks and using the fact that $\mathbf{M}^*\mathbf{M} = \mathbf{0}$ if and only if $\mathbf{M} = \mathbf{0}$ (Exercise 2.4.14, page 194) yields

$$\mathbf{V}_2^*\mathbf{A}^*\mathbf{A}\mathbf{V}_2 = \mathbf{0} \implies \mathbf{A}\mathbf{V}_2 = \mathbf{0}, \tag{3.5.2}$$

$$\mathbf{X}_2^*\mathbf{A}\mathbf{A}^*\mathbf{X}_2 = \mathbf{0} \implies \mathbf{A}^*\mathbf{X}_2 = \mathbf{0}. \tag{3.5.3}$$

These observations lead to the following *singular value decomposition* theorem.

3.5.2. Theorem. For each $\mathbf{A} \in \mathbb{F}^{m\times n}$ whose positive singular values are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, there exist unitary matrices $\mathbf{U}_{m\times m}$ and $\mathbf{V}_{n\times n}$ such that

$$\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{m\times n}\mathbf{V}^*, \quad \text{where} \quad \mathbf{D} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r). \tag{3.5.4}$$

This is called a ***singular value decomposition*** † (or SVD) of $\mathbf{A}$.

Proof. Let $\mathbf{V}_{n\times n} = [\underbrace{\mathbf{V}_1}_{n\times r} \mid \underbrace{\mathbf{V}_2}_{n\times(n-r)}]$ and $\mathbf{X}_{m\times m} = [\underbrace{\mathbf{X}_1}_{m\times r} \mid \underbrace{\mathbf{X}_2}_{m\times(m-r)}]$ be unitary matrices that respectively diagonalize $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ as in (3.5.1). If

$$\mathbf{U} = \left[\mathbf{A}\mathbf{V}_1\mathbf{D}^{-1} \,|\, \mathbf{X}_2\right], \tag{3.5.5}$$

then $\mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^* = \left[\, \mathbf{A}\mathbf{V}_1\mathbf{D}^{-1} \mid \mathbf{X}_2 \,\right]\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{V}_1^* \\ \hline \mathbf{V}_2^* \end{pmatrix} = \mathbf{A}\mathbf{V}_1\mathbf{V}_1^* = \mathbf{A}$. The last equality here follows from (3.5.2) because

$$\mathbf{A} = \mathbf{A}\mathbf{I} = \mathbf{A}\mathbf{V}\mathbf{V}^* = \mathbf{A}(\mathbf{V}_1\mathbf{V}_1^* + \mathbf{V}_2\mathbf{V}_2^*) = \mathbf{A}\mathbf{V}_1\mathbf{V}_1^*.$$

To see that $\mathbf{U}$ is unitary, simply verify that $\mathbf{U}^*\mathbf{U} = \mathbf{I}$ by using $\mathbf{L} = \mathbf{V}_1^*\mathbf{A}^*\mathbf{A}\mathbf{V}_1$ from (3.5.1) along with (3.5.3) and the fact that $\mathbf{X}_2$ has orthonormal columns to write

$$\mathbf{U}^*\mathbf{U} = \begin{pmatrix} \mathbf{D}^{-1}\mathbf{V}_1^*\mathbf{A}^*\mathbf{A}\mathbf{V}_1\mathbf{D}^{-1} & \mathbf{D}^{-1}\mathbf{V}_1^*\mathbf{A}^*\mathbf{X}_2 \\ \mathbf{X}_2^*\mathbf{A}\mathbf{V}_1\mathbf{D}^{-1} & \mathbf{X}_2^*\mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}. \quad \blacksquare$$

Note: The simplicity of proof of the SVD belies the importance of its role in both the theory and applications of linear algebra. While there are several important matrix decompositions, none finds their way into as many different facets of applied mathematics and science as does the SVD. This will become evident as more of the story unfolds with multiple revisitations to the SVD.

Example

Perform a singular value decomposition of $\mathbf{A} = \begin{pmatrix} -13/5 & -2/15 \\ -14/5 & -31/15 \\ -2/5 & -58/15 \end{pmatrix}$ by following the logic in the proof of Theorem 3.5.2. The first step is to diagonalize

$$\mathbf{A}^T\mathbf{A} = \frac{1}{25}\begin{pmatrix} 369 & 192 \\ 192 & 481 \end{pmatrix} \quad \text{and} \quad \mathbf{A}\mathbf{A}^T = \frac{1}{9}\begin{pmatrix} 61 & 68 & 14 \\ 68 & 109 & 82 \\ 14 & 82 & 136 \end{pmatrix}$$

† The SVD has been independently discovered and rediscovered several times. Those credited with the early developments include Eugenio Beltrami (1835–1899) in 1873; M. E. Camille Jordan (1838–1922) in 1875; James J. Sylvester (1814–1897) in 1889; L. Autonne in 1913; and C. Eckart and G. Young in 1936.

with respective orthogonal matrices $\mathbf{V}$ and $\mathbf{X}$. Computing eigenvalues reveals that $\sigma\left(\mathbf{A}^T\mathbf{A}\right) = \{\lambda_1 = 25, \lambda_2 = 9\}$, $\sigma\left(\mathbf{A}\mathbf{A}^T\right) = \{\lambda_1 = 25, \lambda_3 = 9, \lambda_3 = 0\}$, so $r = rank\,(\mathbf{A}) = 2$, and the nonzero singular values for $\mathbf{A}$ are $\sigma_1 = \sqrt{\lambda_1} = 5$ and $\sigma_2 = \sqrt{\lambda_2} = 3$. Thus the matrix $\mathbf{D}$ in (3.5.4) is $\mathbf{D} = \begin{pmatrix} 5 & 0 \\ 0 & 3 \end{pmatrix}$. Compute eigenvectors of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ to be

$$N\left(\mathbf{A}^T\mathbf{A} - 25\mathbf{I}\right) = span\left\{\begin{pmatrix} 3 \\ 4 \end{pmatrix}\right\}, \quad N\left(\mathbf{A}^T\mathbf{A} - 9\mathbf{I}\right) = span\left\{\begin{pmatrix} -4 \\ 3 \end{pmatrix}\right\},$$
$$N\left(\mathbf{A}\mathbf{A}^T - 25\mathbf{I}\right) = span\left\{\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right\}, \quad N\left(\mathbf{A}\mathbf{A}^T - 9\mathbf{I}\right) = span\left\{\begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}\right\},$$
$$N\left(\mathbf{A}\mathbf{A}^T - 0\mathbf{I}\right) = N\left(\mathbf{A}\mathbf{A}^T\right) = span\left\{\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}\right\}.$$

Normalizing the eigenvectors produces the orthogonal matrices

$$\mathbf{V} = \frac{1}{5}\begin{pmatrix} 3 & -4 \\ 4 & 3 \end{pmatrix}, \quad \mathbf{X} = \frac{1}{3}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix} \implies \mathbf{V}_1 = \mathbf{V}, \quad \mathbf{X}_2 = \frac{1}{3}\begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}.$$

The unitary matrix in (3.5.5) is $\mathbf{U} = \left[\mathbf{A}\mathbf{V}_1\mathbf{D}^{-1} \,|\, \mathbf{X}_2\right] = \frac{1}{3}\begin{pmatrix} -1 & 2 & 2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{pmatrix}$, so the resulting SVD is

$$\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} \\ \mathbf{0} \end{pmatrix}\mathbf{V}^* = \begin{pmatrix} -1/3 & 2/3 & 2/3 \\ -2/3 & 1/3 & -2/3 \\ -2/3 & -2/3 & 1/3 \end{pmatrix}\begin{pmatrix} 5 & 0 \\ 0 & 3 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}.$$

Note: The computations in this example are meant to illustrate the proof of Theorem 3.5.2, but this is not the way that singular values are computed in practice. People's careers have revolved around creating practical SVD algorithms, and a significant degree of numerical sophistication is required before being able to understand and appreciate the nature of some of their methods.

Singular Vectors

In a loose sense the singular values for a rectangular matrix play a role similar to eigenvalues for a square matrix. So, are there vectors for rectangular matrices that in some way resemble eigenvectors for square matrices? Well, kind of—if you are willing to stretch things a bit. If $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ is an SVD, and if $\mathbf{v}_j$ and $\mathbf{u}_j$, $1 \le j \le r$, are respective columns in $\mathbf{V}$ and $\mathbf{U}$, then

$$\mathbf{A}\mathbf{V} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Longrightarrow \mathbf{A}\mathbf{v}_j = \sigma_j\mathbf{u}_j \quad \text{and} \quad \mathbf{A}^*\mathbf{U} = \mathbf{V}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Longrightarrow \mathbf{A}^*\mathbf{u}_j = \sigma_j\mathbf{v}_j.$$

For this reason the following terminology is adopted.

3.5.3. Definition. The respective columns $\mathbf{v}_j$ and $\mathbf{u}_j$ in unitary matrices $\mathbf{V}$ and $\mathbf{U}$ from a singular value decomposition $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ are called ***right-hand singular*** vectors and ***left-hand singular vectors*** for $\mathbf{A}$ because

$$\begin{aligned} &\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j \quad \text{and} \quad \mathbf{A}^*\mathbf{u}_j = \sigma_j \mathbf{v}_j \quad \text{for } 1 \le j \le r, \\ &\mathbf{A}\mathbf{v}_j = \mathbf{0} \qquad \text{and} \quad \mathbf{A}^*\mathbf{u}_j = \mathbf{0} \qquad \text{for } j > r. \end{aligned} \tag{3.5.6}$$

Notice that the singular vectors are also legitimate eigenvectors—not for $\mathbf{A}$ but rather for $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ because

$$\mathbf{A}^*\mathbf{A} = \mathbf{V}\begin{pmatrix} \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{n-r} \end{pmatrix}\mathbf{V}^* \quad \text{and} \quad \mathbf{A}\mathbf{A}^* = \mathbf{U}\begin{pmatrix} \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{m-r} \end{pmatrix}\mathbf{U}^*$$

is equivalent to saying that

$$\begin{aligned} &\mathbf{A}^*\mathbf{A}\mathbf{v}_j = \sigma_j^2 \mathbf{v}_j \quad \text{and} \quad \mathbf{A}\mathbf{A}^*\mathbf{u}_j = \sigma_j^2 \mathbf{u}_j \quad \text{for } 1 \le j \le r, \\ &\mathbf{A}^*\mathbf{A}\mathbf{v}_j = \mathbf{0} \qquad \text{and} \quad \mathbf{A}\mathbf{A}^*\mathbf{u}_j = \mathbf{0} \qquad \text{for } j > r. \end{aligned} \tag{3.5.7}$$

Caution! While the right-hand and left-hand singular vectors for a rectangular matrix $\mathbf{A}$ are respective orthonormal eigenvectors for $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$, *the converse is not true*! That is, *respective orthonormal eigenvectors for* $\mathbf{A}^*\mathbf{A}$ *and* $\mathbf{A}\mathbf{A}^*$ *are not necessarily singular vectors for* $\mathbf{A}$. It is a misconception that simply finding any set of orthonormal eigenvectors for $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ is sufficient to construct a $\mathbf{U}$ and $\mathbf{V}$ that will produce an SVD of $\mathbf{A}$, but life is not this easy. Taking $\mathbf{A} = \mathbf{I}$ is a particularly simple example that illustrates this. The columns of any unitary matrix $\mathbf{U}$ will provide orthonormal eigenvectors for $\mathbf{A}\mathbf{A}^* = \mathbf{I}$, and likewise, any other unitary matrix $\mathbf{V}$ will contain orthonormal eigenvectors for $\mathbf{A}^*\mathbf{A} = \mathbf{I}$. All singular values are 1, so $\mathbf{D} = \mathbf{I}$, and the only way that $\mathbf{U}\mathbf{D}\mathbf{V}^*$ can be an SVD for $\mathbf{I}$ (i.e., $\mathbf{U}\mathbf{I}\mathbf{V}^* = \mathbf{I}$) is if $\mathbf{U} = \mathbf{V}$ Thus $\mathbf{U}$ cannot be taken to be just any orthonormal set eigenvectors of $\mathbf{A}\mathbf{A}^*$.

However, there is a duality in the following sense. If $\mathbf{V}$ is an arbitrary orthonormal set of eigenvectors for $\mathbf{A}^*\mathbf{A}$, then $\mathbf{U}$ is fixed by (3.5.5), but if $\mathbf{U}$ is taken as an arbitrary orthonormal set of eigenvectors for $\mathbf{A}\mathbf{A}^*$, then $\mathbf{V}$ is fixed by $\mathbf{V} = \left[\mathbf{A}^*\mathbf{U}_1\mathbf{D}^{-1} \,|\, \mathbf{Y}_2\right]$, where $\mathbf{Y} = [\mathbf{Y}_1 \,|\, \mathbf{Y}_2]$ is a unitary matrix that diagonalizes $\mathbf{A}^*\mathbf{A}$ (see Exercise 3.5.9).

Degree of Uniqueness

The nonzero singular values of $\mathbf{A} \in \mathbb{F}^{m\times n}$ are uniquely defined by $\mathbf{A}$ because they are the positive square roots of the nonzero eigenvalues of $\mathbf{A}^*\mathbf{A}$ (and $\mathbf{A}\mathbf{A}^*$), but what about the singular vectors—to what degree are they uniquely

defined by $\mathbf{A}$? The preceding discussion in which $\mathbf{A} = \mathbf{I}$ shows that singular vectors are not in any way unique because if $\mathbf{U}$ is any unitary matrix, then $\mathbf{UIU}^* = \mathbf{I}$ is a legitimate SVD for $\mathbf{I}$. In other words *an SVD for a given matrix need not be unique.*

However, something a little more specific can be said in the special case of singular vectors that correspond to distinct singular values.

3.5.4. Theorem. If $\mathbf{v}$ and $\widetilde{\mathbf{v}}$ are right-hand singular vectors corresponding to the same non-repeated singular value for $\mathbf{A} \in \mathbb{F}^{m\times n}$, then $\widetilde{\mathbf{v}} = e^{i\theta}\mathbf{v}$ for some θ, and similarly for left-hand singular vectors. In other words, singular vectors corresponding to non-repeated singular values are unique up to a scalar multiple $e^{i\theta}$.

Proof. If $(\sigma, \mathbf{v})$ is a right-hand singular-value pair for $\mathbf{A}$, then it follows from (3.5.6) that $(\sigma^2, \mathbf{v})$ is an eigenpair for $\mathbf{A}^*\mathbf{A}$. If σ is not repeated, then $\textit{alg mult}_{\mathbf{A}^*\mathbf{A}}(\sigma^2) = 1 = \textit{geo mult}_{\mathbf{A}^*\mathbf{A}}(\sigma^2)$ (Theorem 3.2.9, page 306), so if $(\sigma, \widetilde{\mathbf{v}})$ is any other right-hand singular-value pair, then $\widetilde{\mathbf{v}} = \alpha\mathbf{v}$ for some $\alpha \in \mathbb{F}$. Since $\|\widetilde{\mathbf{v}}\|_2 = 1$, it follows that $|\alpha| = 1$, and hence $\widetilde{\mathbf{v}} = e^{i\theta}\mathbf{v}$ for some θ. The logic is similar for left-hand singular-value pairs. ∎

Extremal Singular Values

If $\mathbf{A} \in \mathbb{F}^{n\times n}$ is hermitian with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, then Theorem 3.4.8 (page 342) characterizes the largest and smallest eigenvalues of $\mathbf{A}$ as

$$\lambda_1 = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^*\mathbf{Ax} \quad \text{and} \quad \lambda_n = \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^*\mathbf{Ax}. \tag{3.5.8}$$

This in turn provides expressions for extremal singular values because singular values of $\mathbf{A}$ are simply square roots of eigenvalues of $\mathbf{A}^*\mathbf{A}$. In particular, if σ_1 is the largest singular value for $\mathbf{A}$, then the first part of (3.5.8) ensures that $\sigma_1^2 = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^*\mathbf{A}^*\mathbf{Ax} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2^2$, or equivalently,

$$\sigma_1 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 . \tag{3.5.9}$$

If $\mathbf{A}_{n\times n}$ is nonsingular, then its smallest singular value is $\sigma_n > 0$, so the second part of (3.5.8) implies that $\sigma_n^2 = \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^*\mathbf{A}^*\mathbf{Ax} = \min_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2^2$, or equivalently,

$$\sigma_n = \min_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 . \tag{3.5.10}$$

Matrix 2-Norm

The characterizations (3.5.9) and (3.5.10) reveal important relationships between the euclidean vector 2-norm and singular values that unlock the mystery surrounding the nature of the matrix 2-norm that is induced by the standard vector 2-norm. For $\mathbf{A} \in \mathbb{F}^{m\times n}$ and $\mathbf{x} \in \mathbb{F}^{n\times 1}$, Theorem 1.9.5 on page 84 establishes the fact that every vector norm induces a compatible matrix norm via

$$\|\mathbf{A}\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| \quad \text{and} \quad \left\|\mathbf{A}^{-1}\right\| = \frac{1}{\min_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|} \quad \text{when } \mathbf{A}^{-1} \text{ exists.} \tag{3.5.11}$$

The induced matrix 1-norm and ∞-norm were developed on page 86, but the formulation of the induced matrix 2-norm had to wait until (3.5.9) and (3.5.10) could be developed to complete the picture. Respectively combining these expressions with those in (3.5.11) produces the following theorem.

3.5.5. Theorem. The ***matrix 2-norm*** (some called it the *spectral norm*) that is induced by the vector euclidean 2-norm is

$$\|\mathbf{A}\|_2 = \sigma_1 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 \quad \text{for } \mathbf{A} \in \mathbb{F}^{m\times n},\ \mathbf{x} \in \mathbb{F}^{n\times 1}, \tag{3.5.12}$$

where σ_1 is the largest singular value of $\mathbf{A}$. Compatibility with the euclidean vector norm (i.e., $\|\mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$) is guaranteed by Theorem 1.9.5. Moreover, if $\mathbf{A}_{n\times n}$ is nonsingular with singular values $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n > 0$, then

$$\left\|\mathbf{A}^{-1}\right\|_2 = \frac{1}{\sigma_n} = \frac{1}{\min_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2}. \tag{3.5.13}$$

The figure below updates the illustration on page 85 to reflect the geometric interpretation of the matrix 2-norm in $\mathbb{R}^3$ in terms of singular values.

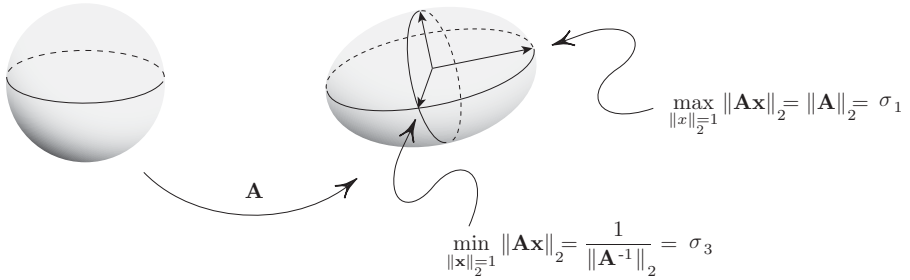

Figure 3.5.1: Induced Matrix 2-norm in $\mathbb{R}^3$.

Example

Let $\mathbf{A} = \begin{pmatrix} -13/5 & -2/15 \\ -14/5 & -31/15 \\ -2/5 & -58/15 \end{pmatrix}$ be the matrix from the example on page 356 where it was determined there that the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$ is $\lambda_{\max}(\mathbf{A}^T\mathbf{A}) = 25$. Consequently,

$$\|\mathbf{A}\|_2 = \sigma_1 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})} = \sqrt{25} = 5.$$

Alternately, $\lambda_{\max}(\mathbf{A}\mathbf{A}^T) = 25$ produces the same result.

Properties of the Matrix 2-Norm

Needless to say, computing $\|\mathbf{A}\|_2$ is generally more involved than computing $\|\mathbf{A}\|_\infty$ (the largest absolute row sum), or $\|\mathbf{A}\|_1$ (the largest absolute column sum), or even $\|\mathbf{A}\|_F$ (the square root of the sum of squares), but the numerous theoretical properties of the matrix 2-norm beyond those of a general matrix norm compensate for the difficulty in computing it.

3.5.6. Theorem. For every $\mathbf{A} \in \mathbb{F}^{m\times n}$, the following statements hold.

- $\|\mathbf{A}\|_2 = \|\mathbf{A}^*\|_2$. (3.5.14)
- $\|\mathbf{A}^*\mathbf{A}\|_2 = \|\mathbf{A}\|_2^2 = \|\mathbf{A}\mathbf{A}^*\|_2$. (3.5.15)
- $\left\| \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix} \right\|_2 = \max\left\{ \|\mathbf{A}\|_2, \|\mathbf{B}\|_2 \right\}$. (3.5.16)
- $\|\mathbf{P}_{p\times m}\mathbf{A}\mathbf{Q}_{n\times q}\|_2 = \|\mathbf{A}\|_2$ when $\mathbf{P}^*\mathbf{P} = \mathbf{I}_m$, $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}_q$. (3.5.17)
- If $\mathbf{B}$ is any submatrix of $\mathbf{A}$, then $\|\mathbf{B}\|_2 \leq \|\mathbf{A}\|_2$. (3.5.18)
- $\|\mathbf{A}\|_2 = \max\limits_{\|\mathbf{x}\|_2=1} \max\limits_{\|\mathbf{y}\|_2=1} |\mathbf{y}^*\mathbf{A}\mathbf{x}|$. (3.5.19)

Proof of (3.5.14). This is true because $\mathbf{A}^*\mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ have the same nonzero eigenvalues, so $\mathbf{A}$ and $\mathbf{A}^*$ have the same singular values.

Proof of (3.5.15). The largest eigenvalue of $(\mathbf{A}^*\mathbf{A})^*(\mathbf{A}^*\mathbf{A}) = (\mathbf{A}^*\mathbf{A})^2$ is the square of the largest eigenvalue of $\mathbf{A}^*\mathbf{A}$ (Exercise 3.1.15, page 298), so the largest singular value of $\mathbf{A}^*\mathbf{A}$ is the square of the largest singular value of $\mathbf{A}$.

Proof of (3.5.16). Let $\mathbf{C} = \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}$ so that $\mathbf{C}^*\mathbf{C} = \begin{pmatrix} \mathbf{A}^*\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^*\mathbf{B} \end{pmatrix}$. It follows from Exercise 3.1.9 on page 297 that $\sigma(\mathbf{C}^*\mathbf{C}) = \sigma(\mathbf{A}^*\mathbf{A}) \cup \sigma(\mathbf{B}^*\mathbf{B})$, and thus

$$\|\mathbf{C}\|_2^2 = \lambda_{\max}(\mathbf{C}^*\mathbf{C}) = \max\{\lambda_{\max}(\mathbf{A}^*\mathbf{A}), \lambda_{\max}(\mathbf{B}^*\mathbf{B})\} = \max\{\|\mathbf{A}\|_2^2, \|\mathbf{B}\|_2^2\}.$$

Proof of (3.5.17). If $\mathbf{P}^*\mathbf{P} = \mathbf{I}_m$, then

$$\|\mathbf{P}\mathbf{A}\mathbf{Q}\mathbf{x}\|_2^2 = \mathbf{x}^*\mathbf{Q}^*\mathbf{A}^*\mathbf{P}^*\mathbf{P}\mathbf{A}\mathbf{Q}\mathbf{x} = \mathbf{x}^*\mathbf{Q}^*\mathbf{A}^*\mathbf{A}\mathbf{Q}\mathbf{x} = \mathbf{y}^*\mathbf{A}^*\mathbf{A}\mathbf{y},$$

where $\mathbf{y} = \mathbf{Qx}$. Now, $\mathbf{Q}^*\mathbf{Q} = \mathbf{I}_q$ ensures that $\|\mathbf{x}\|_2 = 1 \iff \|\mathbf{y}\|_2 = 1$, so

$$\max_{\|\mathbf{x}\|_2=1} \|\mathbf{PAQx}\|_2^2 = \max_{\|\mathbf{y}\|_2=1} \|\mathbf{Ay}\|_2^2 = \|\mathbf{A}\|_2^2 .$$

Proof of (3.5.18). Let $\mathbf{P}$ and $\mathbf{Q}$ be permutation matrices (which are necessarily unitary) such that $\mathbf{PAQ} = \begin{pmatrix} \mathbf{B} & \star \\ \star & \star \end{pmatrix}$. Consequently, $\|\mathbf{A}\|_2 = \left\| \begin{pmatrix} \mathbf{B} & \star \\ \star & \star \end{pmatrix} \right\|_2$. Use this along with $\left\| \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right\|_2 = 1$ to write

$$\begin{aligned} \|\mathbf{B}\|_2 = \left\| \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right\|_2 &= \left\| \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{B} & \star \\ \star & \star \end{pmatrix} \begin{pmatrix} \mathbf{I}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right\|_2 \\ &\le \left\| \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right\|_2 \left\| \begin{pmatrix} \mathbf{B} & \star \\ \star & \star \end{pmatrix} \right\|_2 \left\| \begin{pmatrix} \mathbf{I}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right\|_2 = \|\mathbf{A}\|_2 . \end{aligned}$$

Proof of (3.5.19). Applying the Cauchy–Schwarz (CBS) inequality (Theorem 1.4.5, page 27) yields

$$|\mathbf{y}^*\mathbf{Ax}| \le \|\mathbf{y}\|_2 \|\mathbf{Ax}\|_2 \implies \max_{\substack{\|\mathbf{x}\|_2=1 \\ \|\mathbf{y}\|_2=1}} |\mathbf{y}^*\mathbf{Ax}| \le \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2 = \|\mathbf{A}\|_2 .$$

Equality is attained when $\mathbf{x}$ and $\mathbf{y}$ are respective right-hand and left-hand singular vectors associated with the largest singular value σ_1. ∎

Distance to Lower-Rank Matrices

There are a variety of ways to interpret the meaning of a singular value, but one of the most important is that of the next theorem that shows σ_{k+1} is the 2-norm distance between $\mathbf{A}_{m\times n}$ with $rank\,(\mathbf{A}) = r$ and a closest matrix $\mathbf{B}_{m\times n}$ having $rank\,(\mathbf{B}) = k < r$.

3.5.7. Theorem. Let $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$ be the nonzero singular values of $\mathbf{A} \in \mathbb{F}^{m\times n}$. For each $k < r$, the 2-norm distance from $\mathbf{A}$ to a closest $m \times n$ matrix of rank k is

$$\min_{rank(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_2 = \sigma_{k+1}. \tag{3.5.20}$$

The distance to a nearest rank-k matrix in the Frobenius norm is

$$\min_{rank(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_F = \left[\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_r^2\right]^{1/2}. \tag{3.5.21}$$

Truncating an SVD $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ to become $\mathbf{B} = \mathbf{U}\begin{pmatrix} \mathbf{D}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ produces a closest matrix rank-k matrix in both norms.

Proof of (3.5.20). Let $\mathbf{A}=\mathbf{U}\begin{pmatrix}\mathbf{D}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^*$ be an SVD, and let $rank\,(\mathbf{B}_{m\times n})=k$. If $\mathbf{U}_{m\times m}=[\underbrace{\mathbf{U}_1}_{m\times(k+1)}\,|\,\mathbf{U}_2]$ and $\mathbf{V}_{n\times n}=[\underbrace{\mathbf{V}_1}_{n\times(k+1)}\,|\,\mathbf{V}_2]$, then $\mathbf{BV}_1$ is $m\times(k+1)$. Since $rank\,(\mathbf{BV}_1)\le rank\,(\mathbf{B})=k$, $(\mathbf{BV}_1)\mathbf{x}=\mathbf{0}$ for some $\mathbf{x}\neq\mathbf{0}$, which can be assumed to be normalized so that $\|\mathbf{x}\|_2=1$. For $\mathbf{D}_{k+1}=\text{diag}\,(\sigma_1,\sigma_2,\ldots,\sigma_{k+1})$,

$$\sigma_{k+1}=\min_{\|z\|_2=1}\|\mathbf{D}_{k+1}\mathbf{z}\|_2\le\|\mathbf{D}_{k+1}\mathbf{x}\|_2\le\|\mathbf{U}_1^*(\mathbf{A}-\mathbf{B})\mathbf{V}_1\mathbf{x}\|_2\le\|\mathbf{A}-\mathbf{B}\|_2\,.$$

Equality in the last expression is attained for $\mathbf{B}=\mathbf{U}\begin{pmatrix}\mathbf{D}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^*$. ∎

Proof of (3.5.21). For any rank k matrix $\mathbf{B}$, set $\mathbf{C}=\mathbf{U}^*\mathbf{BV}$ so that

$$\begin{aligned}\|\mathbf{A}-\mathbf{B}\|_F^2&=\|\mathbf{U}^*(\mathbf{A}-\mathbf{B})\mathbf{V}\|_F^2=\left\|\begin{pmatrix}\mathbf{D}_r & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}-\mathbf{C}\right\|_F^2\\&=\sum_{i=1}^{r}|\sigma_i-c_{ii}|^2+\sum_{i>r}|c_{ii}|^2+\sum_{i\neq j}|c_{ij}|^2\ge\sum_{i=1}^{k}|\sigma_i-c_{ii}|^2+\sum_{i>r}|c_{ii}|^2.\end{aligned}\tag{3.5.22}$$

Consequently, a rank k matrix $\mathbf{B}$ that minimizes $\|\mathbf{A}-\mathbf{B}\|_F^2$ must be such that $c_{ij}=0$ for $i\neq j$. Among rank k matrices $\mathbf{C}$ with zero off-diagonal entries, the one that yields a minimum in (3.5.22) has $c_{ii}=\begin{cases}\sigma_i & \text{for } i\le k,\\ 0 & \text{for } i>r.\end{cases}$ Thus the minimum is attained at $\mathbf{B}=\mathbf{U}\begin{pmatrix}\mathbf{D}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^*$, and $\min_{rank(\mathbf{B})=k}\|\mathbf{A}-\mathbf{B}\|_F^2=\sum_{i=k+1}^{r}\sigma_i^2$. ∎

Distance to Singularity

A special case of Theorem 3.5.7 reveals how close each nonsingular matrix is to being singular. It follows from (3.5.20) that $\min_{rank(\mathbf{B})=k<n}\|\mathbf{A}-\mathbf{B}\|_2=\sigma_{k+1}$, and this is smallest when $k+1=n$. Thus the following corollary is produced.

3.5.8. Corollary. If $\mathbf{A}_{n\times n}$ is nonsingular, then the 2-norm distance to a closest singular matrix is the smallest singular value $\sigma_n=1/\|\mathbf{A}^{-1}\|_2$.

Effects of Small Perturbations on Rank

Another important consequence of Theorem 3.5.7 concerns how small perturbations to $\mathbf{A}\in\mathbb{F}^{m\times n}$ can affect its rank. If $rank\,(\mathbf{A})=r<\min\{m,n\}$, then intuition might suggest that for relatively small perturbations $\mathbf{E}$, having $rank\,(\mathbf{A}+\mathbf{E})\ge r$ is just as plausible as $rank\,(\mathbf{A}+\mathbf{E})<r$. But not true!

3.5.9. Theorem. Suppose that $\mathbf{A} \in \mathbb{F}^{m\times n}$ has nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, where $r = rank\,(\mathbf{A})$. If $\|\mathbf{E}\|_2 < \sigma_r$, then $rank\,(\mathbf{A}+\mathbf{E}) \geq rank\,(\mathbf{A})$.

- In other words, *relatively small perturbations cannot decrease rank.*

Proof. Suppose to the contrary that $rank\,(\mathbf{A}+\mathbf{E}) = k < r$. Theorem 3.5.7 then implies that

$$\|\mathbf{E}\|_2 = \|\mathbf{A} - (\mathbf{A}+\mathbf{E})\|_2 \geq \min_{rank(\mathbf{B})=k} \|\mathbf{A}-\mathbf{B}\|_2 = \sigma_{k+1} \geq \sigma_r,$$

which contradicts $\|\mathbf{E}\|_2 < \sigma_r$. Thus $rank\,(\mathbf{A}+\mathbf{E}) \geq r = rank\,(\mathbf{A})$. ∎

Likelihood of a Drop in Rank

If $rank\,(\mathbf{A}_{m\times n}) = r < \min\{m, n\}$, then Theorem 3.5.9 guarantees that perturbations of even the slightest magnitude *can* increase the rank. But what is the likelihood that an increase in rank will actually occur? To answer this question, let $\mathbf{A} = \mathbf{U}\begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^*$ be an SVD of $\mathbf{A}$ in which $\mathbf{D} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$ are the nonzero singular values, and suppose that $\|\mathbf{E}\|_2 < \sigma_r$. Apply $\mathbf{U}^*$ and $\mathbf{V}$ to $\mathbf{E}$ to produce

$$\mathbf{U}^*\mathbf{E}\mathbf{V} = \begin{pmatrix}\mathbf{E}_{11} & \mathbf{E}_{12}\\ \mathbf{E}_{21} & \mathbf{E}_{22}\end{pmatrix}, \quad \text{or equivalently,} \quad \mathbf{E} = \mathbf{U}\begin{pmatrix}\mathbf{E}_{11} & \mathbf{E}_{12}\\ \mathbf{E}_{21} & \mathbf{E}_{22}\end{pmatrix}\mathbf{V}^*,$$

where $\mathbf{E}_{11}$ is $r\times r$ so that $\mathbf{A}+\mathbf{E} = \mathbf{U}\begin{pmatrix}\mathbf{D}+\mathbf{E}_{11} & \mathbf{E}_{12}\\ \mathbf{E}_{21} & \mathbf{E}_{22}\end{pmatrix}\mathbf{V}^*$. Since $\|\mathbf{E}\|_2 < \sigma_r$, it follows from (3.5.18) on page 361 that $\|\mathbf{E}_{11}\|_2 < \sigma_r$, and hence

$$\left\|\mathbf{D}^{-1}\mathbf{E}_{11}\right\|_2 \leq \left\|\mathbf{D}^{-1}\right\|_2 \|\mathbf{E}_{11}\|_2 < \sigma_r^{-1}\sigma_r = 1.$$

Consequently, $\mathbf{I} + \mathbf{D}^{-1}\mathbf{E}_{11}$ is nonsingular (by Theorem 2.3.11 on page 168), which in turn forces $\mathbf{D}+\mathbf{E}_{11} = \mathbf{D}(\mathbf{I}+\mathbf{D}^{-1}\mathbf{E}_{11})$ to be nonsingular. It now follows that

$$\begin{aligned} rank\,(\mathbf{A}+\mathbf{E}) &= rank\,\mathbf{U}\begin{pmatrix}\mathbf{D}+\mathbf{E}_{11} & \mathbf{E}_{12}\\ \mathbf{E}_{21} & \mathbf{E}_{22}\end{pmatrix}\mathbf{V}^* = rank\begin{pmatrix}\mathbf{D}+\mathbf{E}_{11} & \mathbf{E}_{12}\\ \mathbf{E}_{21} & \mathbf{E}_{22}\end{pmatrix}\\ &= rank\,(\mathbf{D}+\mathbf{E}_{11}) + rank\,(\mathbf{S}) = rank\,(\mathbf{A}) + rank\,(\mathbf{S}) \end{aligned}$$

in which $\mathbf{S}$ is the Schur complement

$$\mathbf{S} = \mathbf{E}_{22} - \mathbf{E}_{21}\left(\mathbf{D}+\mathbf{E}_{11}\right)^{-1}\mathbf{E}_{12}. \tag{3.5.23}$$

(Recall Exercise 2.4.26 on page 195.) In order for $rank\,(\mathbf{A}+\mathbf{E})$ to not jump, the perturbation $\mathbf{E}$ must be such that $\mathbf{S}=\mathbf{0}$, which is equivalent to saying $\mathbf{E}_{22}=\mathbf{E}_{21}\left(\mathbf{D}+\mathbf{E}_{11}\right)^{-1}\mathbf{E}_{12}$. This clearly requires the existence of a very specific (and quite special) relationship among the entries of $\mathbf{E}$, and it is extremely unlikely that a random perturbation will produce such a relationship. This is the logic behind following hueristic.

- *The rank of randomly perturbed matrix will almost surely increase if it is not already of maximal rank.* (3.5.24)

Consequences for Linear Systems

Theorem 3.5.9 and the realization in (3.5.24) have important implications for computing solutions of linear systems of equations $\mathbf{Ax}=\mathbf{b}$. When floating-point arithmetic is used (which it almost always is for real-world problems), the computed solution is the exact solution of a different system whose coefficient matrix is $\mathbf{A}+\mathbf{E}$.

Suppose first that $\mathbf{A}$ is nonsingular, and suppose that a stable algorithm is used to solve $\mathbf{Ax}=\mathbf{b}$. Recall from page 236 that being "stable" means that the algorithm returns the exact solution of a nearby system—i.e., the computed solution is the exact solution of $(\mathbf{A}+\mathbf{E})\mathbf{x}=\mathbf{b}$, where the entries in $\mathbf{E}$ have relatively small magnitudes. In particular, if $\|\mathbf{E}\|_2<\sigma_n$ (the smallest singular value), then Theorem 3.5.9 guarantees that $\mathbf{A}+\mathbf{E}$ will also be nonsingular, which means that a stable algorithm applied to a nonsingular system will return a computed solution that is the exact solution to a nearby system that is *again nonsingular*.

On the other hand, if $\mathbf{A}$ is singular (or rank deficient), then perturbations of even the slightest magnitude can increase the rank thereby producing a system with fewer free variables than the original system theoretically demands, so even a stable algorithm can result in a loss of information or degrees of freedom. Although rounding errors are not truly random, they are random enough to make it highly unlikely that the Schur complement $\mathbf{S}$ in (3.5.23) will be zero. Consequently, if $\mathbf{A}$ is rank deficient, then a small perturbation $\mathbf{E}$ due to roundoff will almost certainly cause $rank\,(\mathbf{A}+\mathbf{E})>rank\,(\mathbf{A})$. The moral is to try to avoid floating-point solutions of rank-deficient systems. Such problems can often be distilled down to a nonsingular (or full-rank) core or to full-rank pieces, and these are the components that should be dealt with.

Singular Values of a Perturbed Matrix

Another useful consequence of Theorem 3.5.7 on page 362 is the following result that helps to estimate the singular values of a perturbed matrix.

3.5.10. Theorem. Let $\mathbf{A}, \mathbf{E} \in \mathbb{F}^{m\times n}$, and let $p = \min\{m, n\}$. If $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_p \geq 0$, are *all* of the singular values (nonzero as well as any zero ones) of $\mathbf{A}$ and $\mathbf{A}+\mathbf{E}$, respectively, then

$$|\sigma_k - \beta_k| \leq \|\mathbf{E}\|_2 \quad \text{for each} \quad k = 1, 2, \ldots, p.$$

In particular, small perturbations to $\mathbf{A}$ can only affect small perturbations to its singular values.

Proof. Let $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ be an SVD, where $\mathbf{D}_p = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_p)$. Note that not all zero blocks are present depending on the value of p. Set $\mathbf{A}_{k-1} = \mathbf{U}\begin{pmatrix} \mathbf{D}_{k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$, where $\mathbf{D}_{k-1} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_{k-1})$. The largest singular value of $\mathbf{A} - \mathbf{A}_{k-1}$ is σ_k so adding and subtracting $\mathbf{E}$ and then using the backward triangle inequality (page 83) produces

$$\begin{aligned} \sigma_k &= \|\mathbf{A} - \mathbf{A}_{k-1}\|_2 = \|\mathbf{A} + \mathbf{E} - \mathbf{A}_{k-1} - \mathbf{E}\|_2 \\ &\geq \|(\mathbf{A}+\mathbf{E}) - \mathbf{A}_{k-1}\|_2 - \|\mathbf{E}\|_2 \\ &\geq \beta_k - \|\mathbf{E}\|_2 \quad \text{(by Theorem 3.5.7).} \end{aligned}$$

Couple this with the observation that

$$\begin{aligned} \sigma_k &= \min_{rank(\mathbf{B})=k-1} \|\mathbf{A} - \mathbf{B}\|_2 = \min_{rank(\mathbf{B})=k-1} \|\mathbf{A} + \mathbf{E} - \mathbf{B} - \mathbf{E}\|_2 \\ &\leq \min_{rank(\mathbf{B})=k-1} \|\mathbf{A} + \mathbf{E} - \mathbf{B}\|_2 + \|\mathbf{E}\|_2 = \beta_k + \|\mathbf{E}\|_2 \end{aligned}$$

to conclude that $|\sigma_k - \beta_k| \leq \|\mathbf{E}\|_2$. ∎

Numerical Rank

The rank of a matrix is an easy concept to grasp, and in theory its value can be determined in a variety of ways, one of which is by counting the number of nonzero singular values. But regardless of whether one counts nonzero singular values or uses any other method, computing $rank\,(\mathbf{A})$ using floating-point arithmetic is problematic (see Exercise 2.8.7 on page 242) because $rank\,(\mathbf{A})$ is integer valued. In other words, $rank\,(\mathbf{A})$ is a discontinuous function of $\mathbf{A}$, so, as observed in Theorem 3.5.9 and (3.5.24), the slightest perturbation can (and most likely will) cause the rank to jump if it is not already maximal. This realization motivates the concept of *numerical rank* that is based on Theorem 3.5.10.

When a stable algorithm is used to numerically compute singular values of $\mathbf{A}$, it will return the exact singular values β_k of a nearby matrix $\mathbf{A}+\mathbf{E}$. Considerable effort has gone into the development of stable algorithms for computing singular values, and good SVD algorithms generally have $\|\mathbf{E}\|_2 \approx 5 \times 10^{-t}\|\mathbf{A}\|_2$ when t-digit floating-point arithmetic is used.[†] If $rank\,(\mathbf{A}) = r$, then $p-r$ of the singular values of $\mathbf{A}$ are theoretically zero, where $p = \min\{m,n\}$. Since Theorem 3.5.10 guarantees that $p-r$ of the computed β_k's cannot be larger than $\|\mathbf{E}\|_2$, it is reasonable to make the following definition.

3.5.11. Definition. For $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $p = \min\{m,n\}$ and a given SVD algorithm that returns the exact singular values $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_p$ of $\mathbf{A}+\mathbf{E}$, the ***numerical*** (***or computed***) ***rank*** of $\mathbf{A}$ is defined to be the number $\tilde{r}$ such that

$$\beta_1 \geq \cdots \geq \beta_{\tilde{r}} > \|\mathbf{E}\|_2 \geq \beta_{\tilde{r}+1} \geq \cdots \geq \beta_p.$$

This is the value that many commercial software packages return when $rank\,(\mathbf{A})$ is called for.

Distortion of the Unit Sphere

Important aspects of a matrix $\mathbf{A}$ are revealed by the shape of

$$\mathbf{A}(\mathcal{S}_2) = \{\mathbf{A}\mathbf{x} \,\big|\, \|\mathbf{x}\|_2 = 1\},$$

the image of the unit 2-sphere $\mathcal{S}_2$ under transformation by $\mathbf{A}$. The degree to which $\mathbf{A}$ distorts $\mathcal{S}_2$ measures the ability of $\mathbf{A}$ to lengthen shorter vectors or to shorten longer ones. The expressions in (3.5.9) and (3.5.10) show that the largest and smallest singular values, σ_1 and σ_n, of $\mathbf{A}_{n\times n}$ are the respective lengths of the longest and shortest vectors in $\mathbf{A}(\mathcal{S}_2)$, and the illustration in Figure 3.5.1 on page 360 suggests that $\mathbf{A}(\mathcal{S}_2)$ is an ellipsoid whose respective longest and shortest semi-axes have lengths σ_1 and σ_n. It can now be rigorously established that in general, $\mathbf{A}(\mathcal{S}_2)$ is in fact an ellipsoid in $\mathbb{R}^n$, and moreover, the intermediate singular values of $\mathbf{A}$ provide the lengths of the intermediate semi-axes of $\mathbf{A}(\mathcal{S}_2)$.

To see this suppose that $\mathbf{A} \in \mathbb{R}^{n\times n}$ is nonsingular (singular and rectangular matrices are treated in Exercise 3.5.26), and let $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ be an SVD in which $\mathbf{D}$ contains the the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$ of $\mathbf{A}$. Rather than examining the shape of $\mathbf{A}(\mathcal{S}_2)$, first consider the shape of

$$\mathbf{U}^T\big(\mathbf{A}(\mathcal{S}_2)\big) = \big\{\mathbf{w} \,\big|\, \mathbf{w} = \mathbf{U}^T\mathbf{A}\mathbf{x} = \mathbf{D}\mathbf{V}^T\mathbf{x} \quad \text{for } \|\mathbf{x}\|_2 = 1\big\}.$$

[†] See *"Matrix Computations, 4^{th} Edition"* by G. H. Golub and C. F. Van Loan, The Johns Hopkins University Press, Baltimore, 2013.

Since $\mathbf{U}^T$ is an isometry (a length preserving transformation), the shape of $\mathbf{A}(\mathcal{S}_2)$ under transformation by $\mathbf{U}^T$ is not affected—the effect is only to rotate $\mathbf{A}(\mathcal{S}_2)$ around the origin in $\mathbb{R}^n$ (see page 93). Observing that

$$\frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \cdots + \frac{w_n^2}{\sigma_n^2} = \left\|\mathbf{D}^{-1}\mathbf{w}\right\|_2^2 = \left\|\mathbf{V}^T\mathbf{x}\right\|_2^2 = \|\mathbf{x}\|_2^2 = 1 \tag{3.5.25}$$

shows that $\mathbf{U}^T\big(\mathbf{A}(\mathcal{S}_2)\big)$ is an ellipsoid in $\mathbb{R}^n$ whose k^{th} semi-axis is $\sigma_k\mathbf{e}_k$, or equivalently, $\mathbf{A}(\mathcal{S}_2)$ is the ellipsoid whose k^{th} semi-axis is

$$\mathbf{U}(\sigma_k\mathbf{e}_k) = \sigma_k\mathbf{u}_k, \quad \text{where } \mathbf{u}_k \text{ is } k^{th} \text{ left-hand singular vector.}$$

Furthermore, because $\mathbf{A}\mathbf{v}_k = \sigma_k\mathbf{u}_k$ for the right-hand singular vector $\mathbf{v}_k = \mathbf{V}_{*k}$, it follows that $\mathbf{v}_k$ is a point on $\mathcal{S}_2$ that is mapped to the k^{th} semi-axis vector on $\mathbf{A}(\mathcal{S}_2)$. Figure 3.5.2 below depicts this situation in $\mathbb{R}^3$.

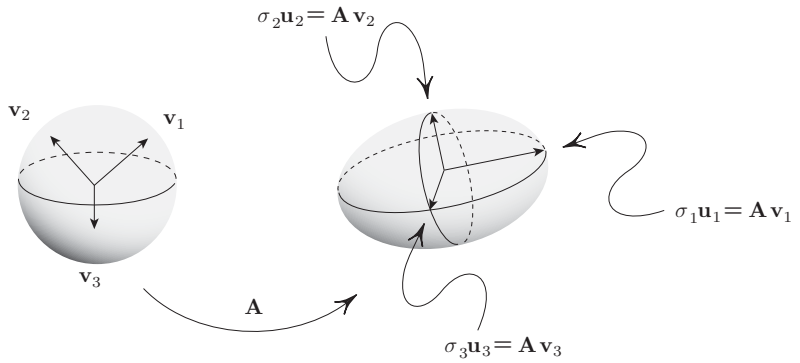

FIGURE 3.5.2: SEMI-AXES AS SINGULAR VECTORS

The degree to which $\mathbf{A}$ distorts the unit 2-sphere is measured by the ratio of the largest stretch to the smallest one, and Theorem 3.5.5 (page 360) guarantees that this "distortion ratio" is given by

$$\kappa_2 = \frac{\max\limits_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2}{\min\limits_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2} = \frac{\sigma_1}{\sigma_3} = \|\mathbf{A}\|_2 \left\|\mathbf{A}^{-1}\right\|_2.$$

Such an expression was called a *condition number* for $\mathbf{A}$ on pages 170 and 249, but it was without regard to a specific norm. The point here is that with respect to the 2-norm, there is an explicit and elegant description of the "2-norm condition number." These observations are formally summarized below.

3.5.12. Theorem. (Distortion of the Unit Sphere) Let $\sigma_1 \geq \cdots \geq \sigma_n > 0$ be the singular values of $\mathbf{A}_{n\times n}$, and let $\mathbf{u}_k$ and $\mathbf{v}_k$ be respective left-hand and right-hand singular vectors for σ_k. The image $\mathbf{A}(\mathcal{S}_2)$ of the unit 2-sphere is an ellipsoid whose k^{th} semi-axis is $\sigma_k\mathbf{u}_k$, and

- $\sigma_1 = \|\mathbf{A}\mathbf{v}_1\|_2 = \max\limits_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{A}\|_2,$ (3.5.26)
- $\sigma_n = \|\mathbf{A}\mathbf{v}_n\|_2 = \min\limits_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2 = 1/\|\mathbf{A}^{-1}\|_2.$ (3.5.27)

The degree of distortion of the unit sphere under transformation by $\mathbf{A}$ is measured by the ***2-norm condition number*** that is defined to be

- $\kappa_2 = \|\mathbf{A}\|_2 \left\|\mathbf{A}^{-1}\right\|_2 = \dfrac{\sigma_1}{\sigma_n} \geq 1.$ (3.5.28)
- $\kappa_2 = 1$ if and only if $\mathbf{A}$ is unitary. (3.5.29)
- Exercise 3.5.17 treats rectangular and rank-degenerate matrices.

Proof. The facts in (3.5.26)–(3.5.28) are from the preceding discussions, and (3.5.29) follows from the observation that $\sigma_1 = \sigma_n$ if and only if $\mathbf{D} = \mathbf{I}$ in an SVD $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ so that $\mathbf{A} = \mathbf{U}\mathbf{V}^T$, which is unitary. ∎

Uncertainties, Errors, and Perturbations in Linear Systems

Uncertainties in a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ arise in various ways. For example, they may emanate from modeling errors (simplifying assumptions are frequently required); they might be due to data collection errors (infinitely precise measurement devices do not exist); there may be data-entry errors (numbers like $\sqrt{2}$, π, and $2/3$ cannot be entered exactly); and finally, errors arising from floating-point computation are a prevalent source of uncertainty. Uncertainties may reside exclusively in $\mathbf{b}$, or exclusively in $\mathbf{A}$, or they may influence both $\mathbf{A}$ and $\mathbf{b}$. A fundamental issue for practical users of linear algebra is to gauge the degree to which uncertainties, errors, or perturbations can affect the solution $\mathbf{x}$. Related problems were treated earlier (pages 170, 174, and 249) with the aid of calculus to provide intuitive insight, but now a more complete and detailed picture can be developed to show how all of these issues boil down to analyzing SVD components.

The simplest case is when only the right-hand side $\mathbf{b}$ is affected. This is also one of the most prevalent situations in practical work because it is often the case that $\mathbf{b}$ contains input data derived from empirical observations whereas $\mathbf{A}$ is known exactly because it is based on underlying physics defining the problem. And sometimes it is possible to aggregate uncertainties and shift all of them to the right-hand side. So, start by assuming that $\mathbf{A}\mathbf{x} = \mathbf{b}$ is a nonsingular system in which only $\mathbf{b}$ is subject to an uncertainty $\mathbf{e}$, and consider $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b} - \mathbf{e} = \tilde{\mathbf{b}}$.

The *relative uncertainty*[†] in $\mathbf{b}$ is $\|\mathbf{e}\|_2 / \|\mathbf{b}\|_2 = \|\mathbf{b}-\tilde{\mathbf{b}}\|_2 / \|\mathbf{b}\|_2$, and the relative uncertainty in $\mathbf{x}$ is $\|\mathbf{x}-\tilde{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2$. The following theorem shows how the relative uncertainty in $\mathbf{x}$ is bounded by the relative uncertainty in $\mathbf{b}$.

Notation: Normally $\mathbf{e}$ designates a vector of all ones, but throughout the remainder of this section the notation is changed to allow $\mathbf{e}$ to denote a vector whose components represent uncertainties or errors.

3.5.13. Theorem. If a nonsingular system $\mathbf{Ax} = \mathbf{b}$ is perturbed to become $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b} - \mathbf{e} = \tilde{\mathbf{b}}$, then the relative change in the solution is bounded above and below as follows.

$$\kappa_2^{-1} \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} \leq \frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_2 \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}, \tag{3.5.30}$$

where $\kappa_2 = \sigma_1/\sigma_n = \|\mathbf{A}\|_2 \left\|\mathbf{A}^{-1}\right\|_2$ is the condition number in (3.5.28).

Proof. Start with $\|\mathbf{b}\|_2 = \|\mathbf{Ax}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$ and $\mathbf{x} - \tilde{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{e}$ to write

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|}{\|\mathbf{x}\|} = \frac{\left\|\mathbf{A}^{-1}\mathbf{e}\right\|_2}{\|\mathbf{x}\|_2} \leq \frac{\|\mathbf{A}\|_2 \left\|\mathbf{A}^{-1}\right\|_2 \|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} = \kappa_2 \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}.$$

Now combine $\|\mathbf{x}\|_2 \leq \left\|\mathbf{A}^{-1}\right\|_2 \|\mathbf{b}\|_2$ and $\|\mathbf{e}\|_2 = \|\mathbf{A}(\mathbf{x}-\tilde{\mathbf{x}})\|_2 \leq \|\mathbf{A}\|_2 \|(\mathbf{x}-\tilde{\mathbf{x}})\|_2$ to produce

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \geq \frac{\|\mathbf{e}\|_2}{\|\mathbf{A}\|_2 \|\mathbf{x}\|_2} \geq \frac{\|\mathbf{e}\|_2}{\|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2 \|\mathbf{b}\|_2} = \frac{1}{\kappa_2} \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}. \quad \blacksquare$$

Equality is Possible

A matrix $\mathbf{A}$ is considered to be well conditioned when κ_2 is small[‡] relative to 1 because in such a case Theorem 3.5.13 means that small relative uncertainties in $\mathbf{b}$ cannot greatly affect the solution, but as $\mathbf{A}$ becomes more ill conditioned (i.e., as κ_2 becomes larger), small relative uncertainties in $\mathbf{b}$ *might* produce larger relative uncertainties in $\mathbf{x}$. To clarify this, it must be determined if equality in (3.5.30) can be realized for every nonsingular $\mathbf{A}$. The next theorem shows that equality in each side of (3.5.30) is indeed possible, and the relationship between $\mathbf{b}$ and its uncertainties $\mathbf{e}$ that will produce equality can be made explicit.

[†] Knowing or estimating *absolute* uncertainties such as $\|\mathbf{e}\|_2$ and $\|\mathbf{x}-\tilde{\mathbf{x}}\|_2$ are generally not meaningful or helpful. For example, an absolute uncertainty of a half of an inch might be excellent when measuring the distance between the earth and the nearest star, but it is not good in the practice of eye surgery.

[‡] See the *rule of thumb* on page 250 to get a feeling of what "small" and "large" might mean in the context of numerical analysis.

3.5.14. Theorem. Let $\mathbf{Ax} = \mathbf{b}$, where $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n > 0$ are the singular values of $\mathbf{A}_{n\times n}$, and let $\mathbf{v}_k$ and $\mathbf{u}_k$ be respective right-hand and left-hand singular vectors associated with σ_k. If $\mathbf{b}$ is perturbed by $\mathbf{e}$ to produce $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b} - \mathbf{e} = \tilde{\mathbf{b}}$, and if $\mathbf{b}$ and $\mathbf{e}$ are respectively directed along $\mathbf{u}_1$ and $\mathbf{u}_n$, then

$$\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} = \kappa_2 \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} \quad \text{(equality in the right-hand side of (3.5.30)).}$$

However if $\mathbf{b}$ and $\mathbf{e}$ are respectively directed along $\mathbf{u}_n$ and $\mathbf{u}_1$, then

$$\kappa_2^{-1} \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} = \frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \quad \text{(equality in the left-hand side of (3.5.30)).}$$

Proof. Suppose first that $\mathbf{b} = \beta\mathbf{u}_1$ and $\mathbf{e} = \epsilon\,\mathbf{u}_n$ for scalars β and ϵ. Since $\mathbf{Av}_k = \sigma_k\mathbf{u}_k$ implies that $\mathbf{v}_k/\sigma_k = \mathbf{A}^{-1}\mathbf{u}_k$ for each k, it follows that

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{A}^{-1}(\beta\mathbf{u}_1) = \frac{\beta\mathbf{v}_1}{\sigma_1} \quad \text{and} \quad \mathbf{x} - \tilde{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{e} = \mathbf{A}^{-1}(\epsilon\mathbf{u}_n) = \frac{\epsilon\mathbf{v}_n}{\sigma_n},$$

and thus

$$\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} = \left(\frac{\sigma_1}{\sigma_n}\right)\frac{|\epsilon|}{|\beta|} = \kappa_2 \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}.$$

On the other hand, if $\mathbf{b} = \beta\mathbf{u}_n$ and $\mathbf{e} = \epsilon\,\mathbf{u}_1$, then the same argument yields $\mathbf{x} = \beta\mathbf{v}_n/\sigma_n$ and $\mathbf{x} - \tilde{\mathbf{x}} = \epsilon\,\mathbf{v}_1/\sigma_1$, so

$$\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} = \left(\frac{\sigma_n}{\sigma_1}\right)\frac{|\epsilon|}{|\beta|} = \kappa_2^{-1} \frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}. \quad \blacksquare$$

In other words, Theorem 3.5.14 guarantees that the worst case as well as the best case scenarios can each be realized for every nonsingular $\mathbf{A}$, so while a small κ_2 ensures that relatively small changes in $\mathbf{b}$ cannot produce relatively large changes in $\mathbf{x}$, it is a certainty that a large κ_2 will perpetuate a huge effect when $\mathbf{b}$ and $\mathbf{e}$ lie in unfortunate directions. But on the other hand, it is also a certainty that if $\mathbf{b}$ and $\mathbf{e}$ lie in more fortunate directions, then changes in $\mathbf{b}$ will have almost no effect on $\mathbf{x}$—in fact, a large κ_2 can actually mitigate the effects of $\mathbf{e}$ resulting in a minimal change in $\mathbf{x}$. Nevertheless, in light of the fact that the direction of $\mathbf{e}$ can rarely be known, the worst case must be guarded against, and increasing caution is required as κ_2 increases.

In hind sight, the results in Theorems 3.5.13 and 3.5.14 could have been anticipated from the geometry illustrated in Figure 3.5.2 on page 368 because

$\mathbf{A}^{-1}$ maps the ellipsoid defined by the singular vectors back to the unit sphere. This means that when a vector $\mathbf{y}$ is in (or near) the span of $\mathbf{u}_1$, its length is reduced by $\mathbf{A}^{-1}$ relative to that when $\mathbf{y}$ is in (or near) the span of $\mathbf{u}_n$. More precisely, if $\mathbf{y}_1 = \alpha\mathbf{u}_1$ and $\mathbf{y}_n = \alpha\mathbf{u}_n$, then $\|\mathbf{y}_1\|_2 = |\alpha| = \|\mathbf{y}_n\|_2$, but $\mathbf{A}^{-1}\mathbf{y}_1 = \alpha\sigma_1^{-1}\mathbf{v}_1$ and $\mathbf{A}^{-1}\mathbf{y}_n = \alpha\sigma_n^{-1}\mathbf{v}_n$, so

$$\frac{\left\|\mathbf{A}^{-1}\mathbf{y}_1\right\|_2}{\left\|\mathbf{A}^{-1}\mathbf{y}_n\right\|_2} = \kappa_2^{-1} \leq 1 \implies \left\|\mathbf{A}^{-1}\mathbf{y}_1\right\|_2 = \kappa_2^{-1}\left\|\mathbf{A}^{-1}\mathbf{y}_n\right\|_2 \leq \left\|\mathbf{A}^{-1}\mathbf{y}_n\right\|_2.$$

Continuity dictates that the same holds when $\mathbf{y}_1$ and $\mathbf{y}_n$ are sufficiently close to the relative spans of $\mathbf{u}_1$ and $\mathbf{u}_n$.

Uncertainty in Both Sides

Suppose that there are uncertainties in both sides of a nonsingular system $\mathbf{Ax} = \mathbf{b}$, and consider $(\mathbf{A} - \mathbf{E})\tilde{\mathbf{x}} = \mathbf{b} - \mathbf{e}$. The aim now is to establish an upper bound on $\|\mathbf{x} - \tilde{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2$ (the relative change in $\mathbf{x}$) in terms of both $\|\mathbf{E}\|_2 / \|\mathbf{A}\|_2$ (the relative change in $\mathbf{A}$) and $\|\mathbf{e}\|_2 / \|\mathbf{b}\|_2$ (the relative change in $\mathbf{b}$) along with the condition number $\kappa_2 = \sigma_1/\sigma_n$. Theorem 2.9.1 on page 250 does this in a continuous sense by using calculus to show that if $\mathbf{A} = \mathbf{A}(t)$ and $\mathbf{b} = \mathbf{b}(t)$ are differentiable functions of a variable t in an interval $[a, b]$ on which $\mathbf{A}(t)$ is nonsingular, and if $\mathbf{x} = \mathbf{x}(t)$ is the solution to $\mathbf{A}(t)\mathbf{x}(t) = \mathbf{b}(t)$, then the relative size of the derivative $\mathbf{x}' = \mathbf{x}'(t)$ is

$$\frac{\|\mathbf{x}'\|}{\|\mathbf{x}\|} \leq \kappa_2 \left(\frac{\|\mathbf{b}'\|}{\|\mathbf{b}\|} + \frac{\|\mathbf{A}'\|}{\|\mathbf{A}\|}\right).$$

While this is mathematically elegant, it lack specificity about the acceptable size of uncertainties that might be tolerable. The discrete bound below is not as pretty as the derivative bound, but it more clearly illuminate things.

3.5.15. Theorem. Let $\mathbf{A} \in \mathbb{F}^{n\times n}$ be nonsingular with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n > 0$, and consider $\mathbf{Ax} = \mathbf{b}$ and $(\mathbf{A} - \mathbf{E})\tilde{\mathbf{x}} = \mathbf{b} - \mathbf{e}$. If $\|\mathbf{E}\|_2 < \sigma_n$, then $\mathbf{A} - \mathbf{E}$ is also nonsingular, and

$$\frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \frac{\kappa_2}{1 - \kappa_2 \|\mathbf{E}\|_2 / \|\mathbf{A}\|_2}\left(\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} + \frac{\|\mathbf{E}\|_2}{\|\mathbf{A}\|_2}\right),$$

where $\kappa_2 = \sigma_1/\sigma_n = \|\mathbf{A}\|_2 \left\|\mathbf{A}^{-1}\right\|_2$.

Proof. The fact that $\mathbf{A}-\mathbf{E}$ is nonsingular when $\|\mathbf{E}\|_2 < \sigma_n$ is a consequence of Theorem 3.5.9 on page 364. For convenience, let $\mathbf{B} = \mathbf{A}^{-1}\mathbf{E}$, and observe that $\|\mathbf{B}\|_2 \le \left\|\mathbf{A}^{-1}\right\|_2 \|\mathbf{E}\|_2 < \sigma_n/\sigma_n = 1$ so that $(\mathbf{I}-\mathbf{B})^{-1} = \sum_{k=0}^{\infty}\mathbf{B}^k$ (by Theorem 2.3.11 on page 168). Starting with

$$\mathbf{x}-\tilde{\mathbf{x}} = \mathbf{x}-(\mathbf{I}-\mathbf{B})^{-1}\mathbf{A}^{-1}(\mathbf{b}-\mathbf{e}) = \left(\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1}\right)\mathbf{x} + (\mathbf{I}-\mathbf{B})^{-1}\mathbf{A}^{-1}\mathbf{e},$$

take norms and apply the inequalities

$$\mathbf{b}=\mathbf{A}\mathbf{x} \implies 1/\|\mathbf{x}\|_2 \le \frac{\|\mathbf{A}\|_2}{\|\mathbf{b}\|_2} \implies \frac{\left\|\mathbf{A}^{-1}\mathbf{e}\right\|_2}{\|\mathbf{x}\|_2} \le \frac{\|\mathbf{A}\|_2\left\|\mathbf{A}^{-1}\right\|_2\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} = \kappa_2\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}$$

to write

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \le \left\|\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1}\right\|_2 + \left\|(\mathbf{I}-\mathbf{B})^{-1}\right\|_2 \kappa_2\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}. \tag{3.5.31}$$

The identity $\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1} = -\mathbf{B}(\mathbf{I}-\mathbf{B})^{-1}$ derived from $\mathbf{I} = (\mathbf{I}-\mathbf{B})(\mathbf{I}-\mathbf{B})^{-1}$ together with the triangle inequality yields

$$\left\|(\mathbf{I}-\mathbf{B})^{-1}\right\|_2 \le \sum_{k=0}^{\infty}\|\mathbf{B}\|_2^k = \frac{1}{1-\|\mathbf{B}\|_2} \quad \text{and} \quad \left\|\mathbf{I}-(\mathbf{I}-\mathbf{B})^{-1}\right\|_2 \le \frac{\|\mathbf{B}\|_2}{1-\|\mathbf{B}\|_2}.$$

Use these in (3.5.31) together with $\|\mathbf{B}\|_2 \le \left\|\mathbf{A}^{-1}\right\|_2\|\mathbf{E}\|_2 = \kappa_2\|\mathbf{E}\|_2/\|\mathbf{A}\|_2$ to conclude that

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \le \frac{\kappa_2}{1-\kappa_2\|\mathbf{E}\|_2/\|\mathbf{A}\|_2}\left(\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}+\frac{\|\mathbf{E}\|_2}{\|\mathbf{A}\|_2}\right). \quad \blacksquare$$

Note: Theorems 3.5.13 and 3.5.15 remain valid for any matrix norm for which $\|\mathbf{I}\| = 1$ and where κ is defined to be $\kappa = \|\mathbf{A}\|\left\|\mathbf{A}^{-1}\right\|$ (Exercise 3.5.23).

Checking an Answer

Suppose that $\tilde{\mathbf{x}}$ is a computed (or approximate) solution for a nonsingular system $\mathbf{A}\mathbf{x}=\mathbf{b}$, and suppose that the accuracy of $\tilde{\mathbf{x}}$ is "checked" by computing the *residual* $\mathbf{r} = \mathbf{b}-\mathbf{A}\tilde{\mathbf{x}}$. If every component of $\mathbf{r}$ is *exactly* zero, then $\tilde{\mathbf{x}}$ must be the exact solution. However, if $\mathbf{r}\neq\mathbf{0}$ but $\|\mathbf{r}\|_2$ is zero to t significant digits, can we be confident that $\tilde{\mathbf{x}}$ be accurate to roughly t significant figures? No! As illustrated in Bob's dance with the Devil on page 246, the signs may not even be correct. Theorems 3.5.13 and 3.5.14 remove the mystery behind using residuals to check an answer.

Rewrite $\mathbf{r} = \mathbf{b}-\mathbf{A}\tilde{\mathbf{x}}$ as $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}-\mathbf{r}$, and apply Theorem 3.5.13 on page 370 with the error term $\mathbf{e}$ replaced by $\mathbf{r}$ to conclude that

$$\kappa_2^{-1}\frac{\|\mathbf{r}\|_2}{\|\mathbf{b}\|_2} \le \frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \le \kappa_2\frac{\|\mathbf{r}\|_2}{\|\mathbf{b}\|_2}. \tag{3.5.32}$$

In other words, this says that for relatively small values of κ_2, a computed (or approximate) solution $\tilde{\mathbf{x}}$ is relatively accurate if and only if $\|\mathbf{r}\|_2$ is relatively small. However, Theorem 3.5.14 on page 371 says that equality on the right-hand side of (3.5.32) is possible, so when this occurs and κ_2 is large enough to overwhelm the size a small $\|\mathbf{r}\|_2$, the computed (or approximate) solution $\widetilde{\mathbf{x}}$ will necessarily be inaccurate. Thus the bottom line concerning "checking an answer" is as follows.

Residuals are reliable indicators of accuracy only when $\mathbf{A}$ *is reasonably well conditioned. As* $\mathbf{A}$ *becomes more ill conditioned, the residuals become increasingly unreliable indicators of accuracy.*

Exercises for section 3.5

3.5.1. Evaluate the matrix 2-norm for each matrix below.

$$\mathbf{A} = \begin{pmatrix} 1 & -2 \\ -1 & 2 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 4 & -2 & 4 \\ -2 & 1 & -2 \\ 4 & -2 & 4 \end{pmatrix}.$$

3.5.2. Determine $\|\mathbf{A}\|_2$ as well as $\|\mathbf{A}^{-1}\|_2$ for $\mathbf{A} = \dfrac{1}{\sqrt{3}} \begin{pmatrix} 3 & -1 \\ 0 & \sqrt{8} \end{pmatrix}$.

3.5.3. Determine a singular value decompositions for each of the following matrices.

(a) $\mathbf{A} = \begin{pmatrix} -4 & -6 \\ 3 & -8 \end{pmatrix}$ (b) $\mathbf{B} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$

3.5.4. Explain why both of the following factorizations represent singular value decompositions of the same matrix.

$$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \\ 0 & 1/\sqrt{3} & -2/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{3} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix} (1) = \begin{pmatrix} 1/\sqrt{2} & 0 & \sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix} (1)$$

3.5.5. If $\mathbf{A}$ is a normal matrix, how are its eigenvalues related to its singular values? How is this different from the case in which $\mathbf{A}$ is hermitian?

3.5.6. If λ is an eigenvalue for $\mathbf{A} \in \mathbb{F}^{n \times n}$, then $1 + \lambda$ is an eigenvalue for $\mathbf{I} + \mathbf{A}$ (by Exercise 3.1.15 on page 298). Is the same true for singular values? That is, if σ is a singular value for $\mathbf{A}$, must $1+\sigma$ be a singular value for $\mathbf{I} + \mathbf{A}$?

3.5.7. Proving that $rank\,(\,\mathbf{A}^*\mathbf{A}\mathbf{A}^*\,) = rank\,(\mathbf{A})$ was Exercise 2.4.21 on page 195. Use an SVD to establish the same result.

3.5.8. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the nonzero singular values of $\mathbf{A} \in \mathbb{F}^{m\times n}$. Explain why the distinct eigenvalues of $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{A}^* & \mathbf{0} \end{pmatrix}$ are $\sigma\,(\mathbf{B}) = \{\pm\sigma_1, \pm\sigma_2, \ldots \pm\,\sigma_r, 0\}$.

3.5.9. For $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $rank\,(\mathbf{A}) = r$, let $\mathbf{U}_{m\times m} = [\underbrace{\mathbf{U}_1}_{m\times r} \,|\, \underbrace{\mathbf{U}_2}_{m\times(m-r)}]$ be a unitary matrix whose columns are a complete orthonormal set of eigenvectors for $\mathbf{A}\mathbf{A}^*$, and let $\mathbf{Y}_{n\times n} = [\underbrace{\mathbf{Y}_1}_{n\times r} \,|\, \underbrace{\mathbf{Y}_2}_{n\times(n-r)}]$ be a unitary matrix that diagonalizes $\mathbf{A}^*\mathbf{A}$. If $\mathbf{V} = \left[\mathbf{A}^*\mathbf{U}_1\mathbf{D}^{-1} \,|\, \mathbf{Y}_2\right]$, where $\mathbf{D} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$ contains the singular values of $\mathbf{A}$, show that $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ is a singular value decomposition of $\mathbf{A}$.

3.5.10. For a convergent sequence $\{\mathbf{A}_k\}_{k=1}^{\infty}$ of matrices, let $\mathbf{A} = \lim_{k\to\infty}\mathbf{A}_k$.

(a) Prove that if each $\mathbf{A}_k$ is singular, then $\mathbf{A}$ is singular.

(b) If each $\mathbf{A}_k$ is nonsingular, must $\mathbf{A}$ be nonsingular? Why?

3.5.11. Let $\mathbf{u}, \mathbf{v} \in \mathbb{F}^{m\times 1}$ be nonzero. Prove that the nonzero singular value of the rank-one matrix $\mathbf{A} = \mathbf{u}\mathbf{v}^*$ is $\sigma = \|\mathbf{u}\|_2\,\|\mathbf{v}\|_2$. **Note:** This means that $\|\mathbf{A}\|_2 = \|\mathbf{u}\mathbf{v}^*\|_2 = \|\mathbf{u}\|_2\,\|\mathbf{v}\|_2$.

3.5.12. Explain why every $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $rank\,(\mathbf{A}) = r$ can be expanded as

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^*,$$

where the σ_i's are the nonzero singular values of $\mathbf{A}$, and the $\mathbf{u}_i$'s and $\mathbf{v}_i$'s are the respective left-hand and right-hand singular vectors.

3.5.13. Give an example to show that there need not be a unique closest matrix of rank k to a given matrix $\mathbf{A}$. In other words, if $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ are the nonzero singular values of $\mathbf{A}$, then for $k < r$ there can be two matrices $\mathbf{B}_1 \neq \mathbf{B}_2$ such that $\|\mathbf{A} - \mathbf{B}_1\|_2 = \|\mathbf{A} - \mathbf{B}_1\|_2 = \sigma_{k+1}$.

3.5.14. SVD and the Pseudo Inverse. Let $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ be an SVD for $\mathbf{A} \in \mathbb{F}^{m\times n}$, where $rank\,(\mathbf{A}) = r$ and $\mathbf{D} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$.

(a) Explain why the pseudo inverse of $\mathbf{A} \in \mathbb{F}^{m\times n}$ is

$$\mathbf{A}^\dagger = \mathbf{V}\begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^* = \sum_{i=1}^{r} \sigma_i^{-1}\mathbf{v}_i\mathbf{u}_i^*.$$

Hint: Recall Exercise 2.4.43, page 197.

(b) Use this SVD formulation to verify that $\mathbf{A}^\dagger$ satisfies the four Penrose equations given in (2.4.16) on page 190—i.e., show that

$$\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A},\ \mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger,\ (\mathbf{A}\mathbf{A}^\dagger)^* = \mathbf{A}\mathbf{A}^\dagger,\ (\mathbf{A}^\dagger\mathbf{A})^* = \mathbf{A}^\dagger\mathbf{A}.$$

3.5.15. Pseudo Inverses Lack Continuity. Theorem 2.3.12 on page 169 guarantees that the inverse of a nonsingular matrix varies continuously with the entries in $\mathbf{A}$. Show that the same is *not* true for the pseudo inverse by considering $\mathbf{A}(x) = \begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix}$.

3.5.16. Let σ_r be the smallest nonzero singular value of $\mathbf{A}_{m\times n}$. Prove that if $|\epsilon| < \sigma_r^2$, then $(\mathbf{A}^T\mathbf{A}+\epsilon\mathbf{I})^{-1}$ exists, and $\lim_{\epsilon\to 0}(\mathbf{A}^T\mathbf{A}+\epsilon\mathbf{I})^{-1}\mathbf{A}^T = \mathbf{A}^\dagger$.

3.5.17. Prove that if $rank\,(\mathbf{A}_{m\times n}) = r$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the nonzero singular values, then (3.5.27) on page 369 generalizes to say

$$\sigma_r = \min_{\substack{\|\mathbf{x}\|_2=1 \\ \mathbf{x}\perp N(\mathbf{A})}} \|\mathbf{A}\mathbf{x}\|_2 = 1/\left\|\mathbf{A}^\dagger\right\|_2.$$

3.5.18. Generalized Condition Number. Let $rank\,(\mathbf{A}_{m\times n}) = r$ with nonzero singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Extend the bounds in Theorem 3.5.13 on page 370 to include singular and rectangular matrices by showing that if $\mathbf{x}$ and $\tilde{\mathbf{x}}$ are the respective minimum 2-norm solutions of consistent systems $\mathbf{A}\mathbf{x} = \mathbf{b}$ and $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}} = \mathbf{b} - \mathbf{e}$ (recall Exercise 2.4.35 on page 196), then

$$\kappa_2^{-1}\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2} \leq \frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \kappa_2\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}, \quad \text{where} \quad \kappa_2 = \frac{\sigma_1}{\sigma_r} = \|\mathbf{A}\|_2\left\|\mathbf{A}^\dagger\right\|_2.$$

Does the same reasoning in the proof of Theorem 3.5.14 on page 371 prove that these upper and lower bounds are attainable for every $\mathbf{A}$?

3.5.19. Let $\mathbf{A} \in \mathbb{F}^{m\times n}$ have $rank\,(\mathbf{A}) = r$. Prove that if $\mathbf{A} = \mathbf{U}\begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^*$ is an SVD in which $\mathbf{D} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$, then the solution of minimum norm for a consistent linear system $\mathbf{Ax} = \mathbf{b}$ is

$$\mathbf{x} = \sum_{i=1}^{r} \frac{\langle \mathbf{u}_i | \mathbf{b}\rangle}{\sigma_i}\mathbf{v}_i = \mathbf{A}^\dagger\mathbf{b}.$$

Hint: Recall Exercise 2.4.35 on page 196.

3.5.20. Normal Equations and SVD. For $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $rank\,(\mathbf{A}) = r$ and for $\mathbf{b} \in \mathbb{F}^{m\times 1}$, the linear system $\mathbf{A}^*\mathbf{Ax} = \mathbf{A}^*\mathbf{b}$ is called the system of *normal equations*, and such a system is always consistent—see Exercise 2.4.34 on page 196. Prove that if $\mathbf{A} = \mathbf{U}\begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^*$ is an SVD in which $\mathbf{D} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$, then a solution of the normal equations is

$$\mathbf{x} = \sum_{i=1}^{r} \frac{\langle \mathbf{u}_i | \mathbf{b}\rangle}{\sigma_i}\mathbf{v}_i = \mathbf{A}^\dagger\mathbf{b}.$$

3.5.21. Show that if κ is the two-norm condition number for $\mathbf{A} \in \mathbb{F}^{m\times n}$, then the two-norm condition number for $\mathbf{A}^*\mathbf{A}$ is κ^2.

Note: In particular, this means that if $\mathbf{A}$ is ill conditioned, then the system of normal equations $\mathbf{A}^*\mathbf{A} = \mathbf{A}^*\mathbf{b}$ is even more so, and hence solving the normal equations may pose numerical difficulties when using float-point arithmetic.

3.5.22. Norms and Singular Values. Let $\mathbf{A} \in \mathbb{F}^{m\times n}$ with $rank\,(\mathbf{A}) = r$.

(a) Show that $\|\mathbf{A}\|_F = \left(\sum_{i=1}^{r}\sigma_i^2\right)^{1/2}$, where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the nonzero singular values of $\mathbf{A}$.

(b) Conclude that

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{n}\,\|\mathbf{A}\|_2\,,$$

and note that this explains the $(2, F)$ and the $(F, 2)$ entry in the matrix

$$\begin{array}{c} \\ 1 \\ 2 \\ \infty \\ F \end{array}\begin{array}{c}\begin{array}{cccc} 1 & 2 & \infty & F \end{array}\\ \begin{pmatrix} * & \sqrt{n} & n & \sqrt{n} \\ \sqrt{n} & * & \sqrt{n} & 1 \\ n & \sqrt{n} & * & \sqrt{n} \\ \sqrt{n} & \sqrt{n} & \sqrt{n} & * \end{pmatrix}\end{array}. \tag{3.5.33}$$

from (1.9.10) on page 87 that provides the multipliers α_{ij} for the relations $\|\mathbf{A}\|_i \leq \alpha_{ij} \|\mathbf{A}\|_j$.

(c) Now derive the remaining entries α_{ij} in (3.5.33) for $\mathbf{A} \in \mathbb{F}^{n\times n}$.
Hint: Recall the relations between the common vector norms from Exercise 1.5.4 on page 37, and revisit the proof of Theorem 1.9.6 on page 86.

Note: Other families of matrix norms are defined in terms of the singular values. For example, using just the first k singular values produces the *Hilbert–Schmidt norms* $\|\mathbf{A}\| = \left(\sum_{i=1}^{k} \sigma_i^2\right)^{1/2}$. The matrix 2-norm $\|\mathbf{A}\|_2 = \sigma_1$ is the most important special case. Another variation called the *Schatten p-norms* are defined by taking $\|\mathbf{A}\| = (\sum_{i=1}^{r} \sigma_i^p)^{1/p}$ for integers $p > 0$.

3.5.23. Verify that the results in Theorem 3.5.13 on page 370 and Theorem 3.5.15 on page 372 hold for any matrix norm for which $\|\mathbf{I}\|_\star = 1$, where κ_2 is replaced by $\kappa_\star = \|\mathbf{A}\|_\star \left\|\mathbf{A}^{-1}\right\|_\star$.

3.5.24. Ky Fan's Extension. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_s$ be the singular values of $\mathbf{A} \in \mathbb{F}^{m\times n}$, where $s = \min\{m, n\}$. Use Ky Fan's Theorem on page 344 to prove that for matrices $\mathbf{X} \in \mathbb{F}^{n\times k}$ such that $\mathbf{X}^T\mathbf{X} = \mathbf{I}_k$ with $k \leq s$,

$$\sum_{j=1}^{k} \sigma_j^2 = \max_{\mathbf{X}^T\mathbf{X}=\mathbf{I}_k} \|\mathbf{AX}\|_F \quad \text{and} \quad \sum_{j=n-k+1}^{n} \sigma_j^2 = \min_{\mathbf{X}^T\mathbf{X}=\mathbf{I}_k} \|\mathbf{AX}\|_F .$$

3.5.25. Pivots and Conditioning. An ill-conditioned matrix might be suspected when a small pivot u_{ii} emerges during the LU factorization of $\mathbf{A}$ because $\left[\mathbf{U}^{-1}\right]_{ii} = 1/u_{ii}$ is then large, and this opens the possibility of $\mathbf{A}^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}$ having large entries. Unfortunately, this is not an absolute test, and no guarantees about conditioning can be made from the pivots alone.

(a) Construct an example of a matrix that is well conditioned but has a small pivot.

(b) Construct an example of a matrix that is ill conditioned but has no small pivots.

3.5.26. Extend the discussion on page 367 concerning the distortion of the unit 2-sphere $\mathcal{S}_2 \subset \mathbb{R}^n$ to include singular and rectangular matrices by showing that if $rank\,(\mathbf{A}_{m\times n}) = r$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the nonzero singular values of $\mathbf{A}$, then the image $\mathbf{A}(\mathcal{S}_2) \subset \mathbb{R}^m$ is an ellipsoid (possibly degenerate) in which the k^{th} semi-axis is $\sigma_k\mathbf{u}_k = \mathbf{A}\mathbf{v}_k$, where $\mathbf{u}_k$ and $\mathbf{v}_k$ are respective left-hand and right-hand singular vectors for $\mathbf{A}$.

Solutions for exercises in section 3. 5

3.5.1. $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{A}^T\mathbf{A}$. Computing these eigenvalues in each case yields $\|\mathbf{A}\|_2 = \sqrt{10}$, $\|\mathbf{B}\|_2 = 1$, and $\|\mathbf{C}\|_2 = 9$.

3.5.2. The characteristic equation of $\mathbf{A}^T\mathbf{A}$ is $\lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda - 2)$, so the singular values of $\mathbf{A}$ are $\sigma_1 = 2$ and $\sigma_2 = \sqrt{2}$, and hence $\|\mathbf{A}\|_2 = \sigma_1 = 2$ and $\|\mathbf{A}^{-1}\|_2 = 1/\sqrt{2}$.

3.5.3. (a) $\sigma_1 = 10$, $\sigma_2 = 5$, and $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \begin{pmatrix} -6/10 & -4/5 \\ -8/10 & 3/5 \end{pmatrix}\begin{pmatrix} 10 & 0 \\ 0 & 5 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

(b) $\sigma_1 = \sqrt{2}$, and $\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T = \begin{pmatrix} 1/\sqrt{2} & 0 & \sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & \sqrt{2} \end{pmatrix}\begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}[1]$.

3.5.4. They are both of the form $\mathbf{U}\begin{pmatrix} \mathbf{D} \\ \mathbf{0} \end{pmatrix}\mathbf{V}^T$ where $\mathbf{U}_{3\times 3}$ and $\mathbf{V}_{1\times 1}$ are orthogonal matrices. Both represent singular value decompositions for $\mathbf{A} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$. This shows that singular vectors are not uniquely defined.

3.5.5. If $\mathbf{A}$ is a normal, then it is unitarily similar to a diagonal matrix—say that $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$, where $\mathbf{D} = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$ contains the eigenvalues of $\mathbf{A}$. Since $\mathbf{A}^*\mathbf{A} = \mathbf{U}\overline{\mathbf{D}}\mathbf{D}\mathbf{U}^*$, where $\overline{\mathbf{D}}\mathbf{D}$ is a diagonal matrix whose diagonal entries are all of the form $|\lambda_i|^2$, these numbers must be the eigenvalues of $\mathbf{A}^*\mathbf{A}$, and thus the singular values of $\mathbf{A}$ are the numbers $\sigma_i = |\lambda_i|$. The situation is exactly the same for hermitian matrices. When $\mathbf{A}$ is hermitian, $\mathbf{A}^*\mathbf{A} = \mathbf{A}^2$, so the eigenvalues of $\mathbf{A}^*\mathbf{A}$ are λ_i^2, where λ_i is an eigenvalue for $\mathbf{A}$, and thus the singular values are of the form $\sigma_i = \sqrt{\lambda_i^2} = |\lambda_i|$.

3.5.6. No—almost any 2×2 matrix can serve as a counterexample. For example, if $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then its singular values are $\{1, 0\}$, but the singular values of $\mathbf{I} + \mathbf{A}$ are $[(3 \pm \sqrt{5})/2]^{1/2}$.

3.5.7. Suppose that $rank\,(\mathbf{A}) = r$, and let $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ be an SVD in which $\mathbf{D}_{r\times r}$ is a nonsingular diagonal matrix. Write

$$\mathbf{A}^*\mathbf{A}\mathbf{A}^* = \mathbf{V}\begin{pmatrix} \mathbf{D}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^*\mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*\mathbf{V}\begin{pmatrix} \mathbf{D}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^* = \mathbf{V}\begin{pmatrix} \mathbf{D}^*\mathbf{D}\mathbf{D}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^*,$$

and use the fact that $rank\,(\mathbf{D}^*\mathbf{D}\mathbf{D}^*) = rank\,(\mathbf{D}) = r$ (because $\mathbf{D}$ is nonsingular) to conclude that that $rank\,(\mathbf{A}^*\mathbf{A}\mathbf{A}^*) = rank\,(\mathbf{A})$.

3.5.8. $\mathbf{B}$ is hermitian, so it has real eigenvalues, and $\mathbf{B}^2 = \mathbf{B}^*\mathbf{B}$ has real and nonnegative eigenvalues. Hence $\lambda \in \sigma\left(\mathbf{B}^2\right) \iff \pm\sqrt{\lambda} \in \sigma\,(\mathbf{B})$. Consequently

$$\begin{aligned} \mathbf{B}^2 = \mathbf{B}^*\mathbf{B} &= \begin{pmatrix} \mathbf{A}\mathbf{A}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^*\mathbf{A} \end{pmatrix} \\ &\Longrightarrow \quad \sigma\left(\mathbf{B}^2\right) = \sigma\,(\mathbf{A}\mathbf{A}^*) \cup \sigma\,(\mathbf{A}^*\mathbf{A}) = \{\sigma_1^2, \sigma_2^2, \ldots \sigma_r^2, 0\} \\ &\Longrightarrow \quad \sigma\,(\mathbf{B}) = \{\pm\sigma_1, \pm\sigma_2, \ldots \pm\sigma_r, 0\}. \end{aligned}$$

3.5.9. The logic is identical to that in the proof of Theorem 3.5.2. Write

$$\mathbf{U}\begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^* = [\mathbf{U}_1\mathbf{D}\,|\,\mathbf{0}]\left(\begin{array}{c}\mathbf{D}^{-1}\mathbf{U}_1^*\mathbf{A}\\ \hline \mathbf{Y}_2^*\end{array}\right) = \mathbf{U}_1\mathbf{U}_1^*\mathbf{A} = \mathbf{A}.$$

The last equality here follows from (3.5.1) and (3.5.3) with $\mathbf{X}_2$ replaced by $\mathbf{U}_2$ because $\mathbf{A}^*\mathbf{U}_2 = \mathbf{0} \implies \mathbf{U}_2^*\mathbf{A} = \mathbf{0}$ so that

$$\mathbf{A} = \mathbf{I}\mathbf{A} = \mathbf{U}\mathbf{U}^*\mathbf{A} = (\mathbf{U}_1\mathbf{U}_1^* + \mathbf{U}_2\mathbf{U}_2^*)\mathbf{A} = \mathbf{U}_1\mathbf{U}_1^*\mathbf{A}.$$

To see that $\mathbf{V}$ is unitary, simply verify that $\mathbf{V}^*\mathbf{V} = \mathbf{I}$ by using $\mathbf{L} = \mathbf{U}_1^*\mathbf{A}\mathbf{A}^*\mathbf{U}_1$ from (3.5.1) (with $\mathbf{X}$ replaced by $\mathbf{U}$) along with (3.5.2) (with $\mathbf{V}$ replaced by $\mathbf{Y}$) and the fact that $\mathbf{Y}_2$ has orthonormal columns to write

$$\mathbf{V}^*\mathbf{V} = \begin{pmatrix}\mathbf{D}^{-1}\mathbf{U}_1^*\mathbf{A}\mathbf{A}^*\mathbf{U}_1\mathbf{D}^{-1} & \mathbf{D}^{-1}\mathbf{U}_1^*\mathbf{A}\mathbf{Y}_2\\ \mathbf{Y}_2^*\mathbf{A}^*\mathbf{U}_1\mathbf{D}^{-1} & \mathbf{Y}_2^*\mathbf{Y}_2\end{pmatrix} = \begin{pmatrix}\mathbf{D}^{-1}\mathbf{L}\mathbf{D}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix} = \begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}.$$

3.5.10. (a) Suppose $\mathbf{A}_{n\times n}$ is nonsingular, and let $\mathbf{E}_k = \mathbf{A}_k - \mathbf{A}$ so that $\lim_{k\to\infty}\mathbf{E}_k = \mathbf{0}$. This together with Theorem 3.5.9 on page 364 implies there exists a sufficiently large value of k such that

$$rank\,(\mathbf{A}_k) = rank\,(\mathbf{A} + \mathbf{E}_k) \geq rank\,(\mathbf{A}) = n,$$

which is impossible because each $\mathbf{A}_k$ is singular. Therefore, the supposition that $\mathbf{A}$ is nonsingular must be false.

(b) No—consider the sequence of 1×1 matrices $[1/k] \to [0]$.

3.5.11. Since $rank\,(\mathbf{A}) = 1$, there is exactly one nonzero singular value σ, and $\lambda = \sigma^2$ is the nonzero eigenvalue of $\mathbf{A}^*\mathbf{A}$. Since

$$\mathbf{A}^*\mathbf{A}\mathbf{v} = (\mathbf{u}\mathbf{v}^*)^*(\mathbf{u}\mathbf{v}^*)\mathbf{v} = (\mathbf{v}\mathbf{u}^*\mathbf{u}\mathbf{v}^*)\mathbf{v} = \|\mathbf{u}\|_2^2\,\|\mathbf{v}\|_2^2\,\mathbf{v},$$

it follows that $\lambda = \|\mathbf{u}\|_2^2\,\|\mathbf{v}\|_2^2$, so $\sigma = \sqrt{\lambda} = \|\mathbf{u}\|_2\,\|\mathbf{v}\|_2$, is the only nonzero singular value of $\mathbf{A}$, and thus $\|\mathbf{A}\|_2 = \sigma = \|\mathbf{u}\|_2\,\|\mathbf{v}\|_2$.

3.5.12. Simply expand an SVD by writing

$$\mathbf{A} = \mathbf{U}\begin{pmatrix}\mathbf{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}\mathbf{V}^* = [\mathbf{u}_1\cdots\mathbf{u}_r\cdots\mathbf{u}_m]\begin{pmatrix}\sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0\end{pmatrix}\begin{pmatrix}\mathbf{v}_1^*\\ \vdots\\ \mathbf{v}_r^*\\ \vdots\\ \mathbf{v}_n^*\end{pmatrix}$$

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

Solutions

3.5.13. If $\mathbf{A} = \begin{pmatrix} 3 & & \\ & 2 & \\ & & 1 \end{pmatrix}$, $\mathbf{B}_1 = \begin{pmatrix} 3 & & \\ & 0 & \\ & & 0 \end{pmatrix}$, and $\mathbf{B}_2 = \begin{pmatrix} 1 & & \\ & 0 & \\ & & 0 \end{pmatrix}$, then

$$\|\mathbf{A} - \mathbf{B}_1\|_2 = 2 = \|\mathbf{A} - \mathbf{B}_1\|_2.$$

In other words, $\mathbf{B}_1$ and $\mathbf{B}_2$ are both rank-one matrices that are a minimal distance from $\mathbf{A}$.

3.5.14. (a) Use the result in Exercise 2.4.43 on page 197 that says $(\mathbf{UXV})^\dagger = \mathbf{V}^*\mathbf{X}^\dagger\mathbf{U}^*$ for unitary $\mathbf{U}$ and $\mathbf{V}$ together with the result of Exercise 2.4.37 on page 196 to conclude that if $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ is an SVD of $\mathbf{A}$, then

$$\mathbf{A}^\dagger = \left(\mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*\right)^\dagger = \mathbf{V}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}^\dagger\mathbf{U}^* = \mathbf{V}\begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^* = \sum_{i=1}^{r}\sigma_i^{-1}\mathbf{v}_i\mathbf{u}_i^*.$$

(b) Straightforward multiplication does the job.

3.5.15. If $\mathbf{A}(x) = \begin{pmatrix} 1 & 0 \\ 0 & x \end{pmatrix}$, then $\mathbf{A}^\dagger(x) = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1/x \end{pmatrix} & \text{for } x \neq 0, \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \text{for } x = 0. \end{cases}$ This shows that not only is $\mathbf{A}^\dagger(x)$ discontinuous because $\lim_{x\to 0}\mathbf{A}^\dagger(x) \neq \mathbf{A}^\dagger(0)$, but it is discontinuous in the worst way because as $\mathbf{A}(x)$ comes closer to $\mathbf{A}(0)$ the matrix $\mathbf{A}^\dagger(x)$ moves farther away from $\mathbf{A}^\dagger(0)$. This type of behavior translates into insurmountable computational difficulties because small errors due to roundoff (or anything else) can produce enormous errors in the computed $\mathbf{A}^\dagger$, and as errors in $\mathbf{A}$ become smaller the resulting errors in $\mathbf{A}^\dagger$ can become greater. The inherent numerical problems coupled with the fact that it is rare for an application to require explicit knowledge of the entries of $\mathbf{A}^\dagger$ constrains the pseudo inverse to being a strictly theoretical or notational tool. But don't underestimate this role—go back and read Laplace's statement quoted on page 2.

3.5.16. If $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^T$ is an SVD, then $\mathbf{A}^T\mathbf{A} + \epsilon\mathbf{I} = \mathbf{U}\begin{pmatrix} \mathbf{D}^2 + \epsilon\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \epsilon\mathbf{I} \end{pmatrix}\mathbf{V}^T$ is an SVD with no zero singular values, so it is nonsingular. Furthermore,

$$(\mathbf{A}^T\mathbf{A} + \epsilon\mathbf{I})^{-1}\mathbf{A}^T = \mathbf{U}\begin{pmatrix} (\mathbf{D}^2 + \epsilon\mathbf{I})^{-1}\mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^T \to \mathbf{U}\begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^T = \mathbf{A}^\dagger.$$

3.5.17. If $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$, where $\mathbf{U}_{m\times m} = [\underbrace{\mathbf{U}_1}_{m\times r} \mid \underbrace{\mathbf{U}_2}_{m\times(m-r)}]$, $\mathbf{V}_{n\times n} = [\underbrace{\mathbf{V}_1}_{n\times r} \mid \underbrace{\mathbf{V}_2}_{n\times(n-r)}]$, and $\mathbf{D} = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$, then $\mathbf{A}[\mathbf{V}_1 \,|\, \mathbf{V}_2] = [\mathbf{U}_1\mathbf{D} \,|\, \mathbf{0}] \Rightarrow \mathbf{AV}_2 = \mathbf{0}$. If $\|\mathbf{x}\|_2 = 1$ and $\mathbf{x} \perp N(\mathbf{A})$, then, in particular, $\mathbf{V}_2^*\mathbf{x} = \mathbf{0}$. Furthermore,

$$\|\mathbf{Ax}\|_2 = \|\mathbf{U}_1\mathbf{DV}_1^*\mathbf{x}\|_2 = \|\mathbf{DV}_1^*\mathbf{x}\|_2 = \|\mathbf{Dy}\|_2, \quad \text{where} \quad \mathbf{y} = \mathbf{V}_1^*\mathbf{x}.$$

Since $\mathbf{I} = \mathbf{V}_1\mathbf{V}_1^* + \mathbf{V}_2\mathbf{V}_2^*$, it follows that

$$1 = \|\mathbf{x}\|_2 = \|\mathbf{Ix}\|_2 = \|\mathbf{V}_1\mathbf{V}_1^*\mathbf{x}\|_2 = \|\mathbf{V}_1\mathbf{y}\|_2 = \|\mathbf{y}\|_2 \quad \text{(by (3.5.17) on page 361)}.$$

Consequently,

$$\min_{\substack{\|\mathbf{x}\|_2=1 \\ \mathbf{x}\perp N(\mathbf{A})}} \|\mathbf{Ax}\|_2 = \min_{\|\mathbf{y}\|_2=1} \|\mathbf{Dy}\|_2 = \sigma_r \quad \text{(by (3.5.10) on page 359).}$$

To prove $\sigma_r = 1/\left\|\mathbf{A}^\dagger\right\|_2$, use $\mathbf{A}^\dagger_{n\times m} = \mathbf{V}\begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^*$ (from Exercise 3.5.14) together with (3.5.12) on page 360 to write $\left\|\mathbf{A}^\dagger\right\|_2 = \max_{\sigma_i\neq 0}\{1/\sigma_i\} = 1/\sigma_r$.

3.5.18. $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$ and $\tilde{\mathbf{x}} = \mathbf{A}^\dagger(\mathbf{b}-\mathbf{e})$ are the respective solutions of minimal 2-norm of $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}} = \mathbf{b}-\mathbf{e}$ by Exercise 2.4.35 on page 196. The development of the more general bounds is essentially the same as in the proof of Theorem 3.5.13.

$$\begin{aligned} &\|\mathbf{x}-\tilde{\mathbf{x}}\|_2 = \|\mathbf{A}^\dagger(\mathbf{b}-\tilde{\mathbf{b}})\|_2 \le \|\mathbf{A}^\dagger\|_2\,\|\mathbf{b}-\tilde{\mathbf{b}}\|_2, \\ &\mathbf{b} = \mathbf{Ax} \implies \|\mathbf{b}\|_2 \le \|\mathbf{A}\|_2\,\|\mathbf{x}\|_2 \implies 1/\|\mathbf{x}\|_2 \le \|\mathbf{A}\|_2/\|\mathbf{b}\|_2, \end{aligned}$$

so

$$\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \le \left(\|\mathbf{A}^\dagger\|_2\,\|\mathbf{b}-\tilde{\mathbf{b}}\|_2\right)\frac{\|\mathbf{A}\|_2}{\|\mathbf{b}\|_2} = \kappa_2\frac{\|\mathbf{e}\|_2}{\|\mathbf{b}\|_2}.$$

Similarly,

$$\begin{aligned} &\|\mathbf{b}-\tilde{\mathbf{b}}\|_2 = \|\mathbf{A}(\mathbf{x}-\tilde{\mathbf{x}})\|_2 \le \|\mathbf{A}\|_2\,\|\mathbf{x}-\tilde{\mathbf{x}}\|_2, \\ &\mathbf{x} = \mathbf{A}^\dagger\mathbf{b} \implies \|\mathbf{x}\|_2 \le \|\mathbf{A}^\dagger\|_2\,\|\mathbf{b}\|_2 \implies 1/\|\mathbf{b}\|_2 \le \|\mathbf{A}^\dagger\|_2/\|\mathbf{x}\|_2, \end{aligned}$$

so

$$\frac{\|\mathbf{b}-\tilde{\mathbf{b}}\|_2}{\|\mathbf{b}\|_2} \le \left(\|\mathbf{A}\|_2\,\|\mathbf{x}-\tilde{\mathbf{x}}\|_2\right)\frac{\|\mathbf{A}^\dagger\|_2}{\|\mathbf{x}\|_2} = \kappa_2\frac{\|\mathbf{x}-\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2}.$$

Equality was attained in the proof of Theorem 3.5.14 by choosing $\mathbf{b}$ and $\mathbf{e}$ to point in special directions. But for these choices $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{A}\tilde{\mathbf{x}} = \tilde{\mathbf{b}} = \mathbf{b}-\mathbf{e}$ may not be consistent for all singular or rectangular matrices $\mathbf{A}$, so the answer to the second part is "no." However, if $\mathbf{AA}^\dagger = \mathbf{I}$ (i.e., if $rank\,(\mathbf{A}_{m\times n}) = m$), then the argument in the proof of Theorem 3.5.14 proves equality for all $\mathbf{A}$.

3.5.19. Exercise 2.4.35 on page 196 shows that the minimal norm solution is $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$, so the result follows from the previous exercise—i.e., Exercise 3.5.14.

3.5.20. Exercise 2.4.34 on page 196 shows that $\mathbf{x} = \mathbf{A}^\dagger\mathbf{b}$ is one solution of the normal equations $\mathbf{A}^*\mathbf{Ax} = \mathbf{A}^*\mathbf{b}$, so the result follows from Exercise 3.5.14.

3.5.21. If $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ with $\mathbf{D} = \text{diag}\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$ is an SVD for $\mathbf{A}$, then $\mathbf{A}^*\mathbf{A} = \mathbf{V}\begin{pmatrix} \mathbf{D}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ is an SVD for $\mathbf{A}^*\mathbf{A}$—i.e., if $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0$ are the nonzero singular values for $\mathbf{A}$, then $\sigma_1^2 \ge \sigma_2^2 \ge \cdots \ge \sigma_r^2 > 0$ are the nonzero singular values for $\mathbf{A}^*\mathbf{A}$. Therefore, the respective (generalized) condition numbers for $\mathbf{A}$ and $\mathbf{A}^*\mathbf{A}$ are

$$\kappa_2 = \frac{\sigma_1}{\sigma_r} = \|\mathbf{A}\|_2\left\|\mathbf{A}^\dagger\right\|_2 \quad \text{and} \quad \kappa_2^2 = \frac{\sigma_1^2}{\sigma_r^2} = \|\mathbf{A}\|_2^2\left\|\mathbf{A}^\dagger\right\|_2^2.$$

Solutions

3.5.22. (a) The Frobenius norm is unitarily invariant (by Theorem 1.10.3 on page 92), so if $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*$ is an SVD, then

$$\|\mathbf{A}\|_F^2 = \left\|\mathbf{U}\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^*\right\|_F^2 = \left\|\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right\|_F^2 = \|\mathbf{D}\|_F^2 = \sum_{i=1}^{r} \sigma_i^2.$$

(b) $\|\mathbf{A}\|_2^2 = \sigma_1^2 \le \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2 = \|\mathbf{A}\|_F^2 \le n\sigma_1^2 = n\|\mathbf{A}\|_2^2$.

(c) For $\mathbf{x} \in \mathbb{F}^{n\times 1}$, Exercise 1.5.4 on page 37 shows that $\|\mathbf{x}\|_i \le \alpha \|\mathbf{x}\|_j$, where α is the (i,j)-entry in the matrix $\begin{array}{c} \\ 1 \\ 2 \\ \infty \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & \infty \end{array} \\ \begin{pmatrix} * & \sqrt{n} & n \\ 1 & * & \sqrt{n} \\ 1 & 1 & * \end{pmatrix} \end{array}$, and the proof of Theorem 1.9.6 on page 86 shows that $\|\mathbf{A}\|_1 = \|\mathbf{A}\mathbf{e}_k\|_1$ for some k, $\|\mathbf{A}\|_\infty = \|\mathbf{A}\mathbf{x}\|_\infty$, where the entries in $\mathbf{x}$ are ± 1, and $\|\mathbf{A}\|_2 = \|\mathbf{A}\mathbf{x}\|_2$ for some $\mathbf{x}$ with $\|\mathbf{x}\|_2 = 1$. Use these to reach the following conclusions.

$$\begin{aligned}
(1,2): &\quad \|\mathbf{A}\|_1 = \|\mathbf{A}\mathbf{e}_k\|_1 \le \sqrt{n}\,\|\mathbf{A}\mathbf{e}_k\|_2 \le \sqrt{n}\,\|\mathbf{A}\|_2\,\|\mathbf{e}_k\|_2 = \sqrt{n}\,\|\mathbf{A}\|_2 \\
(1,\infty): &\quad \|\mathbf{A}\|_1 = \|\mathbf{A}\mathbf{e}_k\|_1 \le n\,\|\mathbf{A}\mathbf{e}_k\|_\infty \le n\,\|\mathbf{A}\|_\infty\,\|\mathbf{e}_k\|_\infty = n\,\|\mathbf{A}\|_\infty \\
(1,F): &\quad \|\mathbf{A}\|_1 = \|\mathbf{A}\mathbf{e}_k\|_1 \le \sqrt{n}\,\|\mathbf{A}\mathbf{e}_k\|_2 \le \sqrt{n}\,\|\mathbf{A}\|_F\,\|\mathbf{e}_k\|_2 = \sqrt{n}\,\|\mathbf{A}\|_F \\
(2,1): &\quad \|\mathbf{A}\|_2 = \|\mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{A}\mathbf{x}\|_1 \le \|\mathbf{A}\|_1\,\|\mathbf{x}\|_1 \le \|\mathbf{A}\|_1\sqrt{n}\,\|\mathbf{x}\|_2 = \|\mathbf{A}\|_1\sqrt{n} \\
(2,\infty): &\quad \|\mathbf{A}\|_2 = \|\mathbf{A}\mathbf{x}\|_2 \le \sqrt{n}\,\|\mathbf{A}\mathbf{x}\|_\infty \le \sqrt{n}\,\|\mathbf{A}\|_\infty\,\|\mathbf{x}\|_\infty \le \sqrt{n}\,\|\mathbf{A}\|_\infty\,\|\mathbf{x}\|_2 \\
&\qquad = \sqrt{n}\,\|\mathbf{A}\|_\infty \\
(2,F): &\quad \text{(see the solution to part b)} \\
(\infty,1): &\quad \|\mathbf{A}\|_\infty = \|\mathbf{A}\mathbf{x}\|_\infty \le \|\mathbf{A}\mathbf{x}\|_1 \le \|\mathbf{A}\|_1\,\|\mathbf{x}\|_1 = \|\mathbf{A}\|_1\,n \\
(\infty,2): &\quad \|\mathbf{A}\|_\infty = \|\mathbf{A}\mathbf{x}\|_\infty \le \|\mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{A}\|_2\,\|\mathbf{x}\|_2 = \|\mathbf{A}\|_2\sqrt{n} \\
(\infty,F): &\quad \|\mathbf{A}\|_\infty = \|\mathbf{A}\mathbf{x}\|_\infty \le \|\mathbf{A}\mathbf{x}\|_2 \le \|\mathbf{A}\|_F\,\|\mathbf{x}\|_2 = \|\mathbf{A}\|_F\sqrt{n} \\
(F,1): &\quad \|\mathbf{A}\|_F^2 = \sum_j \|\mathbf{A}_{*j}\|_2^2 \le \sum_j \|\mathbf{A}_{*j}\|_1^2 \le n \max_j \|\mathbf{A}_{*j}\|_1^2 = n\,\|\mathbf{A}\|_1^2 \\
(F,2): &\quad \text{(see the solution to part b)} \\
(F,\infty): &\quad \|\mathbf{A}\|_F = \|\mathbf{A}^*\|_F \le \sqrt{n}\,\|\mathbf{A}^*\|_1 = \sqrt{n}\,\|\mathbf{A}\|_\infty
\end{aligned}$$

3.5.23. You just need to go through the proofs and verify that nothing is lost by replacing $\|*\|_2$ by another norm $\|*\|_\star$ for which $\|\mathbf{I}\|_\star = 1$.

3.5.24. These results are a corollary of Ky Fan's theorem (Theorem 3.4.10 on page 344) because the eigenvalues of $\mathbf{A}^*\mathbf{A}$ are the squares of the singular values of $\mathbf{A}$, so

$$\|\mathbf{A}\mathbf{X}\|_F^2 = trace\,((\mathbf{A}\mathbf{X})^*(\mathbf{A}\mathbf{X})) = trace\,(\mathbf{X}^*\mathbf{A}^*\mathbf{A}\mathbf{X})$$

$$\Longrightarrow \max_{\mathbf{X}^T\mathbf{X}=\mathbf{I}_k} \|\mathbf{A}\mathbf{X}\|_F = \max_{\mathbf{X}^T\mathbf{X}=\mathbf{I}_k} trace\,(\mathbf{X}^*\mathbf{A}^*\mathbf{A}\mathbf{X}) = \sum_{j=1}^{k} \sigma_j^2.$$

The min part of the statement follows in a similar manner. ∎

3.5.25. (a) Consider $\mathbf{A} = \begin{pmatrix} \epsilon & -1 \\ 1 & 0 \end{pmatrix}$ or $\mathbf{A} = \begin{pmatrix} \epsilon & \epsilon^n \\ 0 & \epsilon \end{pmatrix}$ for small $\epsilon \neq 0$.

(b) For $\alpha > 1$, consider

$$\mathbf{A} = \begin{pmatrix} 1 & -\alpha & 0 & \cdots & 0 \\ 0 & 1 & -\alpha & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -\alpha \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{n \times n} \quad \text{and } \mathbf{A}^{-1} = \begin{pmatrix} 1 & \alpha & \cdots & \alpha^{n-2} & \alpha^{n-1} \\ 0 & 1 & \cdots & \alpha^{n-3} & \alpha^{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \alpha \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

Regardless of which norm is used, $\|\mathbf{A}\| > \alpha$ and $\left\|\mathbf{A}^{-1}\right\| > \alpha^{n-1}$, so $\kappa > \alpha^n$ exhibits exponential growth. Even for moderate values of n and $\alpha > 1$, κ can be quite large.

3.5.26. The argument is almost identical to that given for the nonsingular case except that now you have

$$\mathbf{U}^T\big(\mathbf{A}(\mathcal{S}_2)\big) = \left\{ \mathbf{w} \,\middle|\, \mathbf{w} = \mathbf{U}^T\mathbf{A}\mathbf{x} = \begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^T\mathbf{x} \quad \text{for } \|\mathbf{x}\|_2 = 1 \right\},$$

so the coordinates of $\mathbf{w} \in \mathbf{U}^T\big(\mathbf{A}(\mathcal{S}_2)\big)$ obey the equation

$$\frac{w_1^2}{\sigma_1^2} + \frac{w_2^2}{\sigma_2^2} + \cdots + \frac{w_r^2}{\sigma_r^2} = \left\| \begin{pmatrix} \mathbf{D}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{w} \right\|_2^2 = \left\| \mathbf{V}^T\mathbf{x} \right\|_2^2 = \|\mathbf{x}\|_2^2 = 1,$$

and thus $\mathbf{U}^T\mathbf{A}(\mathcal{S}_2)$ is an ellipsoid (degenerate if $r < n$) whose k^{th} semi-axis is $\sigma_k \mathbf{e}_k$, which means that $\mathbf{A}(\mathcal{S}_2)$ is an ellipsoid (possibly degenerate) whose k^{th} semi-axis is $\mathbf{U}\sigma_k\mathbf{e}_k = \sigma_k\mathbf{u}_k$ for $1 \leq k \leq r$.

Solutions for exercises in section 3. 6

3.6.1. $\mathbf{A}$ and $\mathbf{C}$ are positive definite while $\mathbf{B}$ is positive semidefinite. Computing eigenvalues would show this, but it is easier to use Theorem 2.10.7 on page 270 to check the pivots in the $\mathrm{LDL^T}$ factorization.

3.6.2. They are the same. Recall Exercise 3.5.5 on page 374.

3.6.3. Each diagonal entry is a 1×1 principal submatrix, so Theorems 3.6.4 and 3.6.3 apply. Or you could just say $0 \leq \mathbf{e}_i^T\mathbf{A}\mathbf{e}_i$ when $\mathbf{A}$ is positive semidefinite and $0 < \mathbf{e}_i^T\mathbf{A}\mathbf{e}_i$ when $\mathbf{A}$ is positive definite.

3.6.4. Use Theorem 3.6.1 and the fact that $\lambda \in \sigma(\mathbf{A})$ if and only if $\lambda^{-1} \in \sigma\left(\mathbf{A}^{-1}\right)$.

3.6.5. In either case, $\mathbf{A} = \mathbf{B}^*\mathbf{B}$ by Theorem 3.6.1, so

$$\det(\mathbf{A}) = \det(\mathbf{B}^*\mathbf{B}) = \det(\mathbf{B}^*)\det(\mathbf{B}) = \overline{\det(\mathbf{B})}\det(\mathbf{B}) = \left|\det(\mathbf{B})\right|^2 \geq 0.$$

When $\mathbf{A}$ is positive definite, $\mathbf{B}$ is nonsingular, and thus $\det(\mathbf{A}) > 0$.