Orthogonality

A set of vectors \( \{x_1, x_2, \ldots, x_k\} \) in \( \mathbb{R}^n \) is said to be orthogonal if \( x_i^T x_j = 0 \) whenever \( i \neq j \). If, in addition, \( x_i^T x_i = 1 \) for \( i = 1, 2, \ldots, k \), then we say that the set is orthonormal. In this case, we can also write \( x_i^T x_j = \delta_{ij} \), where

\[
\delta_{ij} = \begin{cases} 
1 & \text{if } i = j \\
0 & \text{if } i \neq j
\end{cases}
\]

is called the Kronecker delta.

**Example** The vectors \( x = [1 \ -3 \ 2]^T \) and \( y = [4 \ 2 \ 1]^T \) are orthogonal, as \( x^T y = 1(4) - 3(2) + 2(1) = 4 - 6 + 2 = 0 \). Using the fact that \( \|x\|_2 = \sqrt{x^T x} \), we can normalize these vectors to obtain orthonormal vectors

\[
\tilde{x} = \frac{x}{\|x\|_2} = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \quad \tilde{y} = \frac{y}{\|y\|_2} = \frac{1}{\sqrt{21}} \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix},
\]

which satisfy \( \tilde{x}^T \tilde{x} = \tilde{y}^T \tilde{y} = 1 \), \( \tilde{x}^T \tilde{y} = 0 \). \( \square \)

The concept of orthogonality extends to subspaces. If \( S_1, S_2, \ldots, S_k \) is a set of subspaces of \( \mathbb{R}^m \), we say that these subspaces are mutually orthogonal if \( x^T y = 0 \) whenever \( x \in S_i \), \( y \in S_j \), for \( i \neq j \). Furthermore, we define the orthogonal complement of a subspace \( S \subseteq \mathbb{R}^n \) to be the set \( S^\perp \) defined by

\[
S^\perp = \{ y \in \mathbb{R}^m \mid x^T y = 0, x \in S \}.
\]

**Example** Let \( S \) be the subspace of \( \mathbb{R}^3 \) defined by \( S = \text{span}\{v_1, v_2\} \), where

\[
v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.
\]

Then, \( S \) has the orthogonal complement \( S^\perp = \text{span}\{[1 \ 0 \ -1]^T\} \).

It can be verified directly, by computing inner products, that the single basis vector of \( S^\perp \), \( v_3 = [1 \ 0 \ -1]^T \), is orthogonal to either basis vector of \( S \), \( v_1 \) or \( v_2 \). It follows that any
multiple of this vector, which describes any vector in the span of \( v_3 \), is orthogonal to any linear combination of these basis vectors, which accounts for any vector in \( S \), due to the linearity of the inner product:

\[
(c_3 v_3)^T (c_1 v_1 + c_2 v_2) = c_3 c_1 v_3^T v_1 + c_3 c_2 v_3^T v_2 = c_3 c_3(0) + c_3 c_2(0) = 0.
\]

Therefore, any vector in \( S^\perp \) is in fact orthogonal to any vector in \( S \). □

It is helpful to note that given a basis \( \{ v_1, v_2 \} \) for any two-dimensional subspace \( S \subseteq \mathbb{R}^3 \), a basis \( \{ v_3 \} \) for \( S^\perp \) can be obtained by computing the cross-product of the basis vectors of \( S \),

\[
v_3 = v_1 \times v_2 = \begin{bmatrix} (v_1)_2(v_2)_3 - (v_1)_3(v_2)_2 \\ (v_1)_3(v_2)_1 - (v_1)_1(v_2)_3 \\ (v_1)_1(v_2)_2 - (v_1)_2(v_2)_1 \end{bmatrix}.
\]

A similar process can be used to compute the orthogonal complement of any \((n - 1)\)-dimensional subspace of \( \mathbb{R}^n \). This orthogonal complement will always have dimension 1, because the sum of the dimension of any subspace of a finite-dimensional vector space, and the dimension of its orthogonal complement, is equal to the dimension of the vector space. That is, if \( \dim(V) = n \), and \( S \subseteq V \) is a subspace of \( V \), then

\[
\dim(S) + \dim(S^\perp) = n.
\]

Let \( S = \text{range}(A) \) for an \( m \times n \) matrix \( A \). By the definition of the range of a matrix, if \( y \in S \), then \( y = Ax \) for some vector \( x \in \mathbb{R}^n \). If \( z \in \mathbb{R}^m \) is such that \( z^T y = 0 \), then we have

\[
z^T A x = z^T (A^T)^T x = (A^T z)^T x = 0.
\]

If this is true for all \( y \in S \), that is, if \( z \in S^\perp \), then we must have \( A^T z = 0 \). Therefore, \( z \in \text{null}(A^T) \), and we conclude that \( S^\perp = \text{null}(A^T) \). That is, the orthogonal complement of the range is the null space of the transpose.

**Example** Let

\[
A = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ 3 & 2 \end{bmatrix}.
\]

If \( S = \text{range}(A) = \text{span}\{a_1, a_2\} \), where \( a_1 \) and \( a_2 \) are the columns of \( A \), then \( S^\perp = \text{span}\{b\} \), where \( b = a_1 \times a_2 = [7 \ 1 \ -3]^T \). We then note that

\[
A^T b = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} 7 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
\]
That is, $b$ is in the null space of $A^T$. In fact, because of the relationships
\[
\text{rank}(A^T) + \text{dim}(\text{null}(A^T)) = 3,
\]
\[
\text{rank}(A) = \text{rank}(A^T) = 2,
\]
we have $\text{dim}(\text{null}(A^T)) = 1$, so $\text{null}(A) = \text{span}\{b\}$. □

A basis for a subspace $S \subseteq \mathbb{R}^m$ is an orthonormal basis if the vectors in the basis form an orthonormal set. For example, let $Q_1$ be an $m \times r$ matrix with orthonormal columns. Then these columns form an orthonormal basis for $\text{range}(Q_1)$. If $r < m$, then this basis can be extended to a basis of all of $\mathbb{R}^m$ by obtaining an orthonormal basis of $\text{range}(Q_1)^\perp = \text{null}(Q_1^T)$.

If we define $Q_2$ to be a matrix whose columns form an orthonormal basis of $\text{range}(Q_1)^\perp$, then the columns of the $m \times m$ matrix $Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$ are also orthonormal. It follows that $Q^T Q = I$. That is, $Q^T = Q^{-1}$. Furthermore, because the inverse of a matrix is unique, $QQ^T = I$ as well. We say that $Q$ is an orthogonal matrix.

One interesting property of orthogonal matrices is that they preserve certain norms. For example, if $Q$ is an orthogonal $n \times n$ matrix, then $\|Qx\|_2^2 = x^T Q^T Q x = x^T x = \|x\|_2^2$. That is, orthogonal matrices preserve $\ell_2$-norms. Geometrically, multiplication by an $n \times n$ orthogonal matrix preserves the length of a vector, and performs a rotation in $n$-dimensional space.

**Example** The matrix
\[
Q = \begin{bmatrix}
\cos \theta & 0 & \sin \theta \\
0 & 1 & 0 \\
-\sin \theta & 0 & \cos \theta
\end{bmatrix}
\]
is an orthogonal matrix, for any angle $\theta$. For any vector $x \in \mathbb{R}^3$, the vector $Qx$ can be obtained by rotating $x$ clockwise by the angle $\theta$ in the $xz$-plane, while $Q^T = Q^{-1}$ performs a counterclockwise rotation by $\theta$ in the $xz$-plane. □

In addition, orthogonal matrices preserve the matrix $\ell_2$ and Frobenius norms. That is, if $A$ is an $m \times n$ matrix, and $Q$ and $Z$ are $m \times m$ and $n \times n$ orthogonal matrices, respectively, then
\[
\|QAZ\|_F = \|A\|_F, \quad \|QAZ\|_2 = \|A\|_2.
\]

**The Singular Value Decomposition**

The matrix $\ell_2$-norm can be used to obtain a highly useful decomposition of any $m \times n$ matrix $A$. Let $x$ be a unit $\ell_2$-norm vector (that is, $\|x\|_2 = 1$ such that $\|Ax\|_2 = \|A\|_2$. If $z = Ax$, and we define $y = z/\|z\|_2$, then we have $Ax = \sigma y$, with $\sigma = \|A\|_2$, and both $x$ and $y$ are unit vectors.
We can extend $x$ and $y$, separately, to orthonormal bases of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, to obtain orthogonal matrices $V_1 = \begin{bmatrix} x & V_2 \end{bmatrix} \in \mathbb{R}^{n \times n}$ and $U_1 = \begin{bmatrix} y & U_2 \end{bmatrix} \in \mathbb{R}^{m \times m}$. We then have

$$U_1^T AV_1 = \begin{bmatrix} y^T U_2^T \\ U_2^T A x \\ U_2^T A y \end{bmatrix} = \begin{bmatrix} \sigma y^T y \\ \sigma U_2^T y \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma w^T \\ 0 \\ B \end{bmatrix} = A_1,$$

where $B = U_2^T A V_2$ and $w = V_2^T A y$. Now, let

$$s = \begin{bmatrix} \sigma \\ w \end{bmatrix}.$$

Then $\|s\|^2_2 = \sigma^2 + \|w\|^2_2$. Furthermore, the first component of $A_1 s$ is $\sigma^2 + \|w\|^2_2$.

It follows from the invariance of the matrix $\ell_2$-norm under orthogonal transformations that

$$\sigma^2 = \|A\|^2_2 = \|A_1\|^2_2 \geq \frac{\|A_1 s\|^2_2}{\|s\|^2_2} \geq \frac{(\sigma^2 + \|w\|^2_2)^2}{\sigma^2 + \|w\|^2_2} = \sigma^2 + \|w\|^2_2,$$

and therefore we must have $w = 0$. We then have

$$U_1^T AV_1 = A_1 = \begin{bmatrix} \sigma & 0 \\ 0 & B \end{bmatrix}.$$

Continuing this process on $B$, and keeping in mind that $\|B\|_2 \leq \|A\|_2$, we obtain the decomposition

$$U^T AV = \Sigma$$

where

$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \in \mathbb{R}^{m \times m}, \quad V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \in \mathbb{R}^{n \times n}$$

are both orthogonal matrices, and

$$\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m \times n}, \quad p = \min\{m, n\}$$

is a diagonal matrix, in which $\Sigma_{ii} = \sigma_i$ for $i = 1, 2, \ldots, p$, and $\Sigma_{ij} = 0$ for $i \neq j$. Furthermore, we have

$$\|A\|_2 = \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0.$$
This decomposition of \( A \) is called the singular value decomposition, or SVD. It is more commonly written as a factorization of \( A \),

\[
A = U \Sigma V^T.
\]

The diagonal entries of \( \Sigma \) are the singular values of \( A \). The columns of \( U \) are the left singular vectors, and the columns of \( V \) are the right singular vectors. It follows from the SVD itself that the singular values and vectors satisfy the relations

\[
A v_i = \sigma_i u_i, \quad A^T u_i = \sigma_i v_i, \quad i = 1, 2, \ldots, \min\{m, n\}.
\]

For convenience, we denote the \( i \)th largest singular value of \( A \) by \( \sigma_i(A) \), and the largest and smallest singular values are commonly denoted by \( \sigma_{\text{max}}(A) \) and \( \sigma_{\text{min}}(A) \), respectively.

The SVD conveys much useful information about the structure of a matrix, particularly with regard to systems of linear equations involving the matrix. Let \( r \) be the number of nonzero singular values. Then \( r \) is the rank of \( A \), and

\[
\text{range}(A) = \text{span}\{u_1, \ldots, u_r\}, \quad \text{null}(A) = \text{span}\{v_{r+1}, \ldots, v_n\}.
\]

That is, the SVD yields orthonormal bases of the range and null space of \( A \).

It follows that we can write

\[
A = \sum_{i=1}^r \sigma_i u_i v_i^T.
\]

This is called the SVD expansion of \( A \). If \( m \geq n \), then this expansion yields the “economy-size” SVD

\[
A = U_1 \Sigma_1 V^T,
\]

where

\[
U_1 = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad \Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{n \times n}.
\]

**Example** The matrix

\[
A = \begin{bmatrix} 11 & 19 & 11 \\ 9 & 21 & 9 \\ 10 & 20 & 10 \end{bmatrix}
\]

has the SVD \( A = U \Sigma V^T \) where

\[
U = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -\sqrt{2/3} \end{bmatrix}, \quad V = \begin{bmatrix} 1/\sqrt{6} & -1/\sqrt{3} & 1/\sqrt{2} \\ \sqrt{2/3} & 1/\sqrt{3} & 0 \\ 1/\sqrt{6} & -1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix},
\]

and

\[
S = \begin{bmatrix} 42.4264 & 0 & 0 \\ 0 & 2.4495 & 0 \\ 0 & 0 & 0 \end{bmatrix}.
\]
Let $U = [\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}]$ and $V = [\begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}]$ be column partitions of $U$ and $V$, respectively. Because there are only two nonzero singular values, we have $\text{rank}(A) = 2$, Furthermore, $\text{range}(A) = \text{span}\{u_1, u_2\}$, and $\text{null}(A) = \text{span}\{v_3\}$. We also have

$$A = 42.4264 u_1 v_1^T + 2.4495 u_2 v_2^T.$$
The absolute error in this approximation is
\[ \| A - A_1 \|_2 = \sigma_2 \| u_2 v_2^T \|_2 = \sigma_2 = 2.4495, \]
while the relative error is
\[ \frac{\| A - A_1 \|_2}{\| A \|_2} = \frac{2.4495}{42.4264} = 0.0577. \]
That is, the rank-one approximation of \( A \) is off by a little less than six percent. □

Suppose that the entries of a matrix \( A \) have been determined experimentally, for example by measurements, and the error in the entries is determined to be bounded by some value \( \epsilon \). Then, if \( \sigma_k > \epsilon \geq \sigma_{k+1} \) for some \( k < \min\{m, n\} \), then \( \epsilon \) is at least as large as the distance between \( A \) and the set of all matrices of rank \( k \). Therefore, due to the uncertainty of the measurement, \( A \) could very well be a matrix of rank \( k \), but it cannot be of lower rank, because \( \sigma_k > \epsilon \). We say that the \( \epsilon \)-rank of \( A \), defined by
\[ \text{rank}(A, \epsilon) = \min_{\| A - B \|_2 \leq \epsilon} \text{rank}(B), \]
is equal to \( k \). If \( k < \min\{m, n\} \), then we say that \( A \) is \textit{numerically rank-deficient}.

**Projections**

Given a vector \( z \in \mathbb{R}^n \), and a subspace \( S \subseteq \mathbb{R}^n \), it is often useful to obtain the vector \( w \in S \) that best approximates \( z \), in the sense that \( z - w \in S^\perp \). In other words, the error in \( w \) is orthogonal to the subspace in which \( w \) is sought. To that end, we define the \textit{orthogonal projection} onto \( S \) to be an \( n \times n \) matrix \( P \) such that if \( w = Pz \), then \( w \in S \), and \( w^T (z - w) = 0 \).

Substituting \( w = Pz \) into \( w^T (z - w) = 0 \), and noting the commutativity of the inner product, we see that \( P \) must satisfy the relations
\[ z^T (P^T - P^T P)z = 0, \quad z^T (P - P^T P)z = 0, \]
for any vector \( z \in \mathbb{R}^n \). It follows from these relations that \( P \) must have the following properties: \( \text{range}(P) = S \), and \( P^T P = P = P^T \), which yields \( P^2 = P \). Using these properties, it can be shown that the orthogonal projection \( P \) onto a given subspace \( S \) is unique. Furthermore, it can be shown that \( P = VV^T \), where \( V \) is a matrix whose columns are an orthonormal basis for \( S \).

**Projections and the SVD**

Let \( A \in \mathbb{R}^{m \times n} \) have the SVD \( A = U \Sigma V^T \), where
\[ U = \begin{bmatrix} U_r & \bar{U}_r \end{bmatrix}, \quad V = \begin{bmatrix} V_r & \bar{V}_r \end{bmatrix}, \]
where \( r = \text{rank}(A) \) and
\[
U_r = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix}, \quad V_r = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix}.
\]

Then, we obtain the following projections related to the SVD:

- \( U_r U_r^T \) is the projection onto the range of \( A \)
- \( \tilde{V}_r \tilde{V}_r^T \) is the projection onto the null space of \( A \)
- \( \tilde{U}_r \tilde{U}_r^T \) is the projection onto the orthogonal complement of the range of \( A \), which is the null space of \( A^T \)
- \( V_r V_r^T \) is the projection onto the orthogonal complement of the null space of \( A \), which is the range of \( A^T \).

These projections can also be determined by noting that the SVD of \( A^T \) is \( A^T = V\Sigma U^T \).

**Example** For the matrix
\[
A = \begin{bmatrix} 11 & 19 & 11 \\ 9 & 21 & 9 \\ 10 & 20 & 10 \end{bmatrix},
\]
that has rank two, the orthogonal projection onto \( \text{range}(A) \) is
\[
U_2 U_2^T = \begin{bmatrix} u_1 u_2 \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \end{bmatrix} = u_1 u_1^T + u_2 u_2^T,
\]
where \( U = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \) is the previously defined column partition of the matrix \( U \) from the SVD of \( A \). \( \square \)

**Sensitivity of Linear Systems**

Let \( A \in \mathbb{R}^{m \times n} \) be nonsingular, and let \( A = U\Sigma V^T \) be the SVD of \( A \). Then the solution \( x \) of the system of linear equations \( Ax = b \) can be expressed as
\[
x = A^{-1}b = V\Sigma^{-1}U^T b = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.
\]

This formula for \( x \) suggests that if \( \sigma_n \) is small relative to the other singular values, then the system \( Ax = b \) can be sensitive to perturbations in \( A \) or \( b \). This makes sense, considering that \( \sigma_n \) is the distance between \( A \) and the set of all singular \( n \times n \) matrices.
In an attempt to measure the sensitivity of this system, we consider the parameterized system

\[(A + \epsilon E)x(\epsilon) = b + \epsilon e,\]

where \(E \in \mathbb{R}^{n \times n}\) and \(e \in \mathbb{R}^n\) are perturbations of \(A\) and \(b\), respectively. Taking the Taylor expansion of \(x(\epsilon)\) around \(\epsilon = 0\) yields

\[x(\epsilon) = x + \epsilon x'(0) + O(\epsilon^2),\]

where

\[x'(\epsilon) = (A + \epsilon E)^{-1}(e - Ex),\]

which yields \(x'(0) = A^{-1}(e - Ex)\).

Using norms to measure the relative error in \(x\), we obtain

\[
\frac{\|x(\epsilon) - x\|}{\|x\|} = |\epsilon| \frac{\|A^{-1}(e - Ex)\|}{\|x\|} + O(\epsilon^2) \leq |\epsilon|\|A^{-1}\| \left\{ \|e\| + \|E\| \right\} + O(\epsilon^2).
\]

Multiplying and dividing by \(\|A\|\), and using \(Ax = b\) to obtain \(\|b\| \leq \|A\|\|x\|\), yields

\[
\frac{\|x(\epsilon) - x\|}{\|x\|} = \kappa(A)|\epsilon| \left( \frac{\|e\|}{\|b\|} + \frac{\|E\|}{\|A\|} \right),
\]

where

\[
\kappa(A) = \|A\|\|A^{-1}\|
\]

is called the condition number of \(A\). We conclude that the relative errors in \(A\) and \(b\) can be amplified by \(\kappa(A)\) in the solution. Therefore, if \(\kappa(A)\) is large, the problem \(Ax = b\) can be quite sensitive to perturbations in \(A\) and \(b\). In this case, we say that \(A\) is ill-conditioned; otherwise, we say that \(A\) is well-conditioned.

The definition of the condition number depends on the matrix norm that is used. Using the \(\ell_2\)-norm, we obtain

\[
\kappa_2(A) = \|A\|_2\|A^{-1}\|_2 = \frac{\sigma_1(A)}{\sigma_n(A)}.
\]

It can readily be seen from this formula that \(\kappa_2(A)\) is large if \(\sigma_n\) is small relative to \(\sigma_1\). We also note that because the singular values are the lengths of the semi-axes of the hyperellipsoid \(\{Ax : \|x\|_2 = 1\}\), the condition number in the \(\ell_2\)-norm measures the elongation of this hyperellipsoid.

**Example** The matrices

\[A_1 = \begin{bmatrix} 0.7674 & 0.0477 \\ 0.6247 & 0.1691 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0.7581 & 0.1113 \\ 0.6358 & 0.0933 \end{bmatrix}\]

do not appear to be very different from one another, but \(\kappa_2(A_1) = 10\) while \(\kappa_2(A_2) = 10^{10}\). That is, \(A_1\) is well-conditioned while \(A_2\) is ill-conditioned.
To illustrate the ill-conditioned nature of $A_2$, we solve the systems $A_2x_1 = b_1$ and $A_2x_2 = b_2$ for the unknown vectors $x_1$ and $x_2$, where

\[
\begin{bmatrix}
0.7662 \\
0.6426
\end{bmatrix}, \quad \begin{bmatrix}
0.7019
\end{bmatrix}.
\]

These vectors differ from one another by roughly 10%, but the solutions

\[
\begin{bmatrix}
0.9894 \\
0.1452
\end{bmatrix}, \quad \begin{bmatrix}
-1.4522 \times 10^8 \\
9.8940 \times 10^8
\end{bmatrix}
\]

differ by several orders of magnitude, because of the sensitivity of $A_2$ to perturbations. □

Just as the largest singular value of $A$ is the $\ell_2$-norm of $A$, and the smallest singular value is the distance from $A$ to the nearest singular matrix in the $\ell_2$-norm, we have, for any $\ell_p$-norm,

\[
\frac{1}{\kappa_p(A)} = \min_{A+\Delta A \text{ singular}} \frac{\|\Delta A\|_p}{\|A\|_p}.
\]

That is, in any $\ell_p$-norm, $\kappa_p(A)$ measures the relative distance in that norm from $A$ to the set of singular matrices.

Because $\det(A) = 0$ if and only if $A$ is singular, it would appear that the determinant could be used to measure the distance from $A$ to the nearest singular matrix. However, this is generally not the case. It is possible for a matrix to have a relatively large determinant, but be very close to a singular matrix, or for a matrix to have a relatively small determinant, but not be nearly singular. In other words, there is very little correlation between $\det(A)$ and the condition number of $A$.

**Example** Let

\[
A = \begin{bmatrix}
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
0 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 1 & -1 & -1 & -1 & -1 & -1 \\
0 & 0 & 0 & 1 & -1 & -1 & -1 & -1 \\
0 & 0 & 0 & 0 & 1 & -1 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

Then $\det(A) = 1$, but $\kappa_2(A) \approx 1,918$, and $\sigma_{10} \approx 0.0029$. That is, $A$ is quite close to a singular matrix, even though $\det(A)$ is not near zero. For example, the nearby matrix $\tilde{A} = A - \sigma_{10}u_{10}v_{10}^T$. 

10
which has entries

\[
\hat{A} \approx \begin{bmatrix}
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-0 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-0 & -0 & 1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-0 & -0 & -0 & 1 & -1 & -1 & -1 & -1 & -1 \\
-0.0001 & -0.0001 & -0 & -0 & -0 & 1 & -1 & -1 & -1 & -1 \\
-0.0003 & -0.0001 & -0.0001 & -0 & -0 & -0 & 1 & -1 & -1 & -1 \\
-0.0005 & -0.0003 & -0.0001 & -0.0001 & -0 & -0 & -0 & 1 & -1 & -1 \\
-0.0011 & -0.0005 & -0.0003 & -0.0001 & -0.0001 & -0 & -0 & -0 & 1 & -1 \\
-0.0022 & -0.0011 & -0.0005 & -0.0003 & -0.0001 & -0.0001 & -0 & -0 & -0 & 1
\end{bmatrix},
\]

is singular. □