Chapter 4 – Singular Value Decomposition

For \( A \in \mathbb{R}^{m \times n} \) with \( m > n \) and \( \text{rank}(A) < n \), the QR decomposition may be difficult to use, since numerically \( A \) may “seem” to have rank \( n \).

Unlike earlier topics in this semester, we’ll save any discussion on how to compute the SVD until later on; for now, we’ll discuss what it is, and how to use it.

4.1 The SVD

We start with the main theoretical result of this section (this should remind you of your introductory class, where this was also presented).

**Theorem 4.1.1** (Singular Value Decomposition). Let \( A \in \mathbb{C}^{m \times n} \) be nonzero, and assume that \( \text{rank}(A) = r > 0 \). Then

\[
A = U \Sigma V^* \tag{4.1.1.1}
\]

where \( U \in \mathbb{C}^{m \times m} \) and \( V \in \mathbb{C}^{n \times n} \) are unitary matrices, \( \Sigma \in \mathbb{R}^{m \times n} \) is all zeros except on the diagonal,

\[
\Sigma = \begin{bmatrix}
\sigma_1 \\
\sigma_2 \\
\ddots \\
\sigma_r \\
0 \\
\ddots \\
0
\end{bmatrix}
\]

and where \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 \).

(proof in homework/exams)

Notes:

- This is called the **Singular Value Decomposition** of \( A \).
- The columns of \( U \) are orthonormal; these are called the **left singular vectors** of \( A \).
- The columns of \( V \) are orthonormal; these are called the **right singular vectors** of \( A \).
- The values \( \sigma_1, \ldots, \sigma_r \) are called the **singular values** of \( A \).
- \( U \) and \( V \) are not unique; however \( \Sigma \) is unique.
• $A^* = (UΣV^*)^* = VΣ^*U^* = VΣ^TU^*$, which is also a singular value decomposition since the outer matrices are again unitary and the inner matrix is diagonal with entries sorted from greatest → smallest down the diagonal (this time for $A^*$ instead of $A$). Hence $A$ and $A^*$ have the same singular values.

• WLOG, we may name $σ_{r+1} = σ_{r+2} = \cdots = σ_k = 0$ for $k = \min m, n$, giving

\[
(Σ)_{i,j} = \begin{cases} 
σ_i, & \text{if } i = j \\
0, & \text{otherwise} 
\end{cases}
\]

• rank($A$) equals the number of nonzero singular values of $A$.

**Definition 4.1.2** (Equivalent matrices). Two matrices $A$ and $B$ are called equivalent if $∃$ nonsingular matrices $X ∈ \mathbb{C}^{m×m}$ and $Y ∈ \mathbb{C}^{n×n}$ such that $A = XBY$.

• If $X$ and $Y$ are orthogonal, then $A$ and $B$ are called orthogonally equivalent.

• If $X$ and $Y$ are unitary, then $A$ and $B$ are called unitarily equivalent.

Hence Theorem 4.1.1 thus equivalently states that every $A ∈ \mathbb{R}^{m×n}$ is orthogonally equivalent to a “diagonal” matrix $Σ ∈ \mathbb{R}^{m×n}$. Similarly, the same theorem states that every $A ∈ \mathbb{C}^{m×n}$ is unitarily equivalent to a “diagonal” matrix $Σ ∈ \mathbb{R}^{m×n}$.

**Theorem 4.1.3** (Geometric SVD). Let $A ∈ \mathbb{C}^{m×n}$ be nonzero, and assume that rank($A$) = $r > 0$. Then $∃$ an orthonormal basis for $\mathbb{C}^n$, $\{v_1, v_2, \ldots, v_n\}$ and $∃$ and orthonormal basis for $\mathbb{C}^m$, $\{u_1, u_2, \ldots, u_m\}$, and $∃$ real values $σ_1 ≥ σ_2 ≥ \cdots ≥ σ_r > 0$ such that

\[
Av_k = \begin{cases} 
σ_ku_k, & 1 \leq k \leq r \\
0, & k > r 
\end{cases} \quad \text{and} \quad A^*u_k = \begin{cases} 
σ_kv_k, & 1 \leq k \leq r \\
0, & k > r 
\end{cases}
\]

(this is just a restatement of Theorem 4.1.1)

Geometric interpretation: if we consider the linear transformation $T : \mathbb{C}^n → \mathbb{C}^n$ defined via the matrix-vector product, $T(x) = Ax$, then Theorem 4.1.3 states that there exists orthonormal bases for $\mathbb{C}^m$ and $\mathbb{C}^n$ such that $T$ maps the $k$th basis vector in $\mathbb{C}^n$ to the $k$th basis vector in $\mathbb{C}^m$, modulo a positive scaling factor. Similarly for the linear transformation $S : \mathbb{C}^m → \mathbb{C}^n$ defined via $S(y) = A^*y$, that maps the $k$th basis vector in $\mathbb{C}^m$ to the $k$th basis vector in $\mathbb{C}^n$, modulo the same positive scaling factor.

Hence $Σ$ is the transformation matrix with respect to these orthonormal bases.

Revisiting the fundamental subspaces for a matrix, we therefore have:

\[
C(A) = \text{span}\{u_1, \ldots, u_r\}, \quad N(A^*) = \text{span}\{u_{r+1}, \ldots, u_m\} \\
C(A^*) = \text{span}\{v_1, \ldots, v_r\}, \quad N(A) = \text{span}\{v_{r+1}, \ldots, v_n\}.
\]
Since the columns of $U$ (and $V$) are orthonormal, then the first two of these immediately show that $(C(A))^\perp = N(A^*)$, and the latter two show that $(C(A^*))^\perp = N(A)$. Hence the Fundamental Theorem of Linear Algebra (Theorem 3.5.3) immediately follows from the SVD.

Question: since we only really care about the first $r$ diagonal values of $\Sigma$, and hence the first $r$ columns of $U$ and $V$, can we store only this relevant information?

**Theorem 4.1.4** (Condensed SVD). Let $A \in \mathbb{C}^{m \times n}$ be nonzero, and assume that $\text{rank}(A) = r > 0$. Then $\exists$ isometries $\hat{U} \in \mathbb{C}^{m \times r}$ and $\hat{V} \in \mathbb{C}^{n \times r}$, and a diagonal matrix $\hat{\Sigma} \in \mathbb{R}^{r \times r}$, where $(\hat{\Sigma})_{i,i} = \sigma_i$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and

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

**Proof.** Rewrite the standard SVD, $A = U \Sigma V^*$ in block form, and consider only the nonzero components. \qed

We have one more result regarding the SVD that is incredibly useful in applications.

**Theorem 4.1.5.** Let $A \in \mathbb{C}^{m \times n}$ be nonzero, and assume that $\text{rank}(A) = r > 0$. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ be the singular values of $A$, with corresponding left/right singular vectors $u_1, \ldots, u_r$ and $v_1, \ldots, v_r$, respectively. Then

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^*$$

(proof is left for homework/exams)

In other words, any matrix $A$ can be rewritten as the sum of $r$ rank-one matrices.

### 4.2 Applications of the SVD

Recall that back in chapter 2, we defined

$$\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2,$$

so $\|A\|_2$ represents the maximum magnification of any $x \in \mathbb{C}^n$ by multiplication with $A$ (in the 2-norm).

We call $\|A\|_2$ the “spectral norm” – it relates intimately with the SVD of $A$.

**Theorem 4.2.1.** Let $A \in \mathbb{C}^{m \times n}$ have singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then $\|A\|_2 = \sigma_1$. 

Proof. First, note that $Av_1 = u_1\sigma_1$ from Theorem 4.1.3, so since $\|v_1\|_2 = \|u_1\|_2 = 1$, then

$$\|Av_1\|_2 = \|\sigma_1 u_1\|_2 = \sigma_1 \|u_1\|_2 = \sigma_1.$$ 

Hence $\|A\|_2 \geq \sigma_1$.

Now let $x \in \mathbb{C}^n$. Then since $\{v_1, \ldots, v_n\}$ is a unitary basis for $\mathbb{C}^n$, we may write

$$x = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$ 

Moreover, since this basis is unitary, then

$$\|x\|_2^2 = \langle x, x \rangle = \left\langle \sum_{j=1}^{n} c_j v_j, \sum_{k=1}^{n} c_k v_k \right\rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} \langle v_j, v_k \rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j \overline{c_k} \delta_{j,k} = \sum_{j=1}^{n} |c_j|^2.$$ 

Also,

$$Ax = A \left( \sum_{k=1}^{n} c_k v_k \right) = \sum_{k=1}^{n} c_k (Av_k) = \sum_{k=1}^{r} c_k \sigma_k u_k$$

where $r = \text{rank}(A)$. Hence using a similar argument as above,

$$\|Ax\|_2^2 = \sum_{j=1}^{r} \sigma_j^2 |c_j|^2.$$ 

Moreover, since $\sigma_1 \geq \sigma_j \forall j$,

$$\|Ax\|_2^2 \leq \sum_{j=1}^{r} \sigma_1^2 |c_j|^2 = \sigma_1^2 \sum_{j=1}^{r} |c_j|^2 \leq \sigma_1^2 \sum_{j=1}^{n} |c_j|^2 = \sigma_1^2 \|x\|_2^2.$$ 

Therefore,

$$\frac{\|Ax\|_2}{\|x\|_2} = \left( \frac{\|Ax\|_2}{\|x\|_2} \right)^{1/2} \leq \left( \frac{\sigma_1^2 \|x\|_2^2}{\|x\|_2^2} \right)^{1/2} = \sigma_1$$

for arbitrary $x \in \mathbb{C}^n$. Hence $\|A\|_2 \leq \sigma_1$.

Since $\sigma_1 \leq \|A\|_2 \leq \sigma_1$, the theorem is proven. 

\[ \square \]

Corollary 4.2.2. Let $A \in \mathbb{C}^{m \times n}$. Then $\|A\|_2 = \|A^*\|_2$.

Proof. Since $A$ and $A^*$ have the same singular values, we merely cite the previous theorem. 

\[ \square \]
Recall from Theorem 4.1.3 that
\[ Av_1 = \sigma_1 u_1, \quad Av_2 = \sigma_2 u_2, \ldots, Av_r = \sigma_r u_r, \]
with each \( \sigma_j > 0 \). Hence if \( A^{-1} \) exists, then
\[ \sigma_1^{-1} v_1 = A^{-1} u_1, \quad \sigma_2^{-1} v_2 = A^{-1} u_2, \ldots, \sigma_r^{-1} v_r = A^{-1} u_r. \]
Moreover, for \( A^{-1} \) to exist, then \( A \in \mathbb{C}^{n \times n} \) (i.e., \( m = n \)) and \( \text{rank}(A) = n \), and hence
\[ \sigma_j^{-1} v_j = A^{-1} u_j, \quad j = 1, \ldots, n. \]
In other words, we immediately have a singular value decomposition for \( A^{-1} \):
\[ A = U \Sigma V^* \quad \Leftrightarrow \quad A^{-1} = (U \Sigma V^*)^{-1} = V^{-1} \Sigma^{-1} U^{-1} = V \Sigma^{-1} U^*, \]
since both \( U \) and \( V \) are unitary, i.e.
\[ U^{-1} = U^* \quad \text{and} \quad V^{-1} = V^* \quad \Leftrightarrow \quad V^{-1} = (V^*)^* = V. \]
Hence \( \|A^{-1}\|_2 = \frac{1}{\sigma_n} \) since
\[ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0 \quad \Leftrightarrow \quad \sigma_1^{-1} \geq \sigma_2^{-1} \geq \cdots \geq \sigma_n^{-1} > 0, \]
and therefore
\[ \kappa_2(A) = \frac{\sigma_1}{\sigma_n}. \]

**Theorem 4.2.3.** Let \( A \in \mathbb{C}^{n \times n} \) be nonsingular, with singular values \( \sigma_1 \geq \cdots \geq \sigma_n > 0 \). Then \( \kappa_2(A) = \frac{\sigma_1}{\sigma_n} \).
(proof given in preceding derivations)

Notes:
- We previously characterized \( \kappa_2(A) = \frac{\maxmag_2(A)}{\minmag_2(A)} \). Since we already have \( \maxmag_2(A) = \sigma_1 \), then clearly \( \minmag_2(A) = \sigma_n \).
- In this way, we can also consider a type of “condition number” for non-square matrices (in following theorem)!

**Theorem 4.2.4.** Let \( A \in \mathbb{C}^{m \times n} \) be nonzero, with \( m \geq n \) and with singular values \( \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0 \) (where here we some \( \sigma_j = 0 \) for \( \text{rank}(A) < n \)). Then \( \maxmag_2(A) = \sigma_1 \), \( \minmag_2(A) = \sigma_n \), and \( \kappa_2(A) = \frac{\sigma_1}{\sigma_n} \).

Note: the above theorem is actually how \texttt{cond} works in Matlab (and likely also in Python), allowing it to work with non-square matrices \( A \).
Theorem 4.2.5. Let $A \in \mathbb{C}^{m \times n}$ with $m \geq n$. Then $\|A^*A\|_2 = \|A\|_2^2$, and $\kappa_2(A^*A) = (\kappa_2(A))^2$.

(proof for homework/exams)

Hence if $\kappa_2(A) \gg 1$, then the normal equations, $A^*Ax = A^*b$ can be seriously ill-conditioned! In this case, you should definitely use some other algorithm for solving the least-squares problem (e.g., QR or SVD, the latter will be shown momentarily).

Theorem 4.2.6. Let $A \in \mathbb{C}^{m \times n}$ with $m \geq n$ and $\text{rank}(A) = n$, and assume that $A$ has singular values $\sigma_1 \geq \cdots \sigma_n > 0$. Then

\[
\| (A^*A)^{-1} \|_2 = \sigma_n^{-2},
\| (A^*A)^{-1} A^* \|_2 = \sigma_n^{-1},
\| A (A^*A)^{-1} \|_2 = \sigma_n^{-1}, \quad \text{and}
\| A (A^*A)^{-1} A^* \|_2 = 1.
\]

(proof for homework/exams)

Definition 4.2.7 (Pseudoinverse (Moore-Penrose inverse)). If a matrix $A \in \mathbb{C}^{m \times n}$ has full column rank (i.e., $\text{rank}(A) = n$), then the matrix $(A^*A)^{-1} A^*$ is called the pseudoinverse of $A$, and is frequently denoted as $A^+$ (a.k.a. A-dagger).

Similarly, the matrix $A (A^*A)^{-1}$ is the pseudoinverse of $A^*$.

Lemma 4.2.8. If $A \in \mathbb{C}^{n \times n}$ is nonsingular, then $A^+ = A^{-1}$.

Proof. If $A$ is nonsingular, then so is $A^*$. Hence we may distribute the inverse within the parentheses in the definition of $A^+$:

\[
A^+ = (A^*A)^{-1} A^* = (A^{-1} A^{-*}) A^* = A^{-1} (A^{-*} A^*) = A^{-1} I = A^{-1}.
\]

We now turn to the question of how to determine $\text{rank}(A)$ numerically. This is notoriously difficult, as we saw in performing Gaussian elimination, or in computing both the LU or QR factorizations of a matrix.

Given a tolerance $\varepsilon > 0$, we say that a matrix $A$ has numerical rank $k$ iff

\[
\sigma_1 \geq \cdots \geq \sigma_k > \varepsilon > \sigma_{k+1} \geq \cdots
\]

The choice of $\varepsilon$ should encode the level of uncertainty in the matrix entries themselves; in the Matlab function \texttt{rank} this tolerance is

\[
\varepsilon = 2^{-51} \max\{m,n\} \|A\|_2.
\]
We can similarly use singular values to answer a related question: out of all matrices with rank \( \leq k \), which matrix is the best approximation of a given matrix, \( A \)?

**Theorem 4.2.9.** Let \( A \in \mathbb{C}^{m \times n} \) have rank \( A = r > 0 \). Let \( A = U \Sigma V^* \) be the SVD of \( A \), with singular values \( \sigma_1 \geq \cdots \geq \sigma_r > 0 \). For any integer \( 1 \leq k \leq r - 1 \), define \( A_k = U \Sigma_k V^* \), where \( \Sigma_k \in \mathbb{R}^{m \times n} \) has diagonal entries \( \{\sigma_1, \ldots, \sigma_k, 0, \ldots, 0\} \). Then rank \( (A_k) = k \), and

\[
\sigma_{k+1} = \|A - A_k\|_2 = \min_{\text{rank}(B) \leq k} \|A - B\|_2
\]

**Proof.** First, rank \( (A_k) = k \) is obvious by construction. Furthermore, by linearity,

\[
A - A_k = U (\Sigma - \Sigma_k) V^*,
\]

which has largest singular value \( \sigma_{k+1} \), and hence \( \|A - A_k\|_2 = \sigma_{k+1} \).

Let \( B \in \mathbb{C}^{m \times n} \) have rank at most \( k \). Then by the Fundamental Theorem of Linear Algebra,

\[
\dim(N(B)) = n - \dim(C(B)) \geq n - k.
\]

Since the columns of \( V \) are orthonormal, then \( \text{span}\{v_1, \ldots, v_{k+1}\} \) has dimension \( k + 1 \). Hence

\[
\dim(N(B)) + \dim(\text{span}\{v_1, \ldots, v_{k+1}\}) \geq n + 1,
\]

so since both subspaces are subsets of \( \mathbb{C}^n \), they must have a nontrivial intersection. Let \( \hat{x} \) be a unit vector in this nontrivial intersection, i.e.

(a) \( \hat{x} \in N(B) \)

(b) \( \hat{x} \in \text{span}\{v_1, \ldots, v_{k+1}\} \)

(c) \( \|\hat{x}\|_2 = 1 \) (and hence \( \hat{x} \neq 0 \))

From (b), we may write

\[
\hat{x} = c_1 v_1 + \cdots + c_{k+1} v_{k+1},
\]

and since the \( v_j \)'s are orthonormal,

\[
1 = \|x\|_2^2 = |c_1|^2 + \cdots + |c_{k+1}|^2.
\]

From (a), we have \( B\hat{x} = 0 \), and hence

\[
(A - B)\hat{x} = A\hat{x} = A \left( \sum_{j=1}^{k+1} c_j v_j \right) = \sum_{j=1}^{k+1} c_j Av_j = \sum_{j=1}^{k+1} \sigma_j c_j u_j.
\]

Now, since the \( u_j \)'s are orthonormal,

\[
\|/(A - B)\hat{x}\|_2^2 = \sum_{j=1}^{k+1} |\sigma_j c_j|^2 \geq \sigma_{k+1}^2 \sum_{j=1}^{k+1} |c_j|^2 = \sigma_{k+1}^2,
\]

\[
\|A - B\|_2^2 = \sum_{j=1}^{k+1} |\sigma_j c_j|^2 \geq \sigma_{k+1}^2 \sum_{j=1}^{k+1} |c_j|^2 = \sigma_{k+1}^2.
\]
or since everything is positive, \( \|(A - B)\hat{x}\|_2 \geq \sigma_{k+1} \). Hence since \( \| \cdot \|_2 \) is an induced matrix norm,

\[
\|A - B\|_2 = \max_{\|x\|_2 = 1} \|(A - B)x\|_2 \geq \|(A - B)\hat{x}\|_2 \geq \sigma_{k+1}.
\]

\[
\square
\]

**Corollary 4.2.10.** Let \( A \in \mathbb{C}^{m \times n} \) have full rank, i.e., \( \text{rank}(A) = r = \min\{m, n\} \). Let \( A \) have singular values \( \sigma_1 \geq \cdots \geq \sigma_r > 0 \). Let \( B \in \mathbb{C}^{m \times n} \) satisfy \( \|A - B\|_2 < \sigma_r \). Then \( B \) also has full rank.

(proof for homework/exams)

In other words, if \( A \) has full rank, then matrices that are “sufficiently close” to \( A \) also have full rank.

### 4.2.1 Closing remarks on SVD applications

The QR decomposition with column pivoting results in

\[
AP = QR \quad \Leftrightarrow \quad A = QRP^T,
\]

where \( Q \) is unitary and \( P \) is orthogonal (unitary and real-valued). Hence this is also a unitary decomposition of \( A \), where the “center” matrix has simple structure (here, upper-triangular).

- The SVD gives a handle on the numerical rank, while the QR does not (+1 for SVD)
- It is much more efficient to compute the pivoted QR factorization than the SVD (+1 for QR)

**Moral:** pick the factorization that “best suits” your needs!

Recall from Chapter 2 that for any nonsingular matrix \( A \in \mathbb{C}^{n \times n} \)

\[
\frac{\|A - A_s\|}{\|A\|} \geq \frac{1}{\kappa(A)},
\]

where \( A_s \) is the “closest” singular matrix to \( A \) – this held for any induced matrix norm. Well in the 2-norm, this is an equality.

**Corollary 4.2.11.** Let \( A \in \mathbb{C}^{n \times n} \) be nonsingular, and let \( A_s \) be the “closest” singular matrix to \( A \) in the 2-norm, i.e., \( \|A - A_s\|_2 \) is minimized. Then \( \|A - A_s\|_2 = \sigma_n \), and

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

(proof directly follows from Theorem 4.2.9)
4.3 SVD and the Least Squares Problem

Let $A \in \mathbb{C}^{m \times n}$ with $\text{rank}(A) = r$. Let $b \in \mathbb{R}^m$ and consider the linear system

$$Ax = b$$

If $m > n$ then this system is overdetermined, so we seek $\bar{x} \in \mathbb{C}^n$ such that $\|b - Ax\|_2$ is minimized. If $r < n$, then $\bar{x}$ is not unique, so we additionally will seek the $\bar{x}$ such that $\|\bar{x}\|_2$ is minimal.

Suppose that $A = U\Sigma V^*$, with $U$ unitary and $V$ unitary and $\Sigma$ diagonal, where

$$\Sigma = \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix}, \quad \hat{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r),$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then

$$\|b - Ax\|_2^2 = \|U^*(b - Ax)\|_2^2 = \|U^*b - \Sigma V^*x\|_2^2 = \|c - \Sigma y\|_2^2 = \sum_{k=1}^r |c_k - \sigma_k y_k|^2 + \sum_{k=r+1}^n |c_k|^2$$

i.e., the $y$ that minimizes the least-squares problem satisfies

$$y_k = \frac{c_k}{\sigma_k}, \quad k = 1, \ldots, r.$$

Moreover, $\|y\|_2$ is minimized by setting $y_{r+1} = y_{r+2} = \cdots = y_n = 0$. Since $x = Vy$, where $V$ is unitary, then $\|y\|_2 = \|x\|_2$, so $\|x\|_2$ is minimal whenever $\|y\|_2$ is also minimal.

As a result, we are guaranteed that the least-squares problem has at least one minimum norm solution.

Partitioning appropriately, we have

$$c = \begin{bmatrix} \hat{c} \\ d \end{bmatrix}, \quad y = \begin{bmatrix} \hat{y} \\ z \end{bmatrix}, \quad \text{with} \quad \hat{c}, \hat{y} \in \mathbb{C}^r,$$

so

$$c - \Sigma y = \begin{bmatrix} \hat{c} \\ d \end{bmatrix} - \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{y} \\ z \end{bmatrix} = \begin{bmatrix} \hat{c} - \hat{\Sigma} \hat{y} \\ d \end{bmatrix},$$

and $\|b - Ax\|_2^2 = \|c - \Sigma y\|_2^2 = \|\hat{c} - \hat{\Sigma} \hat{y}\|_2^2 + \|d\|_2^2$. While we can solve this via $\hat{y} = \hat{\Sigma}^{-1} \hat{c}$, or equivalently, $\hat{y}_k = \frac{\hat{c}_k}{\sigma_k}$, the minimum norm for $\|y\|_2$ occurs with $z = 0$. Hence the norm of the minimized residual is therefore $\|d\|_2$.

Steps for computing the minimum norm solution, $x$: 

D.R. Reynolds, SMU Mathematics
1. Compute \( c = U^*b \), and partition this into \( c = \begin{bmatrix} \hat{c} \\ d \end{bmatrix} \).

2. Compute \( \hat{y} = \hat{\Sigma}^{-1}\hat{c} \).

3. Construct \( y = \begin{bmatrix} \hat{y} \\ 0 \end{bmatrix} \).

4. Compute \( x = Vy \).

Unfortunately, the big “gotcha” here is that if we do not know \( \text{rank}(A) = r \), then how can we accomplish steps 2 and 3? Answer: use the numerical rank instead.

### 4.3.1 The rank-deficient pseudoinverse

Recall that given the singular value decomposition of a matrix, \( A = U\Sigma V^* \), in Definition 4.2.7 we previously defined the pseudoinverse of a \( A \in \mathbb{C}^{m \times n} \) with full column rank as

\[
A^\dagger = (A^*A)^{-1}A^*.
\]

If instead we only have a rank-\( r \) matrix \( A \), with \( r < \min\{m, n\} \), then we can extend this definition to similarly provide the solution to the least-squares problem.

Returning to our “geometric” version of the SVD, Theorem 4.1.3,

\[
Av_k = \sigma_k u_k, \quad k = 1, \ldots, r.
\]

We would therefore “like” that even when \( A \) is a singular matrix, we could convert the above equations to

\[
A^\dagger u_k = \sigma_k^{-1} v_k, \quad k = 1, \ldots, r,
\]

and for the remaining \( u_k \) and \( v_k \) entries \((k = r + 1, \ldots)\), we choose

\[
A^\dagger u_k = 0, \quad k = r + 1, \ldots, n.
\]

The pseudoinverse of \( A \in \mathbb{C}^{m \times n} \) is the unique matrix \( A^\dagger \in \mathbb{C}^{n \times n} \) such that both of the above goals are true, i.e.,

\[
A^\dagger u_k = \begin{cases} 
\sigma_k^{-1} v_k, & k = 1, \ldots, r \\
0, & k = r + 1, \ldots, n
\end{cases}
\]

Note: Since the vectors \( u_k \) and \( v_k \) are orthonormal, then \( \text{rank}(A^\dagger) = \text{rank}(A) = r \), where \( \{u_1, \ldots, u_n\} \) and \( \{v_1, \ldots, v_n\} \) are right/left singular vectors for \( A^\dagger \), respectively, with singular values \( \sigma_r^{-1} \geq \sigma_{r-1}^{-1} \geq \cdots \geq \sigma_1^{-1} > 0 \).

If we then have \( \Sigma = \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n} \) with \( \hat{\Sigma} = \text{diag}(\sigma_1, \ldots, \sigma_r) \), then we may define \( \Sigma^\dagger = \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{n \times m} \). Then since \( U \) and \( V \) are unitary, we’ll have

\[
A^\dagger = V \Sigma^\dagger U^*.
\]
While this formula looks unlike our previous Definition 4.2.7 for the case that $A$ has full column rank, in fact these formulas are identical. First, note that if $A$ has full column rank then $r = n$, and so $\Sigma = \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix}$. Hence

(a) $\Sigma^T \Sigma = \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix} \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix} = \hat{\Sigma}^2$, which is an invertible, diagonal matrix in $\mathbb{R}^{n \times n}$

(b) $(\Sigma^T \Sigma)^{-1} = \hat{\Sigma}^{-2}$, and

(c) $(\Sigma^T \Sigma)^{-1} \Sigma^T = \hat{\Sigma}^{-2} \begin{bmatrix} \hat{\Sigma} & 0 \end{bmatrix} = \begin{bmatrix} \hat{\Sigma}^{-1} & 0 \end{bmatrix} = \Sigma^\dagger \in \mathbb{R}^{n \times m}$.

We may use the above results to connect the two definitions of the pseudoinverse:

$$A_{\text{orig}}^\dagger = (A^* A)^{-1} A^* \quad \text{[def. 4.2.7]}$$

$$= ((U \Sigma V^*)^* U \Sigma V^*)^{-1} (U \Sigma V^*)^* \quad [A = U \Sigma V^*]$$

$$= (V \Sigma^T U^* U \Sigma V^*)^{-1} V \Sigma^T U^* \quad \text{[distributing the Hermitian-transpose]}$$

$$= V (\Sigma^T \Sigma)^{-1} V^* V \Sigma^T U^* \quad [V^* = V^{-1}, \text{and } \Sigma^T \Sigma \text{ invertible}]$$

$$= V \Sigma^\dagger U^* \quad [(c) \text{ above}]$$

$$= A_{\text{new}}^\dagger \quad \text{[equation (1)]}.$$  

**Theorem 4.3.1.** Let $A \in \mathbb{C}^{m \times n}$ and $b \in \mathbb{C}^m$. If $x \in \mathbb{C}^n$ is the minimum-norm solution of

$$\|b - Ax\|_2 = \min_{w \in \mathbb{C}^n} \|b - Aw\|_2,$$

then $x = A^\dagger b$.

**Proof.** By our earlier algorithm, we may compute $x$ via

$$x = V y = V \begin{bmatrix} \hat{y} \\ 0 \end{bmatrix} = V \begin{bmatrix} \hat{\Sigma}^{-1} \hat{c} \\ 0 \end{bmatrix} = V \begin{bmatrix} \hat{\Sigma}^{-1} \hat{c} \\ 0 \end{bmatrix} \begin{bmatrix} \hat{d} \\ 0 \end{bmatrix} = V \Sigma^\dagger c = V \Sigma^\dagger U^* b = A^\dagger b.$$

Matlab note: `pinv` returns the pseudoinverse of $A \in \mathbb{C}^{m \times n}$.