Section 4.1. Recall the SVD: for any matrix $A \in \mathbb{R}^{m,n}$ with rank $r > 0$, there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ as well as a “diagonal” matrix $\Sigma \in \mathbb{R}^{m \times n}$ with diagonal entries $\Sigma(i,i) = \sigma_i$, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, and all the other entries 0.

Throughout this chapter, we will assume for simplicity that $m \geq n$. For the complementary case, it is sufficient to note that the SVD of $A^T$ is $A^T = V \Sigma^T U^T$.

Recall “rank” is the

- largest number of linearly independent rows
- largest number of linearly independent columns
- dimension of range of $A$
- number of nonzero entries on the diagonal of $R$ (for the reduced QR factorization)
- number of nonzero eigenvalues
- number of nonzero singular values.

If one looks carefully at the SVD above, one may see that due to the pattern of zeros in $\Sigma$, only the first $r$ columns of $U$ are involved in the calculation of $A$ (the entries in all the other columns get multiplied by 0 and vanish), only the first $r$ columns of $V$ are involved in the calculation (for the same reason), and $\Sigma$ could be boiled down to its upper left $r \times r$ corner.

Let $U = [u_1, \ldots, u_m]$ and $V = [v_1, \ldots, v_n]$; let $U_r = [u_1, \ldots, u_r]$ and $V_r = [v_1, \ldots, v_r]$; similarly let $\sigma_r$ be the top left $r \times r$ corner of $\Sigma$, so a square diagonal matrix with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$ on the diagonal. The above paragraph gives us the reduced (or, as the textbook likes to call it, condensed) SVD, which represents a “chopping off” of the columns and rows that do not play any role in the final product, $A$.

**Theorem 1.** Let $A \in \mathbb{R}^{m \times n}$ be a nonzero matrix of rank $r$. Then, with the definitions above,

$$A = U_r \Sigma_r V_r^T.$$  \hspace{1cm} (1)

Neither the full SVD nor the reduced SVD are unique, but the singular values $\sigma_1, \ldots, \sigma_r$ are, since the way $A$ maps the unit sphere is unique (for any $1 \leq i \leq r$, we can replace any $u_i$ with $-u_i$ and effect the same change in the corresponding $v_i$, so this leads to non-uniqueness).

Finally, another way to interpret Theorem 1 above is that

$$A = \sum_{i=1}^r \sigma_i u_i v_i^T.$$  \hspace{1cm} (1)

This is a very useful decomposition, which is employed in a variety of fields from machine learning to image compression and from statistics and principal component analysis to matrix completion. We will see some other consequences of this below.
The spectral norm and the SVD. Recall

**Theorem 2.** For any matrix \( A \in \mathbb{R}^{m \times n} \), define \( \|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \). Then \( \|A\|_2 = \sigma_1 \), where \( \sigma_1 \) is the largest singular value of \( A \).

For an \( n \times n \) square, invertible matrix (\( r = n \)), is \( A = U\Sigma V^T \) (then \( U, V, \Sigma \) are all \( n \times n \)), \( A^{-1} = V\Sigma^{-1}U^T \), and note that this means the singular values of \( A^{-1} \) are \( \frac{1}{\sigma_n} \geq \frac{1}{\sigma_{n-1}} \geq \ldots \geq \frac{1}{\sigma_1} \), and thus

\[
\|A^{-1}\|_2 = \frac{1}{\sigma_n}.
\]

Thus,

**Theorem 3.** For an \( n \times n \) invertible matrix \( A \), in the 2-norm,

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

Low-rank approximation. Let us write \( A \) as the sum of rank-one matrices from (1), and we will denote by \( A_k \) the truncated sum \( A_k = \sum_{i=1}^k \sigma_i u_i v_i^T \).

**Theorem 4.** For any \( k = 1, \ldots, r \), let \( \mathcal{V}_k \) be the set of all \( m \times n \) matrices of rank at most \( k \). Then, with the notation above,

\[
\sigma_{k+1} = \|A - A_k\|_2 = \min_{B \in \mathcal{V}_k} \|A - B\|_2.
\]

That is, among all matrices of rank \( k \) or less, \( A_k \) is “closest” to \( A \), and the difference in norm is \( \sigma_{k+1} \).

**Remark 1.** Note that \( \{0\} \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \ldots \subset \mathcal{V}_{n-1} \subset \mathcal{V}_n = \mathbb{R}^{n \times n} \), and that \( \mathcal{V}_{n-1} \) is the set of all singular matrices (if a matrix is singular, then its rank is at most \( n - 1 \)).

The proof for this theorem is in the textbook (4.2.15), and I invite you all to take a look. Although it’s a bit long, it is quite elegant.

As a consequence, we have the following two results.

**Corollary 1.** If \( A \) is square and full-rank (\( m = r = n \)) and \( B \) is a matrix for which \( \|B - A\|_2 < \sigma_n \), then \( B \) is full-rank.

**Corollary 2.** If \( A \) is square and non-singular (\( m = n = r \) and \( \sigma_n > 0 \)), let \( B \) be the matrix that is singular and closest to \( A \) in 2-norm (so that \( \|A - B\|_2 \) is minimal among all singular matrices). Then, with the notations above, \( B = A_{n-1} \) and

\[
\frac{\|A - A_{n-1}\|_2}{\|A\|_2} = \frac{1}{\kappa(A)}.
\]

Note that the second corollary shows that the result we got in 2.3 about the distance to singularity and the condition number is, in fact, tight.