Section 5.8. Although we could talk of the SVD of a complex matrix, we will limit ourselves to $A$ being real for the purposes of this chapter.

We have so far seen that one way to compute the SVD of an $n \times m$ real matrix $A$ is to form $A^T A$ and compute its eigenvalues and eigenvectors; the square roots of the eigenvalues will be singular values, and the $m$ eigenvectors will be the right singular vectors. We can get $m$ left-singular vectors from them, and the rest by finding the eigenvectors corresponding to eigenvalue 0 for $AA^T$. (Or, we may just find the eigenvectors of $AA^T$.)

The problem with this approach is that the condition number of the matrix $A^T A$ is the square of the condition number of $A$:

$$\kappa_2(A^T A) = \left(\frac{\sigma_1}{\sigma_m}\right)^2 \left(\frac{\sigma_1}{\sigma_m}\right)^2 = (\kappa_2(A))^2.$$  

Although we have only seen the way in which condition numbers affect the solution under perturbations for systems $Ax = b$, it should be said that they affect similarly eigenvalue computations, and as such the squaring of a condition number (which may already be large) has a negative impact on accuracy.

So this method is not necessarily desirable, and in fact there aren’t many SVD algorithms based on it.

There are other, better methods that are used in practice, and they involve finding the bidiagonal form of $A$. That is, finding two orthogonal matrices $\tilde{U}$ ($n \times n$) and $\tilde{V}$ ($m \times m$), as well as a bidiagonal $n \times m$ matrix $B$, such that $A = \tilde{U}B\tilde{V}^T$. Here $B$ will have the following explicit shape:

$$B = \begin{bmatrix} * & * & \cdots & * \ \ * & * & \cdots & * \ \ \vdots & \vdots & \ddots & \vdots \ \ * & * & \cdots & * \end{bmatrix}.$$  

Lemma 1. Suppose that $n \times m$ matrices $A$ and $B$ are related in the following way: there exist orthogonal matrices $U$ and $V$, $n \times n$, respectively $m \times m$, such that $A = UBV^T$. Then $A$ and $B$ have the same singular values.

Proof. Let $B = \tilde{U}\Sigma\tilde{V}^T$; then

$$A = \left(U\tilde{U}\right) \Sigma \left(\tilde{V}^T V^T\right) = \left(U\tilde{U}\right) \Sigma \left(V\tilde{V}\right)^T.$$  

The last expression is an SVD for $A$, since products of orthogonal matrices are orthogonal. So $A$ and $B$ have the same singular values.
Just as we showed the existence of the Hessenberg form of a matrix (which shares the same eigenvalues), we can use Householder reflectors to prove the existence of a bidiagonal form of a matrix, which shares the same singular values.

One very important thing to note is that before, the process could be done by first “zeroing out” the lower triangular part of the matrix, then multiplying the resulting matrices on the right side. Now, however, the process will have a very clear prescribed order: we will have to first zero out the first column, then work on zeroing our all but two entries in the first row; then we zero out the under-diagonal of the second column, followed by the zeroing out of all but 3 entries of the second row; basically, we keep moving from left to right and back again, or else we cannot guarantee that the next multiplication by a reflector matrix will not mess up the zeros we created at the previous step.

Moreover, note that since \( B = \begin{bmatrix} \tilde{B} \\ 0 \end{bmatrix} \), we might as well work only with the upper, square \( m \times m \) matrix \( \tilde{B} \). From now on, \( \tilde{B} \) will be square.

**Next: computing the SVD of a bidiagonal matrix.** One obvious thing to do here would be to examine \( \tilde{B}^T \tilde{B} \); indeed, a quick calculation shows that if \( A = \hat{U} \hat{B} \hat{V}^T \), then
\[
A^T A = \hat{V} \tilde{B}^T \tilde{B} \hat{V}^T,
\]
and so, \( A^T A \) and \( \tilde{B}^T \tilde{B} \) have the same eigenvalues. In fact, the following result is very easy and can be left as an exercise.

**Lemma 2.** \( \tilde{B}^T \tilde{B} \) is tridiagonal and symmetric.

**Example.** Let \( \tilde{B} = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 \end{bmatrix} \); then
\[
\tilde{B}^T \tilde{B} = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 5 & -1 & 0 \\ 0 & -1 & 2 & 2 \\ 0 & 0 & 2 & 5 \end{bmatrix}.
\]

**Remark 1.** It should come as no surprise that \( \tilde{B}^T \tilde{B} \) is the Hessenberg form of \( A^T A \).

This would more or less equivalent to first computing \( A^T A \), then the Hessenberg form of \( A \), and so it shares its drawback. We have already mentioned that \( \kappa_2(A^T A) = (\kappa(A))^2 \) (indeed, you had to prove it as part of a homework). This means that calculating anything based on \( A^T A \) potentially amplifies errors MUCH more than algorithms based on \( A \). Note that \( \kappa_2(\tilde{B}) = \kappa_2(A) \), and \( \kappa_2(A^T A) = \kappa_2(\tilde{B}^T \tilde{B}) \).

What can we do to avoid this pitfall? The answer seems simple, don’t multiply with the transpose. Surprisingly, this turns out to be realizable. First, we need a very important lemma.

**Lemma 3.** Assume that \( A \) (and thus, \( \tilde{B} \)) is full-rank. The matrix \( C = \begin{bmatrix} 0 & \tilde{B}^T \\ \tilde{B} & 0 \end{bmatrix} \) has eigenvalues that come in pairs: For any \( 1 \leq i \leq m \), \((\sigma_i, [v_i^T, u_i^T]^T)\) is an eigenpair; so is \((-\sigma, [-v_i^T, -u_i^T]^T)\). As the matrix \( C \) has size \( 2m \times 2m \), these are all the eigenvalues.
Proof. This is actually pretty easy to check, as

\[
C \begin{bmatrix} v_i \\ u_i \end{bmatrix} = \begin{bmatrix} A^T u_i \\ Av_i \end{bmatrix} = \begin{bmatrix} \sigma_i v_i \\ \sigma_i u_i \end{bmatrix} = \sigma_i \begin{bmatrix} v_i \\ u_i \end{bmatrix},
\]

as well as

\[
C \begin{bmatrix} v_i \\ -u_i \end{bmatrix} = \begin{bmatrix} -A^T u_i \\ Av_i \end{bmatrix} = \begin{bmatrix} -\sigma_i v_i \\ \sigma_i u_i \end{bmatrix} = -\sigma_i \begin{bmatrix} v_i \\ -u_i \end{bmatrix}.
\]

While working directly with \( C \) is undesirable, as every eigenvalue has a counterpart that is equal in absolute value, the use of shifts and some smart permutations makes it possible to compute the eigenvalues of the matrix \( C \) in \( O(n^2) \), and so the hard part, just as it was the case with the Hessenberg reduction, is in the bidiagonalization of \( A \).

**PCA and the SVD.** We can revisit the PCA now, with the insights we have gained into the relationship between eigenvalues and singular values, and see that the eigenvalues of the sample covariance matrix \( C \) are the singular values of \( X \), and that the eigenvectors of \( C \) are the right singular vectors of \( X \).

We can think of PCA as either an eigenvalue problem (involving \( C = X^T X \), the sample covariance matrix) or as a singular value problem (involving \( X \), the data matrix).