2 Singular Value Decomposition

The *singular value decomposition* (SVD) allows us to transform a matrix $A \in \mathbb{C}^{m\times n}$ to diagonal form using unitary matrices, i.e.,

$$A = \hat{U}\hat{\Sigma}V^*. \tag{4}$$

Here $\hat{U} \in \mathbb{C}^{m\times n}$ has orthonormal columns, $\hat{\Sigma} \in \mathbb{C}^{n\times n}$ is diagonal, and $V \in \mathbb{C}^{n\times n}$ is unitary. This is the practical version of the SVD also known as the *reduced SVD*. We will discuss the *full SVD* later. It is of the form

$$A = U\Sigma V^*$$

with unitary matrices U and V and $\Sigma \in \mathbb{C}^{m\times n}$.

Before we worry about how to find the matrix factors of A we give a geometric interpretation. First note that since V is unitary (i.e., $V^* = V^{-1}$) we have the equivalence

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

Considering each column of V separately the latter is the same as

$$A\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j, \qquad j = 1, \ldots, n. \tag{5}$$

Thus, the unit vectors of an orthogonal coordinate system $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ are mapped under A onto a new "scaled" orthogonal coordinate system $\{\sigma_1\boldsymbol{u}_1, \ldots, \sigma_n\boldsymbol{u}_n\}$. In other words, the unit sphere with respect to the matrix 2-norm (which is a perfectly round sphere in the $\boldsymbol{v}$-system) is transformed to an ellipsoid with semi-axes $\sigma_j\boldsymbol{u}_j$ (see Figure 2). We will see below that, depending on the rank of A, some of the σ_j may be zero. Therefore, yet another geometrical interpretation of the SVD is: Any $m \times n$ matrix A maps the 2-norm unit sphere in $\mathbb{R}^n$ to an ellipsoid in $\mathbb{R}^r$ $(r \leq \min(m, n))$.

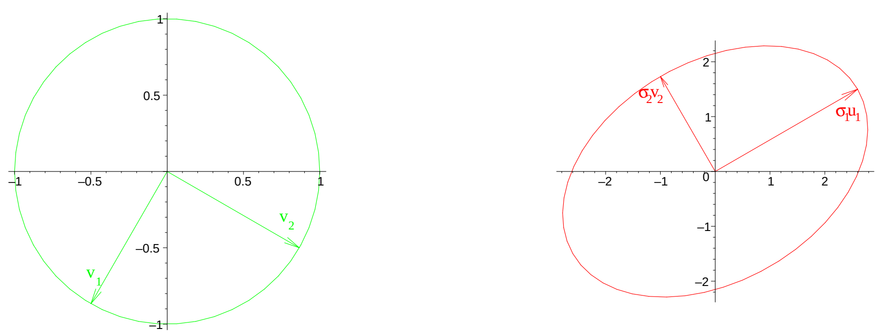

Figure 2: Geometrical interpretation of singular value decomposition.

In (5) we refer to the σ_j as *singular values* of A (the diagonal entries of $\hat{\Sigma}$). They are usually ordered such that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$. The orthonormal vectors $\boldsymbol{u}_j$ (the columns of $\hat{U}$) are called the *left singular vectors* of A, and the orthonormal vectors $\boldsymbol{v}_j$ (the columns of V) are called the *right singular vectors* of A).

Remark For most practical purposes it suffices to compute the reduced SVD (4). We will give examples of its use, and explain how to compute it later.

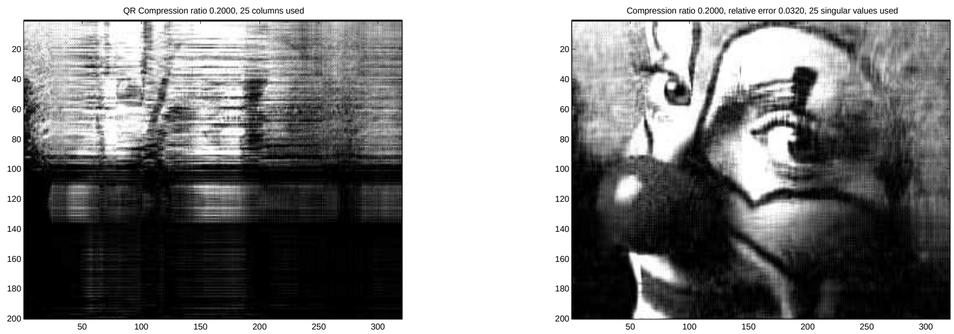

Figure 3: Image compressed using QR factorization (left) and SVD (right).

Besides applications to inconsistent and underdetermined linear systems and least squares problems, the SVD has important applications in image and data compression (see our discussion of low-rank approximation below). Figure 3 shows the difference between using the SVD and the QR factorization (to be introduced later) for compression of the same image. In both cases the same amount (20%) of information was retained. Clearly, the SVD does a much better job in picking out what information is "important". We will also see below that a number of theoretical facts about the matrix A can be obtained via the SVD.

2.0.4 Full SVD

The idea is to extend $\hat{U}$ to an orthonormal basis of $\mathbb{C}^{m\times m}$ by adding appropriate orthogonal (but otherwise arbitrary) columns and call this new matrix U. This will also force $\hat{\Sigma}$ to be extended to an $m \times n$ matrix Σ. Since we do not want to alter the product of the factors, the additional rows (or columns – depending on whether $m > n$ or $m < n$) of Σ will be all zero. Thus, in the case of $m \geq n$ we have

$$\begin{aligned} A &= U\Sigma V^* \qquad (6) \\ &= \begin{bmatrix} \hat{U} & \widetilde{U} \end{bmatrix} \begin{bmatrix} \hat{\Sigma} \\ O \end{bmatrix} V^*. \end{aligned}$$

Since U is now also a unitary matrix we have

$$U^*AV = \Sigma,$$

i.e., unitary transformations (reflections or rotations) are applied from the left and right to A in order to obtain a diagonal matrix Σ.

Remark Note that the "diagonal" matrix Σ is in many cases rectangular and will contain extra rows/columns of all zeros.

It is clear that the SVD will simplify the solution of many problems since the transformed system matrix is diagonal, and thus trivial to work with.

2.0.5 Existence and Uniqueness Theorem

Theorem 2.1 *Let A be a complex $m \times n$ matrix. A has a singular value decomposition of the form*

$$A = U\Sigma V^*,$$

where Σ is a uniquely determined $m \times n$ (real) diagonal matrix, U is an $m \times m$ unitary matrix, and V is an $n \times n$ unitary matrix.

Proof We prove only existence. The uniqueness part of the proof follows directly from the geometric interpretation. A (more rigorous?) algebraic argument can be found, e.g., in [Trefethen/Bau].

We use induction on the dimensions of A. All of the following arguments assume $m \geq n$ (the case $m < n$ can be obtained by transposing the arguments).

For $n = 1$ (and any m) the matrix A is a column vector. We take $V = 1$, $\hat{\Sigma} = \|A\|_2$ and $\hat{U} = \frac{A}{\|A\|_2}$. Then, clearly, we have found a reduced SVD, i.e., $A = \hat{U}\hat{\Sigma}V^*$. The full SVD is obtained by extending $\hat{U}$ to U by the Gram-Schmidt algorithm and adding the necessary zeros to $\hat{\Sigma}$.

We now assume an SVD exists for the case $(m-1, n-1)$ and show it also exists for (m, n). To this end we pick $\boldsymbol{v}_1 \in \mathbb{C}^n$ such that $\|\boldsymbol{v}_1\|_2 = 1$ and

$$\|A\|_2 = \sup_{\substack{\boldsymbol{v}_1 \in \mathbb{C}^n \\ \|\boldsymbol{v}_1\|_2 = 1}} \|A\boldsymbol{v}_1\|_2 > 0.$$

Now we take

$$\boldsymbol{u}_1 = \frac{A\boldsymbol{v}_1}{\|A\boldsymbol{v}_1\|_2}. \tag{7}$$

Next, we use the Gram-Schmidt algorithm to arbitrarily extend $\boldsymbol{u}_1$ and $\boldsymbol{v}_1$ to unitary matrices by adding columns $\widetilde{U}_1$ and $\widetilde{V}_1$, i.e.,

$$U_1 = \begin{bmatrix} \boldsymbol{u}_1 & \widetilde{U}_1 \end{bmatrix} \quad V_1 = \begin{bmatrix} \boldsymbol{v}_1 & \widetilde{V}_1 \end{bmatrix}.$$

This results in

$$\begin{aligned} U_1^* A V_1 &= \begin{bmatrix} \boldsymbol{u}_1^* \\ \widetilde{U}_1^* \end{bmatrix} A \begin{bmatrix} \boldsymbol{v}_1 & \widetilde{V}_1 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{u}_1^* A \boldsymbol{v}_1 & \boldsymbol{u}_1^* A \widetilde{V}_1 \\ \widetilde{U}_1^* A \boldsymbol{v}_1 & \widetilde{U}_1^* A \widetilde{V}_1 \end{bmatrix}. \end{aligned}$$

We now look at three of these four blocks:

- Using (7) and the specific choice of $\boldsymbol{v}_1$ we have

$$\begin{aligned} \boldsymbol{u}_1^* A \boldsymbol{v}_1 &= \frac{(A\boldsymbol{v}_1)^*}{\|A\boldsymbol{v}_1\|_2} A\boldsymbol{v}_1 \\ &= \frac{\|A\boldsymbol{v}_1\|_2^2}{\|A\boldsymbol{v}_1\|_2} = \|A\boldsymbol{v}_1\|_2 \\ &= \|A\|_2. \end{aligned}$$

  For this quantity we introduce the abbreviation $\sigma_1 = \|A\|_2$.

- Again, using (7) we get
$$\widetilde{U}_1^* A\boldsymbol{v}_1 = \widetilde{U}_1^* \boldsymbol{u}_1 \|A\boldsymbol{v}_1\|_2.$$
This, however, is a zero vector since U_1 has orthonormal columns, i.e., $\widetilde{U}_1^* \boldsymbol{u}_1 = \boldsymbol{0}$.

- We show that $\boldsymbol{u}_1^* A\widetilde{V}_1 = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}$ by contradiction. If it were nonzero then we could look at the first row of the block matrix $U_1^* A V_1$ and see that
$$U_1^* A V_1(1,:) = \begin{bmatrix} \sigma_1 & \boldsymbol{u}_1^* A\widetilde{V}_1 \end{bmatrix}$$
with $\| \begin{bmatrix} \sigma_1 & \boldsymbol{u}_1^* A\widetilde{V}_1 \end{bmatrix} \|_2 > \sigma_1$. On the other hand, we know that unitary matrices leave the 2-norm invariant, i.e.,
$$\|U_1^* A V_1\|_2 = \|A\|_2 = \sigma_1.$$
Since the norm of the first row of the block matrix $U_1^* A V_1$ cannot exceed that of the entire matrix we have reached a contradiction.

We now abbreviate the fourth block with $\widetilde{A} = \widetilde{U}_1^* A \widetilde{V}_1$ and can write the block matrix as
$$U_1^* A V_1 = \begin{bmatrix} \sigma_1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \widetilde{A} \end{bmatrix}.$$
To complete the proof we apply the induction hypothesis to $\widetilde{A}$, i.e., we use the SVD $\widetilde{A} = U_2 \Sigma_2 V_2^*$. Then
$$\begin{aligned} U_1^* A V_1 &= \begin{bmatrix} \sigma_1 & \boldsymbol{0}^T \\ \boldsymbol{0} & U_2 \Sigma_2 V_2^* \end{bmatrix} \\ &= \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & V_2 \end{bmatrix}^* \end{aligned}$$
or
$$A = U_1 \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & V_2 \end{bmatrix}^* V_1^*,$$
another SVD (since the product of unitary matrices is unitary). ■

2.1 SVD as a Change of Basis

We now discuss the use of the SVD to diagonalize systems of linear equations. Consider the linear system
$$A\boldsymbol{x} = \boldsymbol{b}$$
with $A \in \mathbb{C}^{m \times n}$. Using the SVD we can write
$$A\boldsymbol{x} = U\Sigma V^* \boldsymbol{x} \quad \Longleftrightarrow \quad \boldsymbol{b} = U\boldsymbol{b}'.$$
Thus, we can express $\boldsymbol{b} \in \text{range}(A)$ in terms of range(U):
$$U\boldsymbol{b}' = \boldsymbol{b} \quad \Longleftrightarrow \quad \boldsymbol{b}' = U^* \boldsymbol{b}$$

where we have used the columns of U as an orthonormal basis for range(A).

Similarly, any $\boldsymbol{x} \in \mathbb{C}^n$ (the domain of A) can be written in terms of range(V):

$$\boldsymbol{x}' = V^*\boldsymbol{x}.$$

Now

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} &\iff U^*A\boldsymbol{x} = U^*\boldsymbol{b} \\ &\iff U^*U\Sigma V^*\boldsymbol{x} = U^*\boldsymbol{b} \\ &\iff I\Sigma V^*\boldsymbol{x} = U^*\boldsymbol{b} \\ &\iff \Sigma\boldsymbol{x}' = \boldsymbol{b}', \end{aligned}$$

and we have diagonalized the linear system.

In summary, expressing the range space of A in terms of the columns of U and the domain space of A in terms of the columns of V converts $A\boldsymbol{x} = \boldsymbol{b}$ to a diagonal system.

2.1.1 Connection to Eigenvalues

If $A \in \mathbb{C}^{m\times m}$ is square with a linearly independent set of eigenvectors (i.e., nondefective), then

$$AX = \Lambda X \quad \iff \quad X^{-1}AX = \Lambda,$$

where X contains the eigenvectors of A as its columns and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$ is a diagonal matrix of the eigenvalues of A.

If we compare this eigen-decomposition of A to the SVD we see that the SVD is a generalization: A need not be square, and the SVD always exists (whereas even a square matrix need not have an eigen-decomposition). The price we pay is that we require two unitary matrices U and V instead of only X (which is in general not unitary).

2.1.2 Theoretical Information via SVD

A number of theoretical facts about the matrix A can be obtained via the SVD. They are summarized in

Theorem 2.2 *Assume $A \in \mathbb{C}^{m\times n}$, $p = \min(m, n)$, and $r \leq p$ denotes the number of positive singular values of A. Then*

1. $\text{rank}(A) = r$
2. $\text{range}(A) = \text{range}(U(:, 1 : r))$
   $\text{null}(A) = \text{range}(V(:, r + 1, : n))$
3. $\|A\|_2 = \sigma_1$
   $\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2}$
4. *The eigenvalues of A^*A are the σ_i^2 and the $\boldsymbol{v}_i$ are the corresponding (orthonormalized) eigenvectors. The eigenvalues of AA^* are the σ_i^2 and possibly $m - n$ zeros. The corresponding orthonormalized eigenvectors are given by the $\boldsymbol{u}_i$.*

5. *If $A = A^*$ (Hermitian or real symmetric), then the eigen-decomposition $A = X\Lambda X^*$ and the SVD $A = U\Sigma V^*$ are almost identical. We have $U = X$, $\sigma_i = |\lambda_i|$, and $\boldsymbol{v}_i = \mathrm{sign}(\lambda_i)\boldsymbol{u}_i$.*

6. *If $A \in \mathbb{C}^{m\times m}$ then $|\det(A)| = \prod_{i=1}^{m} \sigma_i$.*

Proof We discuss items 1–3 and 6.

1. Since U and V are unitary matrices of full rank and $\mathrm{rank}(\Sigma) = r$ the statement follows from the SVD $A = U\Sigma V^*$.

2. Both statements follow from the fact that the range of Σ is spanned by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r$ and that U and V are full-rank unitary matrices whose ranges are $\mathbb{C}^m$ and $\mathbb{C}^n$, respectively.

3. The invariance of the 2-norm and Frobenius norm under unitary transformations imply $\|A\|_2 = \|\Sigma\|_2$ and $\|A\|_F = \|\Sigma\|_F$. Since Σ is diagonal we clearly have $\|\Sigma\|_2 = \max_{\substack{\boldsymbol{x}\in\mathbb{C}^n \\ \|\boldsymbol{x}\|_2=1}} \|\Sigma\boldsymbol{x}\|_2 = \max_{1\le i\le r} \sigma_i = \sigma_1$. The formula for $\|\Sigma\|_F$ follows directly from the definition of the Frobenius norm.

6. We know that the determinant of a unitary matrix is either plus or minus 1, and that of a diagonal matrix is the product of the diagonal entries. Finally, the determinant of a product of matrices is given by the product of their determinant. Thus, the SVD yields the stated result.

■

2.1.3 Low-rank Approximation

Theorem 2.3 *The $m \times n$ matrix A can be decomposed into a sum of r rank-one matrices:*

$$A = \sum_{j=1}^{r} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^*. \tag{8}$$

Moreover, the best 2-norm approximation of rank ν ($0 \le \nu \le r$) to A is given by

$$A_\nu = \sum_{j=1}^{\nu} \sigma_j \boldsymbol{u}_j \boldsymbol{v}_j^*.$$

In fact,

$$\|A - A_\nu\|_2 = \sigma_{\nu+1}. \tag{9}$$

Proof The representation (8) of the SVD follows immediately from the full SVD (6) by splitting Σ into a sum of diagonal matrices $\Sigma_j = \mathrm{diag}(0, \ldots, 0, \sigma_j, 0, \ldots, 0)$.

Formula (9) for the approximation error follows from the fact that $U^*AV = \Sigma$ and the expansion for A_ν so that $U^*(A - A_\nu)V = \mathrm{diag}(0, \ldots, 0, \sigma_{\nu+1}, \ldots)$ and $\|A - A_\nu\|_2 = \sigma_{\nu+1}$ by the invariance of the 2-norm under unitary transformations and item 3 of the previous theorem.

The claim regarding the best approximation property is a little more involved, and omitted. ■

Remark There are many possible rank-ν decompositions of A (e.g., by taking partial sums of the LU or QR factorization). Theorem 2.3, however, says that the ν-th partial sum of the SVD captures as much of the *energy* of A (measured in the 2-norm) as possible. This fact gives rise to many applications in image processing, data compression, data mining, and other fields. See, e.g., the Matlab scripts `svd_compression.m` and `qr_compression.m`.

A geometric interpretation of Theorem 2.3 is given by the best approximation of a hyperellipsoid by lower-dimensional ellipsoids. For example, the best approximation of a given hyperellipsoid by a line segment is given by the line segment corresponding to the hyperellipsoids longest axis. Similarly, the best approximation by an ellipse is given by that ellipse whose axes are the longest and second-longest axis of the hyperellipsoid.

2.1.4 Computing the SVD by Hand

We now list a simplistic algorithm for computing the SVD of a matrix A. It can be used fairly easily for manual computation of small examples. For a given $m \times n$ matrix A the procedure is as follows:

1. Form A^*A.

2. Find the eigenvalues and orthonormalized eigenvectors of A^*A, i.e.,

$$A^*A = V\Lambda V^*.$$

3. Sort the eigenvalues according to their magnitude, and let $\sigma_j = \sqrt{\lambda_j}$, $j = 1, \ldots, n$.

4. Find the first r columns of U via

$$\boldsymbol{u}_j = \sigma_j^{-1} A \boldsymbol{v}_j, \qquad j = 1, \ldots, r.$$

Pick the remaining columns such that U is unitary.

Example Find the SVD for

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{bmatrix}.$$

1.

$$A^*A = \begin{bmatrix} 9 & 8 \\ 8 & 9 \end{bmatrix}.$$

2. The eigenvalues (in order of decreasing magnitude) are $\lambda_1 = 17$ and $\lambda_2 = 1$, and the corresponding eigenvectors

$$\widetilde{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \widetilde{\boldsymbol{v}}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

so that (after normalization)

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

3. $\sigma_1 = \sqrt{17}$ and $\sigma_2 = 1$, so that

$$\Sigma = \begin{bmatrix} \sqrt{17} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

4. The first two columns of U can be computed as

$$\begin{aligned} \boldsymbol{u}_1 &= \frac{1}{\sqrt{17}} A \boldsymbol{v}_1 \\ &= \frac{1}{\sqrt{17}} \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{34}} \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}, \end{aligned}$$

and

$$\begin{aligned} \boldsymbol{u}_2 &= \frac{1}{1} A \boldsymbol{v}_2 \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 2 \\ 2 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}. \end{aligned}$$

Thus far we have

$$U = \begin{bmatrix} \frac{3}{\sqrt{34}} & \frac{-1}{\sqrt{2}} & \boldsymbol{u}_3(1) \\ \frac{4}{\sqrt{34}} & 0 & \boldsymbol{u}_3(2) \\ \frac{3}{\sqrt{34}} & \frac{1}{\sqrt{2}} & \boldsymbol{u}_3(3) \end{bmatrix}.$$

In order to determine $\boldsymbol{u}_3(i)$, $i = 1, 2, 3$, we need to satisfy

$$\boldsymbol{u}_j^* \boldsymbol{u}_3 = \delta_{j3}, \qquad j = 1, 2, 3.$$

The following choice satisfies this requirement

$$\boldsymbol{u}_3 = \frac{1}{\sqrt{17}} \begin{bmatrix} 2 \\ -3 \\ 2 \end{bmatrix},$$

so that

$$A = U \Sigma V^* = \begin{bmatrix} \frac{3}{\sqrt{34}} & \frac{-1}{\sqrt{2}} & \frac{2}{\sqrt{17}} \\ \frac{4}{\sqrt{34}} & 0 & \frac{-3}{\sqrt{17}} \\ \frac{3}{\sqrt{34}} & \frac{1}{\sqrt{2}} & \frac{2}{\sqrt{17}} \end{bmatrix} \begin{bmatrix} \sqrt{17} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

The reduced SVD is given by

$$A = \hat{U}\hat{\Sigma}V^* = \begin{bmatrix} \frac{3}{\sqrt{34}} & \frac{-1}{\sqrt{2}} \\ \frac{4}{\sqrt{34}} & 0 \\ \frac{3}{\sqrt{34}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{17} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Remark A practical computer implementation of the SVD will require an algorithm for finding eigenvalues. We will study this later. The two most popular SVD implementations use either a method called Golub-Kahan-Bidiagonalization (GKB) (from 1965), or some Divide-and-Conquer strategy (which have been studied since around 1980).