Chapter 19
The Singular Value Decomposition

The matrix decomposition introduced in this chapter is very important in many practical applications, since it yields the best possible approximation (in a certain sense) of a given matrix by a matrix of low rank. A low rank approximation can be considered a "compression" of the data represented by the given matrix. We illustrate this below with an example from image processing.

We first prove the existence of the decomposition.

Theorem 19.1 *Let $A \in \mathbb{C}^{n,m}$ with $n \geq m$ be given. Then there exist unitary matrices $V \in \mathbb{C}^{n,n}$ and $W \in \mathbb{C}^{m,m}$ such that*

$$A = V \Sigma W^H \quad \textit{with} \quad \Sigma = \begin{bmatrix} \Sigma_r & 0_{r,m-r} \\ 0_{n-r,r} & 0_{n-r,m-r} \end{bmatrix} \in \mathbb{R}^{n,m}, \quad \Sigma_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r), \tag{19.1}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $r = \mathrm{rank}(A)$.

Proof If $A = 0$, then we set $V = I_n$, $\Sigma = 0 \in \mathbb{C}^{n,m}$, $\Sigma_r = [\,]$, $W = I_m$, and we are finished.

Let $A \neq 0$ and $r := \mathrm{rank}(A)$. Since $n \geq m$, we have $1 \leq r \leq m$, and since $A^H A \in \mathbb{C}^{m,m}$ is Hermitian, there exists a unitary matrix $W = [w_1, \ldots, w_m] \in \mathbb{C}^{m,m}$ with

$$W^H(A^H A)W = \mathrm{diag}(\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^{m,m}$$

(cp. (2) in Corollary 18.18). Without loss of generality we assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. For every $j = 1, \ldots, m$ then $A^H A w_j = \lambda_j w_j$, and hence

$$\lambda_j w_j^H w_j = w_j^H A^H A w_j = \|A w_j\|_2^2 \geq 0,$$

i.e., $\lambda_j \geq 0$ for $j = 1, \ldots, m$. Then $\mathrm{rank}(A^H A) = \mathrm{rank}(A) = r$ (to see this, modify the proof of Lemma 10.25 for the complex case). Therefore, the matrix $A^H A$ has exactly r positive eigenvalues $\lambda_1, \ldots, \lambda_r$ and $m - r$ times the eigenvalue 0. We then

J. Liesen and V. Mehrmann, *Linear Algebra*, Springer Undergraduate Mathematics Series, DOI 10.1007/978-3-319-24346-7_19

define $\sigma_j := \lambda_j^{1/2}$, $j = 1, \dots, r$, and have $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Let Σ_r be as in (19.1),

$$D := \begin{bmatrix} \Sigma_r & 0 \\ 0 & I_{m-r} \end{bmatrix} \in GL_m(\mathbb{R}), \quad X = [x_1, \dots, x_m] := AWD^{-1},$$

$V_r := [x_1, \dots, x_r]$, and $Z := [x_{r+1}, \dots, x_m]$. Then

$$\begin{bmatrix} V_r^H V_r & V_r^H Z \\ Z^H V_r & Z^H Z \end{bmatrix} = \begin{bmatrix} V_r^H \\ Z^H \end{bmatrix} [V_r,\ Z] = X^H X = D^{-1} W^H A^H A W D^{-1} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

which implies, in particular, that $Z = 0$ and $V_r^H V_r = I_r$. We extend the vectors $x_1, \dots, x_r$ to an orthonormal basis $\{x_1, \dots, x_r, \widetilde{x}_{r+1}, \dots, \widetilde{x}_n\}$ of $\mathbb{C}^{n,1}$ with respect to the standard scalar product. Then the matrix

$$V := [V_r, \widetilde{x}_{r+1}, \dots, \widetilde{x}_n] \in \mathbb{C}^{n,n}$$

is unitary. From $X = AWD^{-1}$ and $X = [V_r,\ Z] = [V_r,\ 0]$ we finally obtain $A = [V_r,\ 0]DW^H$ and $A = V\Sigma W^H$ with Σ as in (19.1). □

As the proof shows, Theorem 19.1 can be formulated analogously for real matrices $A \in \mathbb{R}^{n,m}$ with $n \geq m$. In this case the two matrices V and W are orthogonal. If $n < m$ we can apply the theorem to A^H (resp. A^T in the real case).

Definition 19.2 A decomposition of the form (19.1) is called a *singular value decomposition* or short *SVD*[1] of the matrix A. The diagonal entries of the matrix Σ_r are called *singular values* and the columns of V resp. W are called *left* resp. *right singular vectors* of A.

From (19.1) we obtain the unitary diagonalizations of the matrices $A^H A$ and AA^H,

$$A^H A = W \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} W^H \quad \text{and} \quad AA^H = V \begin{bmatrix} \Sigma_r^2 & 0 \\ 0 & 0 \end{bmatrix} V^H.$$

The singular values of A are therefore uniquely determined as the positive square roots of the positive eigenvalues of $A^H A$ or AA^H. The unitary matrices V and W in the singular value decomposition, however, are (as the eigenvectors in general) not uniquely determined.

[1] In the development of this decomposition from special cases in the middle of the 19th century to its current general form many important players of the history of Linear Algebra played a role. In the historical notes concerning the singular value decomposition in [HorJ91] one finds contributions of Jordan (1873), Sylvester (1889/1890) and Schmidt (1907). The current form was shown in 1939 by Carl Henry Eckart (1902–1973) and Gale Young.

If we write the SVD of A in the form

$$A = V\Sigma W^H = \left(V\begin{bmatrix} I_m \\ 0_{n-m,m}\end{bmatrix}W^H\right)\left(W\begin{bmatrix}\Sigma_r & 0 \\ 0 & 0_{m-r}\end{bmatrix}W^H\right) =: UP,$$

then $U \in \mathbb{C}^{n,m}$ has orthonormal columns, i.e., $U^HU = I_m$, and $P = P^H \in \mathbb{C}^{m,m}$ is positive semidefinite with the inertia $(r, 0, m-r)$. The factorization $A = UP$ is called a *polar decomposition* of A. It can be viewed as a generalization of the polar representation of complex numbers, $z = e^{i\varphi}|z|$.

Lemma 19.3 *Suppose that the matrix $A \in \mathbb{C}^{n,m}$ with* $\operatorname{rank}(A) = r$ *has an SVD of the form (19.1) with $V = [v_1, \ldots, v_n]$ and $W = [w_1, \ldots, w_m]$. Considering A as an element of $\mathcal{L}(\mathbb{C}^{m,1}, \mathbb{C}^{n,1})$, we then have* $\operatorname{im}(A) = \operatorname{span}\{v_1, \ldots, v_r\}$ *and* $\ker(A) = \operatorname{span}\{w_{r+1}, \ldots, w_m\}$.

Proof For $j = 1, \ldots, r$ we have $Aw_j = V\Sigma W^H w_j = V\Sigma e_j = \sigma_j v_j \neq 0$, since $\sigma_j \neq 0$. Hence these r linear independent vectors satisfy $v_1, \ldots, v_r \in \operatorname{im}(A)$. Now $r = \operatorname{rank}(A) = \dim(\operatorname{im}(A))$ implies that $\operatorname{im}(A) = \operatorname{span}\{v_1, \ldots, v_r\}$.

For $j = r+1, \ldots, m$ we have $Aw_j = 0$, and hence these $m-r$ linear independent vectors satisfy $w_{r+1}, \ldots, w_m \in \ker(A)$. Then $\dim(\ker(A)) = m - \dim(\operatorname{im}(A)) = m - r$ implies that $\ker(A) = \operatorname{span}\{w_{r+1}, \ldots, w_m\}$. □

An SVD of the form (19.1) can be written as

$$A = \sum_{j=1}^{r} \sigma_j v_j w_j^H.$$

Thus, A can be written as a sum of r matrices of the form $\sigma_j v_j w_j^H$, where $\operatorname{rank}\left(\sigma_j v_j w_j^H\right) = 1$. Let

$$A_k := \sum_{j=1}^{k} \sigma_j v_j w_j^H \quad \text{for some } k, \quad 1 \leq k \leq r. \tag{19.2}$$

Then $\operatorname{rank}(A_k) = k$ and, using that the matrix 2-norm is unitarily invariant (cp. Exercise 19.1), we get

$$\|A - A_k\|_2 = \|\operatorname{diag}(\sigma_{k+1}, \ldots, \sigma_r)\|_2 = \sigma_{k+1}. \tag{19.3}$$

Hence A is approximated by the matrix A_k, where the rank of the approximating matrix and the approximation error in the matrix 2-norm are explicitly known. The singular value decomposition, furthermore, yields the *best possible* approximation of A by a matrix of rank k with respect to the matrix 2-norm.

Theorem 19.4 *With A_k as in (19.2), we have $\|A - A_k\|_2 \leq \|A - B\|_2$ for every matrix $B \in \mathbb{C}^{n,m}$ with* $\operatorname{rank}(B) = k$.

Proof The assertion is clear for $k = \text{rank}(A)$, since then $A_k = A$ and $\|A - A_k\|_2 = 0$.

Let $k < \text{rank}(A) \leq m$. Let $B \in \mathbb{C}^{n,m}$ with $\text{rank}(B) = k$ be given, then $\dim(\ker(B)) = m - k$, where we consider B as an element of $\mathcal{L}(\mathbb{C}^{m,1}, \mathbb{C}^{n,1})$. If $w_1, \ldots, w_m$ are the right singular vectors of A from (19.1), then $\mathcal{U} := \text{span}\{w_1, \ldots, w_{k+1}\}$ has the dimension $k + 1$. Since $\ker(B)$ and $\mathcal{U}$ are subspaces of $\mathbb{C}^{m,1}$ with $\dim(\ker(B)) + \dim(\mathcal{U}) = m + 1$, we have $\ker(B) \cap \mathcal{U} \neq \{0\}$.

Let $v \in \ker(B) \cap \mathcal{U}$ with $\|v\|_2 = 1$ be given. Then there exist $\alpha_1, \ldots, \alpha_{k+1} \in \mathbb{C}$ with $v = \sum_{j=1}^{k+1} \alpha_j w_j$ and $\sum_{j=1}^{k+1} |\alpha_j|^2 = \|v\|_2^2 = 1$. Hence

$$(A - B)v = Av - \underbrace{Bv}_{=0} = \sum_{j=1}^{k+1} \alpha_j A w_j = \sum_{j=1}^{k+1} \alpha_j \sigma_j v_j$$

and, therefore,

$$\begin{aligned}
\|A - B\|_2 = \max_{\|y\|_2=1} \|(A - B)y\|_2 &\geq \|(A - B)v\|_2 = \Big\| \sum_{j=1}^{k+1} \alpha_j \sigma_j v_j \Big\|_2 \\
&= \Big(\sum_{j=1}^{k+1} |\alpha_j \sigma_j|^2 \Big)^{1/2} \quad \text{(since } v_1, \ldots, v_{k+1} \text{ are pairwise orthonormal)} \\
&\geq \sigma_{k+1} \Big(\sum_{j=1}^{k+1} |\alpha_j|^2 \Big)^{1/2} \quad \text{(since } \sigma_1 \geq \cdots \geq \sigma_{k+1}) \\
&= \sigma_{k+1} = \|A - A_k\|_2,
\end{aligned}$$

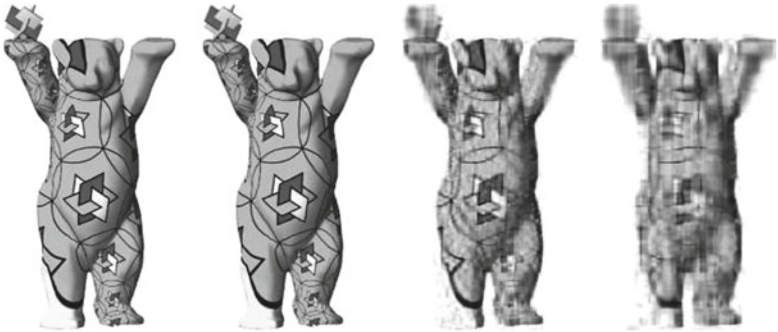
which completes the proof. □

MATLAB-Minute.
The command `A=magic(n)` generates for $n \geq 3$ an $n \times n$ matrix `A` with entries from 1 to n^2, so that all row, column and diagonal sums of `A` are equal. The entries of `A` therefore from a "magic square".
Compute the SVD of `A=magic(10)` using the command `[V,S,W]=svd(A)`. What can be said about the singular values of `A` and what is rank(`A`)? Form A_k for $k = 1, 2, \ldots, \text{rank}(A)$ as in (19.2) and verify numerically the equation (19.3).

The SVD is one of the most important and practical mathematical tools in almost all areas of science, engineering and social sciences, in medicine and even in psychology. Its great importance is due to the fact that the SVD allows to distinguish between "important" and "non-important" information in a given data. In practice, the latter corresponds, e.g., to measurement errors, noise in the transmission of data, or fine details in a signal or an image that do not play an important role. Often, the "important" information corresponds to the large singular values, and the "non-important" information to the small ones.

In many applications one sees, furthermore, that the singular values of a given matrix decay rapidly, so that there exist only few large and many small singular values. If this is the case, then the matrix can be approximated well by a matrix with low rank, since already for a small k the approximation error $\|A - A_k\|_2 = \sigma_{k+1}$ is small. A *low rank approximation* A_k requires little storage capacity in the computer; only k scalars and $2k$ vectors have to be stored. This makes the SVD a powerful tool in all applications where data compression is of interest.

Example 19.5 We illustrate the use of the SVD in image compression with a picture that we obtained from the research center MATHEON: Mathematics for Key Technologies[2]. The greyscale picture is shown on the left of the figure below. It consists of 286×152 pixels, where each of the pixels is given by a value between 0 and 64. These values are stored in a real 286×152 matrix A which has (full) rank 152.



We compute an SVD $A = V\Sigma W^T$ using the command `[V,S,W]=svd(A)` in MATLAB. The diagonal entries of the matrix `S`, i.e., the singular values of A, are ordered decreasingly by MATLAB (as in Theorem 19.1). For $k = 100, 20, 10$ we now compute matrices A_k with rank k as in (19.2) using the command `Ak=V(:,1:k)*S(1:k,1:k)*W(:,1:k)'`. These matrices represent approximations of the original picture based on the k largest singular values and the corresponding singular vectors. The three approximations are shown next to the original picture above. The quality of the approximation decreases with decreasing k, but even the approximation for $k = 10$ shows the essential features of the "MATHEON bear".

Another important application of the SVD arises in the solution of linear systems of equations. If $A \in \mathbb{C}^{n,m}$ has an SVD of the form (19.1), we define the matrix

$$A^\dagger := W\Sigma^\dagger V^H \in \mathbb{C}^{m,n}, \quad \text{where} \quad \Sigma^\dagger := \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m,n}. \qquad (19.4)$$

[2] We thank Falk Ebert for his help. The original bear can be seen in front of the Mathematics building of the TU Berlin. More information on MATHEON can be found at `www.matheon.de`.

One easily sees that

$$A^{\dagger}A = W \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} W^H \in \mathbb{R}^{m,m}.$$

If $r = m = n$, then A is invertible and the right hand side of the above equation is equal to the identity matrix I_n. In this case we have $A^{\dagger} = A^{-1}$. The matrix $A^{\dagger}$ can therefore be viewed as a *generalized inverse*, that in the case of an invertible matrix A is equal to the inverse of A.

Definition 19.6 The matrix $A^{\dagger}$ in (19.4) is called *Moore-Penrose inverse*[3] or *pseudo-inverse* of A.

Let $A \in \mathbb{C}^{n,m}$ and $b \in \mathbb{C}^{n,1}$ be given. If the linear system of equations $Ax = b$ has no solution, then we can try to find an $\widehat{x} \in \mathbb{C}^{m,1}$ such that $A\widehat{x}$ is "as close as possible" to b. Using the Moore-Penrose inverse we obtain the best possible approximation with respect to the Euclidean norm.

Theorem 19.7 *Let $A \in \mathbb{C}^{n,m}$ with $n \geq m$ and $b \in \mathbb{C}^{n,1}$ be given. If $A = V\Sigma W^H$ is an SVD, and $A^{\dagger}$ is as in (19.4), then $\widehat{x} = A^{\dagger}b$ satisfies*

$$\|b - A\widehat{x}\|_2 \leq \|b - Ay\|_2 \quad \textit{for all} \quad y \in \mathbb{C}^{m,1},$$

and

$$\|\widehat{x}\|_2 = \left(\sum_{j=1}^{r} \left| \frac{v_j^H b}{\sigma_j} \right|^2 \right)^{1/2} \leq \|y\|_2$$

for all $y \in \mathbb{C}^{m,1}$ with $\|b - A\widehat{x}\|_2 = \|b - Ay\|_2$.

Proof Let $y \in \mathbb{C}^{m,1}$ be given and let $z = [\xi_1, \ldots, \xi_m]^T := W^H y$. Then

$$\begin{aligned} \|b - Ay\|_2^2 = \|b - V\Sigma W^H y\|_2^2 &= \|V(V^H b - \Sigma z)\|_2^2 = \|V^H b - \Sigma z\|_2^2 \\ &= \sum_{j=1}^{r} \left| v_j^H b - \sigma_j \xi_j \right|^2 + \sum_{j=r+1}^{n} \left| v_j^H b \right|^2 \\ &\geq \sum_{j=r+1}^{n} \left| v_j^H b \right|^2. \end{aligned} \tag{19.5}$$

Equality holds if and only if $\xi_j = \left(v_j^H b\right)/\sigma_j$ for all $j = 1, \ldots, r$. This is satisfied if $z = W^H y = \Sigma^{\dagger} V^H b$. The last equation holds if and only if

$$y = W\Sigma^{\dagger}V^H b = A^{\dagger}b = \widehat{x}.$$

[3] Eliakim Hastings Moore (1862–1932) and Sir Roger Penrose (1931–).

The vector $\widehat{x}$ therefore attains the lower bound (19.5).

The equation

$$\|\widehat{x}\|_2 = \left(\sum_{j=1}^{r} \left| \frac{v_j^H b}{\sigma_j} \right|^2 \right)^{1/2}$$

is easily checked. Every vector $y \in \mathbb{C}^{m,1}$ that attains the lower bound (19.5) must have the form

$$y = W \left[\frac{v_1^H b}{\sigma_1}, \ldots, \frac{v_r^H b}{\sigma_r}, y_{r+1}, \ldots, y_m \right]^T$$

for some $y_{r+1}, \ldots, y_m \in \mathbb{C}$, which implies that $\|y\|_2 \geq \|\widehat{x}\|_2$. □

The minimization problem for the vector $\widehat{x}$ can be written as

$$\|b - A\widehat{x}\|_2 = \min_{y \in \mathbb{C}^{m,1}} \|b - Ay\|_2.$$

If

$$A = \begin{bmatrix} \tau_1 & 1 \\ \vdots & \vdots \\ \tau_m & 1 \end{bmatrix} \in \mathbb{R}^{m,2}$$

for (pairwise distinct) $\tau_1, \ldots, \tau_m \in \mathbb{R}$, then this minimization problem corresponds to the problem of linear regression and the least squares approximation in Example 12.16, that we have solved with the QR-decomposition of A. If $A = QR$ is this decomposition, then $A^\dagger = (A^H A)^{-1} A^H$ (cp. Exercise 19.5) and we have

$$A^\dagger = (R^H Q^H QR)^{-1} R^H Q^H = R^{-1} (R^H)^{-1} R^H Q^H = R^{-1} Q^H.$$

Thus, the solution of the least-squares approximation in Example 12.16 is identical to the solution of the above minimization problem using the SVD of A.

Exercises

19.1 Show that the Frobenius norm and the matrix 2-norm are *unitarily invariant*, i.e., that $\|PAQ\|_F = \|A\|_F$ and $\|PAQ\|_2 = \|A\|_2$ for all $A \in \mathbb{C}^{n,m}$ and unitary matrices $P \in \mathbb{C}^{n,n}$, $Q \in \mathbb{C}^{m,m}$.
(*Hint:* For the Frobenius norm one can use that $\|A\|_F^2 = \text{trace}(A^H A)$.)

19.2 Use the result of Exercise 19.1 to show that $\|A\|_F = \left(\sigma_1^2 + \ldots + \sigma_r^2\right)^{1/2}$ and $\|A\|_2 = \sigma_1$, where $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are the singular values of $A \in \mathbb{C}^{n,m}$.

19.3 Show that $\|A\|_2 = \|A^H\|_2$ and $\|A\|_2^2 = \|A^H A\|_2$ for all $A \in \mathbb{C}^{n,m}$.

19.4 Show that $\|A\|_2^2 \leq \|A\|_1 \, \|A\|_\infty$ for all $A \in \mathbb{C}^{n,m}$.

19.5 Let $A \in \mathbb{C}^{n,m}$ and let $A^\dagger$ be the Moore-Penrose inverse of A. Show the following assertions:

(a) If $\text{rank}(A) = m$, then $A^\dagger = (A^H A)^{-1} A^H$.
(b) The matrix $X = A^\dagger$ is the uniquely determined matrix that satisfies the following four matrix equations:
(1) $AXA = A$,
(2) $XAX = X$,
(3) $(AX)^H = AX$,
(4) $(XA)^H = XA$.

19.6 Let

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & -2 \end{bmatrix} \in \mathbb{R}^{3,2}, \quad b = \begin{bmatrix} 5 \\ 2 \\ -5 \end{bmatrix} \in \mathbb{R}^{3,1}.$$

Compute the Moore-Penrose inverse of A and a vector $\widehat{x} \in \mathbb{R}^{2,1}$ such that

(a) $\|b - A\widehat{x}\|_2 \le \|b - Ay\|_2$ for all $y \in \mathbb{R}^{2,1}$, and
(b) $\|\widehat{x}\|_2 \le \|y\|_2$ for all $y \in \mathbb{R}^{2,1}$ with $\|b - Ay\|_2 = \|b - A\widehat{x}\|_2$.

19.7 Prove the following theorem:

Let $A \in \mathbb{C}^{n,m}$ *and* $B \in \mathbb{C}^{\ell,m}$ *with* $m \le n \le \ell$. *Then* $A^H A = B^H B$ *if and only if* $B = UA$ for a matrix $U \in \mathbb{C}^{\ell,n}$ with $U^H U = I_n$. If A *and* B *are real, then* U *can also be chosen to be real.*

(*Hint:* One direction is trivial. For the other direction consider the unitary diagonalization of $A^H A = B^H B$. This yields the matrix W in the SVD of A and of B. Show the assertion using these two decompositions. This theorem and its applications can be found in the article [HorO96].)