Chapter 22

Singular Value Decomposition and Polar Form

22.1 Properties of $f^* \circ f$

In this section we assume that we are dealing with real Euclidean spaces. Let $f\colon E \to E$ be any linear map. In general, it may not be possible to diagonalize f. We show that every linear map can be diagonalized if we are willing to use *two* orthonormal bases. This is the celebrated *singular value decomposition (SVD)*. A close cousin of the SVD is the *polar form* of a linear map, which shows how a linear map can be decomposed into its purely rotational component (perhaps with a flip) and its purely stretching part.

The key observation is that $f^* \circ f$ is self-adjoint since

$$\langle (f^* \circ f)(u), v\rangle = \langle f(u), f(v)\rangle = \langle u, (f^* \circ f)(v)\rangle.$$

Similarly, $f \circ f^*$ is self-adjoint.

The fact that $f^* \circ f$ and $f \circ f^*$ are self-adjoint is very important, because by Theorem 17.8, it implies that *$f^* \circ f$ and $f \circ f^*$ can be diagonalized and that they have real eigenvalues.* In fact, these *eigenvalues are all nonnegative* as shown in the following proposition.

Proposition 22.1. *The eigenvalues of $f^* \circ f$ and $f \circ f^*$ are nonnegative.*

Proof. If u is an eigenvector of $f^* \circ f$ for the eigenvalue λ, then

$$\langle (f^* \circ f)(u), u\rangle = \langle f(u), f(u)\rangle$$

and

$$\langle (f^* \circ f)(u), u\rangle = \lambda\langle u, u\rangle,$$

and thus

$$\lambda\langle u, u\rangle = \langle f(u), f(u)\rangle,$$

which implies that $\lambda \geq 0$, since $\langle -, -\rangle$ is positive definite. A similar proof applies to $f \circ f^*$. $\square$

Thus, the eigenvalues of $f^* \circ f$ are of the form $\sigma_1^2, \ldots, \sigma_r^2$ or 0, where $\sigma_i > 0$, and similarly for $f \circ f^*$.

The above considerations also apply to any linear map $f\colon E \to F$ between two Euclidean spaces $(E, \langle -, - \rangle_1)$ and $(F, \langle -, - \rangle_2)$. Recall that the adjoint $f^*\colon F \to E$ of f is the unique linear map f^* such that

$$\langle f(u), v \rangle_2 = \langle u, f^*(v) \rangle_1, \quad \text{for all } u \in E \text{ and all } v \in F.$$

Then $f^* \circ f$ and $f \circ f^*$ are self-adjoint (the proof is the same as in the previous case), and the eigenvalues of $f^* \circ f$ and $f \circ f^*$ are nonnegative.

Proof. If λ is an eigenvalue of $f^* \circ f$ and u $(\neq 0)$ is a corresponding eigenvector, we have

$$\langle (f^* \circ f)(u), u \rangle_1 = \langle f(u), f(u) \rangle_2,$$

and also

$$\langle (f^* \circ f)(u), u \rangle_1 = \lambda \langle u, u \rangle_1,$$

so

$$\lambda \langle u, u \rangle_1, = \langle f(u), f(u) \rangle_2,$$

which implies that $\lambda \geq 0$. A similar proof applies to $f \circ f^*$. $\square$

The situation is even better, since we will show shortly that $f^* \circ f$ and $f \circ f^*$ have the same nonzero eigenvalues.

Remark: Given any two linear maps $f\colon E \to F$ and $g\colon F \to E$, where $\dim(E) = n$ and $\dim(F) = m$, it can be shown that

$$\lambda^m \det(\lambda\, I_n - g \circ f) = \lambda^n \det(\lambda\, I_m - f \circ g),$$

and thus $g \circ f$ and $f \circ g$ always have the same nonzero eigenvalues; see Problem 15.14.

Definition 22.1. Given any linear map $f\colon E \to F$, the square roots $\sigma_i > 0$ of the positive eigenvalues of $f^* \circ f$ (and $f \circ f^*$) are called the *singular values of f*.

Definition 22.2. A self-adjoint linear map $f\colon E \to E$ whose eigenvalues are nonnegative is called *positive semidefinite* (or *positive*), and if f is also invertible, f is said to be *positive definite*. In the latter case, every eigenvalue of f is strictly positive.

The following proposition shows that the conditions on the eigenvalues of a self-adjoint linear map used to define the notion of a positive definite linear map is equivalent to the condition used in Definition 8.4. A similar but weaker condition is equivalent to the notion of self-adjoint positive semidefinite linear map.

Proposition 22.2. *Let $f\colon E \to E$ be a self-adjoint linear map, where E is a Euclidean space of finite dimension with inner product $\langle -.- \rangle$.*

(1) The eigenvalues of f are strictly positive iff

$$\langle f(u), u\rangle > 0 \quad \text{for all } u \neq 0.$$

(2) The eigenvalues of f are nonnegative iff

$$\langle f(u), u\rangle \geq 0 \quad \text{for all } u \neq 0.$$

Proof. Since f is self-adjoint, by the spectral theorem (Theorem 17.8), f has real eigenvalues $\lambda_1, \ldots, \lambda_n$, and there is some orthonormal basis $(e_1, \ldots, e_n)$, where e_i is an eigenvector for λ_i. With respect to this basis, every vector $u \in E$ can be written in a unique way as $u = \sum_{i=1}^n x_i u_i$ for some $x_i \in \mathbb{R}$. Since each e_i is eigenvector associated with $\lambda_i \in \mathbb{R}$, we have

$$f\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i f(e_i) = \sum_{i=1}^n \lambda_i x_i e_i,$$

and using the bilinearity of the inner product, we have

$$\begin{aligned}\langle f(u), u\rangle &= \left\langle f\left(\sum_{i=1}^n x_i e_i\right), \sum_{j=1}^n x_j e_j\right\rangle \\ &= \left\langle \sum_{i=1}^n \lambda_i x_i e_i, \sum_{j=1}^n x_j e_j\right\rangle \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i x_i x_j \langle e_i, e_j\rangle,\end{aligned}$$

and since $(e_1, \ldots, e_n)$, is an orthonormal basis, we obtain

$$\langle f(u), u\rangle = \sum_{i=1}^n \lambda_i x_i^2. \tag{$\dagger$}$$

(1) If $\lambda_i > 0$ for $i = 1, \ldots, n$, for any $u \neq 0$, we have $x_i \neq 0$ for some i, so $\langle f(u), u\rangle = \sum_{i=1}^n \lambda_i x_i^2 > 0$.

Conversely, if $\langle f(u), u\rangle > 0$ for all $u \neq 0$, by picking $u = e_i$, we get

$$\langle f(e_i), e_i\rangle = \langle \lambda_i e_i, e_i\rangle = \lambda_i \langle e_i, e_i\rangle = \lambda_i,$$

so $\lambda_i > 0$ for $i = 1, \ldots, n$.

(2) If $\lambda_i \geq 0$ for $i = 1, \ldots, n$, for any $u \neq 0$, then $\langle f(u), u\rangle = \sum_{i=1}^n \lambda_i x_i^2 \geq 0$.

Conversely, if $\langle f(u), u\rangle \geq 0$ for all $u \neq 0$, by picking $u = e_i$, we get

$$\langle f(e_i), e_i\rangle = \langle \lambda_i e_i, e_i\rangle = \lambda_i \langle e_i, e_i\rangle = \lambda_i,$$

so $\lambda_i \geq 0$ for $i = 1, \ldots, n$. ∎

Proposition 22.2 also holds for self-adjoint linear maps on a complex vector space with a Hermitian inner product. The proof is essentially the same and is left as an exercise to the reader.

The version of Proposition 22.2 for matrices follows immediately.

Proposition 22.3. *Let A be a real $n \times n$ symmetric matrix.*

(1) The eigenvalues of A are strictly positive iff

$$u^\top A u > 0 \quad \textit{for all } u \neq 0.$$

(2) The eigenvalues of A are nonnegative iff

$$u^\top A u \geq 0 \quad \textit{for all } u \neq 0.$$

It is important to note that Proposition 22.3 is false for nonsymmetric matrices.

Example 22.1. The matrix

$$A = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$$

has the positive eigenvalues $(1, 1)$, but

$$\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} -3 \\ -1 \end{pmatrix} = -2.$$

Example 22.2. The matrix

$$A = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$$

has the complex eigenvalues $1 + 2i, 1 - 2i$, and yet

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} x - 2y \\ 2x + y \end{pmatrix} = x^2 + y^2,$$

so $u^\top A u > 0$ for all $u \neq 0$.

Since $u^\top A u$ is a scalar, if A is a skew symmetric matrix ($A^\top = -A$), then we see that

$$u^\top A u = 0 \quad \text{for all } u \in \mathbb{R}.$$

Therefore, if A is a real $n \times n$ matrix then

$$u^\top A u = u^\top H(A) u \quad \text{for all } u \in \mathbb{R},$$

where $H(A) = (1/2)(A + A^\top)$ is the symmetric part of A. This explains why the notion of a positive definite matrix is only interesting for symmetric matrices. But but one should also be aware that even if a nonsymmetric matrix A has "well-behaved" eigenvalues, its symmetric part $H(A)$ may *not* be positive definite.

Example 22.3. The matrix
$$A = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}$$
of Example 22.1 has positive eigenvalues $(1, 1)$, but its symmetric part
$$H(A) = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$
is *not* positive definite, since its eigenvalues are $-1, 3$.

Beware that if A is a complex skew-Hermitian matrix, which means that $A^* = -A$, then
$$(u^* A u)^* = -u^* A u,$$
but this only implies that *the real part of $u^* A u$ is zero*. So for any arbitrary complex square matrix A, in general,
$$u^* A u \neq u^* H(A) u,$$
where $H(A) = (1/2)(A + A^*)$.

If $f \colon E \to F$ is any linear map, we just showed that $f^* \circ f$ and $f \circ f^*$ are positive semidefinite self-adjoint linear maps. This fact has the remarkable consequence that every linear map has two important decompositions:

1. The polar form.
2. The singular value decomposition (SVD).

The wonderful thing about the singular value decomposition is that there exist two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ such that, with respect to these bases, f is a diagonal matrix consisting of the singular values of f or 0. Thus, in some sense, f can always be diagonalized with respect to *two* orthonormal bases. The SVD is also a useful tool for solving overdetermined linear systems in the least squares sense and for data analysis, as we show later on.

First we show some useful relationships between the kernels and the images of f, f^*, $f^* \circ f$, and $f \circ f^*$. Recall that if $f \colon E \to F$ is a linear map, the *image* $\mathrm{Im}\, f$ *of* f is the subspace $f(E)$ of F, and the *rank of* f is the dimension $\dim(\mathrm{Im}\, f)$ of its image. Also recall that (Theorem 6.16)
$$\dim(\mathrm{Ker}\, f) + \dim(\mathrm{Im}\, f) = \dim(E),$$
and that (Propositions 12.11 and 14.13) for every subspace W of E,
$$\dim(W) + \dim(W^{\perp}) = \dim(E).$$

Proposition 22.4. *Given any two Euclidean spaces E and F, where E has dimension n and F has dimension m, for any linear map $f\colon E \to F$, we have*

$$\begin{aligned}
\operatorname{Ker} f &= \operatorname{Ker}(f^* \circ f),\\
\operatorname{Ker} f^* &= \operatorname{Ker}(f \circ f^*),\\
\operatorname{Ker} f &= (\operatorname{Im} f^*)^\perp,\\
\operatorname{Ker} f^* &= (\operatorname{Im} f)^\perp,\\
\dim(\operatorname{Im} f) &= \dim(\operatorname{Im} f^*),
\end{aligned}$$

and f, f^*, $f^* \circ f$, and $f \circ f^*$ have the same rank.

Proof. To simplify the notation, we will denote the inner products on E and F by the same symbol $\langle -, - \rangle$ (to avoid subscripts). If $f(u) = 0$, then $(f^* \circ f)(u) = f^*(f(u)) = f^*(0) = 0$, and so $\operatorname{Ker} f \subseteq \operatorname{Ker}(f^* \circ f)$. By definition of f^*, we have

$$\langle f(u), f(u) \rangle = \langle (f^* \circ f)(u), u \rangle$$

for all $u \in E$. If $(f^* \circ f)(u) = 0$, since $\langle -, - \rangle$ is positive definite, we must have $f(u) = 0$, and so $\operatorname{Ker}(f^* \circ f) \subseteq \operatorname{Ker} f$. Therefore,

$$\operatorname{Ker} f = \operatorname{Ker}(f^* \circ f).$$

The proof that $\operatorname{Ker} f^* = \operatorname{Ker}(f \circ f^*)$ is similar.

By definition of f^*, we have

$$\langle f(u), v \rangle = \langle u, f^*(v) \rangle \quad \text{for all } u \in E \text{ and all } v \in F. \tag{$*$}$$

This immediately implies that

$$\operatorname{Ker} f = (\operatorname{Im} f^*)^\perp \quad \text{and} \quad \operatorname{Ker} f^* = (\operatorname{Im} f)^\perp.$$

Let us explain why $\operatorname{Ker} f = (\operatorname{Im} f^*)^\perp$, the proof of the other equation being similar.

Because the inner product is positive definite, for every $u \in E$, we have

- $u \in \operatorname{Ker} f$
- iff $f(u) = 0$
- iff $\langle f(u), v \rangle = 0$ for all v,
- by $(*)$ iff $\langle u, f^*(v) \rangle = 0$ for all v,
- iff $u \in (\operatorname{Im} f^*)^\perp$.

Since
$$\dim(\operatorname{Im} f) = n - \dim(\operatorname{Ker} f)$$
and
$$\dim(\operatorname{Im} f^*) = n - \dim((\operatorname{Im} f^*)^{\perp}),$$
from
$$\operatorname{Ker} f = (\operatorname{Im} f^*)^{\perp}$$
we also have
$$\dim(\operatorname{Ker} f) = \dim((\operatorname{Im} f^*)^{\perp}),$$
from which we obtain
$$\dim(\operatorname{Im} f) = \dim(\operatorname{Im} f^*).$$
Since
$$\dim(\operatorname{Ker}(f^* \circ f)) + \dim(\operatorname{Im}(f^* \circ f)) = \dim(E),$$
$\operatorname{Ker}(f^* \circ f) = \operatorname{Ker} f$ and $\operatorname{Ker} f = (\operatorname{Im} f^*)^{\perp}$, we get
$$\dim((\operatorname{Im} f^*)^{\perp}) + \dim(\operatorname{Im}(f^* \circ f)) = \dim(E).$$
Since
$$\dim((\operatorname{Im} f^*)^{\perp}) + \dim(\operatorname{Im} f^*) = \dim(E),$$
we deduce that
$$\dim(\operatorname{Im} f^*) = \dim(\operatorname{Im}(f^* \circ f)).$$
A similar proof shows that
$$\dim(\operatorname{Im} f) = \dim(\operatorname{Im}(f \circ f^*)).$$
Consequently, f, f^*, $f^* \circ f$, and $f \circ f^*$ have the same rank. $\square$

22.2 Singular Value Decomposition for Square Matrices

We will now prove that every square matrix has an SVD. Stronger results can be obtained if we first consider the polar form and then derive the SVD from it (there are uniqueness properties of the polar decomposition). For our purposes, uniqueness results are not as important so we content ourselves with existence results, whose proofs are simpler. Readers interested in a more general treatment are referred to Gallier [72].

The early history of the singular value decomposition is described in a fascinating paper by Stewart [165]. The SVD is due to Beltrami and Camille Jordan independently (1873, 1874). Gauss is the grandfather of all this, for his work on least squares (1809, 1823) (but Legendre also published a paper on least squares!). Then come Sylvester, Schmidt, and

Hermann Weyl. Sylvester's work was apparently "opaque." He gave a computational method to find an SVD. Schmidt's work really has to do with integral equations and symmetric and asymmetric kernels (1907). Weyl's work has to do with perturbation theory (1912). Autonne came up with the polar decomposition (1902, 1915). Eckart and Young extended SVD to rectangular matrices (1936, 1939).

Theorem 22.5. *(Singular value decomposition) For every real $n \times n$ matrix A there are two orthogonal matrices U and V and a diagonal matrix D such that $A = VDU^\top$, where D is of the form*

$$D = \begin{pmatrix} \sigma_1 & & \ldots & \\ & \sigma_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \sigma_n \end{pmatrix},$$

where $\sigma_1, \ldots, \sigma_r$ are the singular values of A, i.e., the (positive) square roots of the nonzero eigenvalues of $A^\top A$ and $A\,A^\top$, and $\sigma_{r+1} = \cdots = \sigma_n = 0$. The columns of U are eigenvectors of $A^\top A$, and the columns of V are eigenvectors of $A\,A^\top$.

Proof. Since $A^\top A$ is a symmetric matrix, in fact, a positive semidefinite matrix, there exists an orthogonal matrix U such that

$$A^\top A = UD^2U^\top,$$

with $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$, where $\sigma_1^2, \ldots, \sigma_r^2$ are the nonzero eigenvalues of $A^\top A$, and where r is the rank of A; that is, $\sigma_1, \ldots, \sigma_r$ are the singular values of A. It follows that

$$U^\top A^\top AU = (AU)^\top AU = D^2,$$

and if we let f_j be the jth column of AU for $j = 1, \ldots, n$, then we have

$$\langle f_i, f_j \rangle = \sigma_i^2 \delta_{ij}, \quad 1 \leq i, j \leq r$$

and

$$f_j = 0, \quad r + 1 \leq j \leq n.$$

If we define $(v_1, \ldots, v_r)$ by

$$v_j = \sigma_j^{-1} f_j, \quad 1 \leq j \leq r,$$

then we have

$$\langle v_i, v_j \rangle = \delta_{ij}, \quad 1 \leq i, j \leq r,$$

so complete $(v_1, \ldots, v_r)$ into an orthonormal basis $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_n)$ (for example, using Gram–Schmidt). Now since $f_j = \sigma_j v_j$ for $j = 1 \ldots, r$, we have

$$\langle v_i, f_j \rangle = \sigma_j \langle v_i, v_j \rangle = \sigma_j \delta_{i,j}, \quad 1 \leq i \leq n,\ 1 \leq j \leq r$$

and since $f_j = 0$ for $j = r+1, \ldots, n$,

$$\langle v_i, f_j \rangle = 0 \quad 1 \leq i \leq n,\ r+1 \leq j \leq n.$$

If V is the matrix whose columns are $v_1, \ldots, v_n$, then V is orthogonal and the above equations prove that

$$V^\top A U = D,$$

which yields $A = VDU^\top$, as required.

The equation $A = VDU^\top$ implies that

$$A^\top A = UD^2U^\top, \quad AA^\top = VD^2V^\top,$$

which shows that $A^\top A$ and $AA^\top$ have the same eigenvalues, that the columns of U are eigenvectors of $A^\top A$, and that the columns of V are eigenvectors of $AA^\top$. $\square$

Example 22.4. Here is a simple example of how to use the proof of Theorem 22.5 to obtain an SVD decomposition. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. Then $A^\top = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, $A^\top A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, and $AA^\top = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$. A simple calculation shows that the eigenvalues of $A^\top A$ are 2 and 0, and for the eigenvalue 2, a unit eigenvector is $\begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, while a unit eigenvector for the eigenvalue 0 is $\begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$. Observe that the singular values are $\sigma_1 = \sqrt{2}$ and $\sigma_2 = 0$. Furthermore, $U = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} = U^\top$. To determine V, the proof of Theorem 22.5 tells us to first calculate

$$AU = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix},$$

and then set

$$v_1 = (1/\sqrt{2}) \begin{pmatrix} \sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Once v_1 is determined, since $\sigma_2 = 0$, we have the freedom to choose v_2 such that (v_1, v_2) forms an orthonormal basis for $\mathbb{R}^2$. Naturally, we chose $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and set $V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The columns of V are unit eigenvectors of $AA^\top$, but finding V by computing unit eigenvectors of $AA^\top$ does not guarantee that these vectors are consistent with U so that $A = V\Sigma U^\top$. Thus one has to use AU instead. We leave it to the reader to check that

$$A = V \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} U^\top.$$

Theorem 22.5 suggests the following definition.

Definition 22.3. A triple (U, D, V) such that $A = V D U^\top$, where U and V are orthogonal and D is a diagonal matrix whose entries are nonnegative (it is positive semidefinite) is called a *singular value decomposition (SVD) of* A. If $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, it is customary to assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$.

The `Matlab` command for computing an SVD $A = VDU^\top$ of a matrix A is
`[V, D, U] = svd(A)`.

The proof of Theorem 22.5 shows that there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$, where $(u_1, \ldots, u_n)$ are eigenvectors of $A^\top A$ and $(v_1, \ldots, v_n)$ are eigenvectors of $AA^\top$. Furthermore, $(u_1, \ldots, u_r)$ is an orthonormal basis of $\mathrm{Im}\, A^\top$, $(u_{r+1}, \ldots, u_n)$ is an orthonormal basis of $\mathrm{Ker}\, A$, $(v_1, \ldots, v_r)$ is an orthonormal basis of $\mathrm{Im}\, A$, and $(v_{r+1}, \ldots, v_n)$ is an orthonormal basis of $\mathrm{Ker}\, A^\top$.

Using a remark made in Chapter 4, if we denote the columns of U by $u_1, \ldots, u_n$ and the columns of V by $v_1, \ldots, v_n$, then we can write

$$A = V D U^\top = \sigma_1 v_1 u_1^\top + \cdots + \sigma_r v_r u_r^\top,$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. As a consequence, if r is a lot smaller than n (we write $r \ll n$), we see that A can be reconstructed from U and V using a much smaller number of elements. This idea will be used to provide "low-rank" approximations of a matrix. The idea is to keep only the k top singular values for some suitable $k \ll r$ for which $\sigma_{k+1}, \ldots, \sigma_r$ are very small.

Remarks:

(1) In Strang [170] the matrices U, V, D are denoted by $U = Q_2$, $V = Q_1$, and $D = \Sigma$, and an SVD is written as $A = Q_1 \Sigma Q_2^\top$. This has the advantage that Q_1 comes before Q_2 in $A = Q_1 \Sigma Q_2^\top$. This has the disadvantage that A maps the columns of Q_2 (eigenvectors of $A^\top A$) to multiples of the columns of Q_1 (eigenvectors of $A\, A^\top$).

(2) Algorithms for actually computing the SVD of a matrix are presented in Golub and Van Loan [80], Demmel [48], and Trefethen and Bau [176], where the SVD and its applications are also discussed quite extensively.

(3) If A is a symmetric matrix, then in general, there is no SVD $V \Sigma U^\top$ of A with $V = U$. However, if A is positive semidefinite, then the eigenvalues of A are nonnegative, and so the nonzero eigenvalues of A are equal to the singular values of A and SVDs of A are of the form
$$A = V \Sigma V^\top.$$

(4) The SVD also applies to complex matrices. In this case, for every complex $n \times n$ matrix A, there are two unitary matrices U and V and a diagonal matrix D such that
$$A = V D U^*,$$
where D is a diagonal matrix consisting of real entries $\sigma_1, \ldots, \sigma_n$, where $\sigma_1 \geq \cdots \geq \sigma_r$ are the singular values of A, i.e., the positive square roots of the nonzero eigenvalues of $A^* A$ and $A\, A^*$, and $\sigma_{r+1} = \ldots = \sigma_n = 0$.

22.3 Polar Form for Square Matrices

A notion closely related to the SVD is the polar form of a matrix.

Definition 22.4. A pair (R, S) such that $A = RS$ with R orthogonal and S symmetric positive semidefinite is called a *polar decomposition of* A.

Theorem 22.5 implies that for every real $n \times n$ matrix A, there is some orthogonal matrix R and some positive semidefinite symmetric matrix S such that

$$A = RS.$$

This is easy to show and we will prove it below. Furthermore, R, S are unique if A is invertible, but this is harder to prove; see Problem 22.9.

For example, the matrix

$$A = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

is both orthogonal and symmetric, and $A = RS$ with $R = A$ and $S = I$, which implies that some of the eigenvalues of A are negative.

Remark: In the complex case, the polar decomposition states that for every complex $n \times n$ matrix A, there is some unitary matrix U and some positive semidefinite Hermitian matrix H such that

$$A = UH.$$

It is easy to go from the polar form to the SVD, and conversely.

Given an SVD decomposition $A = V D U^\top$, let $R = V U^\top$ and $S = U D U^\top$. It is clear that R is orthogonal and that S is positive semidefinite symmetric, and

$$RS = V U^\top U D U^\top = V D U^\top = A.$$

Example 22.5. Recall from Example 22.4 that $A = VDU^\top$ where $V = I_2$ and

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \qquad D = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}.$$

Set $R = VU^\top = U$ and

$$S = UDU^\top = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Since $S = \frac{1}{\sqrt{2}} A^\top A$, S has eigenvalues $\sqrt{2}$ and 0. We leave it to the reader to check that $A = RS$.

Going the other way, given a polar decomposition $A = R_1 S$, where R_1 is orthogonal and S is positive semidefinite symmetric, there is an orthogonal matrix R_2 and a positive semidefinite diagonal matrix D such that $S = R_2 D\, R_2^\top$, and thus

$$A = R_1 R_2 D\, R_2^\top = V D\, U^\top,$$

where $V = R_1 R_2$ and $U = R_2$ are orthogonal.

Example 22.6. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and $A = R_1 S$, where $R_1 = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$ and $S = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$. This is the polar decomposition of Example 22.5. Observe that

$$S = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = R_2 D R_2^\top.$$

Set $U = R_2$ and $V = R_1 R_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ to obtain the SVD decomposition of Example 22.4.

The eigenvalues and the singular values of a matrix are typically not related in any obvious way. For example, the $n \times n$ matrix

$$A = \begin{pmatrix} 1 & 2 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 2 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 & 2 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 & 2 \\ 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{pmatrix}$$

has the eigenvalue 1 with multiplicity n, but its singular values, $\sigma_1 \geq \cdots \geq \sigma_n$, which are the positive square roots of the eigenvalues of the matrix $B = A^\top A$ with

$$B = \begin{pmatrix} 1 & 2 & 0 & 0 & \dots & 0 & 0 \\ 2 & 5 & 2 & 0 & \dots & 0 & 0 \\ 0 & 2 & 5 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 2 & 5 & 2 & 0 \\ 0 & 0 & \dots & 0 & 2 & 5 & 2 \\ 0 & 0 & \dots & 0 & 0 & 2 & 5 \end{pmatrix}$$

have a wide spread, since

$$\frac{\sigma_1}{\sigma_n} = \mathrm{cond}_2(A) \geq 2^{n-1}.$$

If A is a complex $n \times n$ matrix, the eigenvalues $\lambda_1, \ldots, \lambda_n$ and the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ of A are not unrelated, since

$$\sigma_1^2 \cdots \sigma_n^2 = \det(A^* A) = |\det(A)|^2$$

and

$$|\lambda_1| \cdots |\lambda_n| = |\det(A)|,$$

so we have

$$|\lambda_1| \cdots |\lambda_n| = \sigma_1 \cdots \sigma_n.$$

More generally, Hermann Weyl proved the following remarkable theorem:

Theorem 22.6. *(Weyl's inequalities, 1949) For any complex $n \times n$ matrix, A, if $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ are the eigenvalues of A and $\sigma_1, \ldots, \sigma_n \in \mathbb{R}_+$ are the singular values of A, listed so that $|\lambda_1| \geq \cdots \geq |\lambda_n|$ and $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$, then*

$$\begin{aligned} &|\lambda_1| \cdots |\lambda_n| = \sigma_1 \cdots \sigma_n \quad \textit{and} \\ &|\lambda_1| \cdots |\lambda_k| \leq \sigma_1 \cdots \sigma_k, \quad \textit{for} \quad k = 1, \ldots, n-1. \end{aligned}$$

A proof of Theorem 22.6 can be found in Horn and Johnson [96], Chapter 3, Section 3.3, where more inequalities relating the eigenvalues and the singular values of a matrix are given.

Theorem 22.5 can be easily extended to rectangular $m \times n$ matrices, as we show in the next section. For various versions of the SVD for rectangular matrices, see Strang [170] Golub and Van Loan [80], Demmel [48], and Trefethen and Bau [176].

22.4 Singular Value Decomposition for Rectangular Matrices

Here is the generalization of Theorem 22.5 to rectangular matrices.

Theorem 22.7. *(Singular value decomposition) For every real $m \times n$ matrix A, there are two orthogonal matrices U $(n \times n)$ and V $(m \times m)$ and a diagonal $m \times n$ matrix D such that $A = VDU^\top$, where D is of the form*

$$D = \begin{pmatrix} \sigma_1 & & \ldots & \\ & \sigma_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \sigma_n \\ 0 & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \ldots & 0 \end{pmatrix} \quad \textit{or} \quad D = \begin{pmatrix} \sigma_1 & & \ldots & & 0 & \ldots & 0 \\ & \sigma_2 & \ldots & & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \vdots & 0 \\ & & \ldots & \sigma_m & 0 & \ldots & 0 \end{pmatrix},$$

where $\sigma_1, \ldots, \sigma_r$ are the singular values of A, i.e. the (positive) square roots of the nonzero eigenvalues of $A^\top A$ and $A\,A^\top$, and $\sigma_{r+1} = \ldots = \sigma_p = 0$, where $p = \min(m, n)$. The columns of U are eigenvectors of $A^\top A$, and the columns of V are eigenvectors of $A\,A^\top$.

Proof. As in the proof of Theorem 22.5, since $A^\top A$ is symmetric positive semidefinite, there exists an $n \times n$ orthogonal matrix U such that

$$A^\top A = U\Sigma^2 U^\top,$$

with $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$, where $\sigma_1^2, \ldots, \sigma_r^2$ are the nonzero eigenvalues of $A^\top A$, and where r is the rank of A. Observe that $r \leq \min\{m, n\}$, and AU is an $m \times n$ matrix. It follows that

$$U^\top A^\top AU = (AU)^\top AU = \Sigma^2,$$

and if we let $f_j \in \mathbb{R}^m$ be the jth column of AU for $j = 1, \ldots, n$, then we have

$$\langle f_i, f_j\rangle = \sigma_i^2 \delta_{ij}, \quad 1 \leq i, j \leq r$$

and

$$f_j = 0, \quad r + 1 \leq j \leq n.$$

If we define $(v_1, \ldots, v_r)$ by

$$v_j = \sigma_j^{-1} f_j, \quad 1 \leq j \leq r,$$

then we have

$$\langle v_i, v_j\rangle = \delta_{ij}, \quad 1 \leq i, j \leq r,$$

so complete $(v_1, \ldots, v_r)$ into an orthonormal basis $(v_1, \ldots, v_r, v_{r+1}, \ldots, v_m)$ (for example, using Gram–Schmidt).

Now since $f_j = \sigma_j v_j$ for $j = 1 \ldots, r$, we have

$$\langle v_i, f_j\rangle = \sigma_j \langle v_i, v_j\rangle = \sigma_j \delta_{i,j}, \quad 1 \leq i \leq m,\ 1 \leq j \leq r$$

and since $f_j = 0$ for $j = r + 1, \ldots, n$, we have

$$\langle v_i, f_j\rangle = 0 \quad 1 \leq i \leq m,\ r + 1 \leq j \leq n.$$

If V is the matrix whose columns are $v_1, \ldots, v_m$, then V is an $m \times m$ orthogonal matrix and if $m \geq n$, we let

$$D = \begin{pmatrix} \Sigma \\ 0_{m-n} \end{pmatrix} = \begin{pmatrix} \sigma_1 & & \ldots & \\ & \sigma_2 & \ldots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \ldots & \sigma_n \\ 0 & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \vdots & \ldots & 0 \end{pmatrix},$$

else if $n \geq m$, then we let

$$D = \begin{pmatrix} \sigma_1 & & \ldots & & 0 & \ldots & 0 \\ & \sigma_2 & \ldots & & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 & \vdots & 0 \\ & & \ldots & \sigma_m & 0 & \ldots & 0 \end{pmatrix}.$$

In either case, the above equations prove that

$$V^\top AU = D,$$

which yields $A = VDU^\top$, as required.

The equation $A = VDU^\top$ implies that

$$A^\top A = UD^\top DU^\top = U\mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2, \underbrace{0, \ldots, 0}_{n-r})U^\top$$

and

$$AA^\top = VDD^\top V^\top = V\mathrm{diag}(\sigma_1^2, \ldots, \sigma_r^2, \underbrace{0, \ldots, 0}_{m-r})V^\top,$$

which shows that $A^\top A$ and $AA^\top$ have the same nonzero eigenvalues, that the columns of U are eigenvectors of $A^\top A$, and that the columns of V are eigenvectors of $AA^\top$. $\square$

A triple (U, D, V) such that $A = V D U^\top$ is called a *singular value decomposition (SVD) of* A. If $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ (with $p = \min(m, n)$), it is customary to assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$.

Example 22.7. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$. Then $A^\top = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ $A^\top A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, and $AA^\top = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. The reader should verify that $A^\top A = U\Sigma^2 U^\top$ where $\Sigma^2 = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ and $U = U^\top = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}$. Since $AU = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$, set $v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, and complete an orthonormal basis for $\mathbb{R}^3$ by assigning $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, and $v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Thus $V = I_3$, and the reader should verify that $A = VDU^\top$, where $D = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Even though the matrix D is an $m \times n$ rectangular matrix, since its only nonzero entries are on the descending diagonal, we still say that D is a diagonal matrix.

The `Matlab` command for computing an SVD $A = VDU^\top$ of a matrix A is `[V, D, U] = svd(A)`. Beware that `Matlab` uses the convention that the SVD of a matrix A is written as $A = UDV^\top$, and so the call for this version of the SVD is `[U, D, V] = svd(A)`.

If we view A as the representation of a linear map $f \colon E \to F$, where $\dim(E) = n$ and $\dim(F) = m$, the proof of Theorem 22.7 shows that there are two orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ for E and F, respectively, where $(u_1, \ldots, u_n)$ are eigenvectors of $f^* \circ f$ and $(v_1, \ldots, v_m)$ are eigenvectors of $f \circ f^*$. Furthermore, $(u_1, \ldots, u_r)$ is an orthonormal basis of $\mathrm{Im}\, f^*$, $(u_{r+1}, \ldots, u_n)$ is an orthonormal basis of $\mathrm{Ker}\, f$, $(v_1, \ldots, v_r)$ is an orthonormal basis of $\mathrm{Im}\, f$, and $(v_{r+1}, \ldots, v_m)$ is an orthonormal basis of $\mathrm{Ker}\, f^*$.

The SVD of matrices can be used to define the pseudo-inverse of a rectangular matrix; we will do so in Chapter 23. The reader may also consult Strang [170], Demmel [48], Trefethen and Bau [176], and Golub and Van Loan [80].

One of the spectral theorems states that a symmetric matrix can be diagonalized by an orthogonal matrix. There are several numerical methods to compute the eigenvalues of a symmetric matrix A. One method consists in *tridiagonalizing* A, which means that there exists some orthogonal matrix P and some symmetric tridiagonal matrix T such that $A = PTP^\top$. In fact, this can be done using Householder transformations; see Theorem 18.2. It is then possible to compute the eigenvalues of T using a bisection method based on Sturm sequences. One can also use Jacobi's method. For details, see Golub and Van Loan [80], Chapter 8, Demmel [48], Trefethen and Bau [176], Lecture 26, Ciarlet [41], and Chapter 18. Computing the SVD of a matrix A is more involved. Most methods begin by finding orthogonal matrices U and V and a *bidiagonal* matrix B such that $A = VBU^\top$; see Problem 13.8 and Problem 22.3. This can also be done using Householder transformations. Observe that $B^\top B$ is symmetric tridiagonal. Thus, in principle, the previous method to diagonalize a symmetric tridiagonal matrix can be applied. However, it is unwise to compute $B^\top B$ explicitly, and more subtle methods are used for this last step; the matrix of Problem 22.1 can be used, and see Problem 22.3. Again, see Golub and Van Loan [80], Chapter 8, Demmel [48], and Trefethen and Bau [176], Lecture 31.

The polar form has applications in continuum mechanics. Indeed, in any deformation it is important to separate stretching from rotation. This is exactly what QS achieves. The orthogonal part Q corresponds to rotation (perhaps with an additional reflection), and the symmetric matrix S to stretching (or compression). The real eigenvalues $\sigma_1, \ldots, \sigma_r$ of S are the stretch factors (or compression factors) (see Marsden and Hughes [120]). The fact that S can be diagonalized by an orthogonal matrix corresponds to a natural choice of axes, the principal axes.

The SVD has applications to data compression, for instance in image processing. The idea is to retain only singular values whose magnitudes are significant enough. The SVD can also be used to determine the rank of a matrix when other methods such as Gaussian elimination produce very small pivots. One of the main applications of the SVD is the computation of the pseudo-inverse. Pseudo-inverses are the key to the solution of various optimization problems, in particular the method of least squares. This topic is discussed in the next chapter (Chapter 23). Applications of the material of this chapter can be found in Strang [170, 169]; Ciarlet [41]; Golub and Van Loan [80], which contains many other references; Demmel [48]; and Trefethen and Bau [176].

22.5 Ky Fan Norms and Schatten Norms

The singular values of a matrix can be used to define various norms on matrices which have found recent applications in quantum information theory and in spectral graph theory. Following Horn and Johnson [96] (Section 3.4) we can make the following definitions:

Definition 22.5. For any matrix $A \in \mathrm{M}_{m,n}(\mathbb{C})$, let $q = \min\{m, n\}$, and if $\sigma_1 \geq \cdots \geq \sigma_q$ are the singular values of A, for any k with $1 \leq k \leq q$, let

$$N_k(A) = \sigma_1 + \cdots + \sigma_k,$$

called the *Ky Fan k-norm* of A.

More generally, for any $p \geq 1$ and any k with $1 \leq k \leq q$, let

$$N_{k;p}(A) = (\sigma_1^p + \cdots + \sigma_k^p)^{1/p},$$

called the *Ky Fan p-k-norm* of A. When $k = q$, $N_{q;p}$ is also called the *Schatten p-norm*.

Observe that when $k = 1$, $N_1(A) = \sigma_1$, and the Ky Fan norm N_1 is simply the *spectral norm* from Chapter 9, which is the subordinate matrix norm associated with the Euclidean norm. When $k = q$, the Ky Fan norm N_q is given by

$$N_q(A) = \sigma_1 + \cdots + \sigma_q = \mathrm{tr}((A^*A)^{1/2})$$

and is called the *trace norm* or *nuclear norm*. When $p = 2$ and $k = q$, the Ky Fan $N_{q;2}$ norm is given by

$$N_{k;2}(A) = (\sigma_1^2 + \cdots + \sigma_q^2)^{1/2} = \sqrt{\mathrm{tr}(A^*A)} = \|A\|_F\,,$$

which is the *Frobenius norm* of A.

It can be shown that N_k and $N_{k;p}$ are unitarily invariant norms, and that when $m = n$, they are matrix norms; see Horn and Johnson [96] (Section 3.4, Corollary 3.4.4 and Problem 3).

22.6 Summary

The main concepts and results of this chapter are listed below:

- For any linear map $f\colon E \to E$ on a Euclidean space E, the maps $f^* \circ f$ and $f \circ f^*$ are self-adjoint and positive semidefinite.
- The *singular values* of a linear map.
- *Positive semidefinite* and *positive definite* self-adjoint maps.
- Relationships between $\mathrm{Im}\, f$, $\mathrm{Ker}\, f$, $\mathrm{Im}\, f^*$, and $\mathrm{Ker}\, f^*$.
- The *singular value decomposition theorem* for square matrices (Theorem 22.5).
- The *SVD* of matrix.
- The *polar decomposition* of a matrix.
- The *Weyl inequalities*.
- The *singular value decomposition theorem* for $m \times n$ matrices (Theorem 22.7).
- Ky Fan k-norms, Ky Fan p-k-norms, Schatten p-norms.

22.7 Problems

Problem 22.1. (1) Let A be a real $n \times n$ matrix and consider the $(2n) \times (2n)$ real symmetric matrix

$$S = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix}.$$

Suppose that A has rank r. If $A = V\Sigma U^\top$ is an SVD for A, with $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and $\sigma_1 \geq \cdots \geq \sigma_r > 0$, denoting the columns of U by u_k and the columns of V by v_k, prove that σ_k is an eigenvalue of S with corresponding eigenvector $\begin{pmatrix} v_k \\ u_k \end{pmatrix}$ for $k = 1, \ldots, n$, and that $-\sigma_k$ is an eigenvalue of S with corresponding eigenvector $\begin{pmatrix} v_k \\ -u_k \end{pmatrix}$ for $k = 1, \ldots, n$.

Hint. We have $Au_k = \sigma_k v_k$ for $k = 1, \ldots, n$. Show that $A^\top v_k = \sigma_k u_k$ for $k = 1, \ldots, n$.

(2) Prove that the $2n$ eigenvectors of S in (1) are pairwise orthogonal. Check that if A has rank r, then S has rank $2r$.

(3) Now assume that A is a real $m \times n$ matrix and consider the $(m + n) \times (m + n)$ real symmetric matrix

$$S = \begin{pmatrix} 0 & A \\ A^\top & 0 \end{pmatrix}.$$