Chapter 16

Singular Value Decomposition and Polar Form

16.1 Properties of $f^* \circ f$

Let $f\colon E \to E$ be any linear map, where E is a Euclidean space.

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 it implies that $f^* \circ f$ and $f \circ f^*$ *can be diagonalized and that they have real eigenvalues*.

Proposition 16.1. *The eigenvalues of $f^* \circ f$ and $f \circ f^*$ are nonnegative.*

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 $f \colon E \to F$ betwen 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, and the eigenvalues of $f^* \circ f$ and $f \circ f^*$ are nonnegative (the proof is the same as in the previous case), The situation is even better, since we will show shortly that $f^* \circ f$ and $f \circ f^*$ have the *same 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!

Definition 16.1. 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 16.2. A self-adjoint linear map $f: E \to E$ 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.

Recall that Proposition 14.24 implies that a symmetric matrix is positive definite iff its eigenvalues are positive, so Definition 16.2 is equivalent to Definition 6.1 (for matrices).

If $f: E \to F$ 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.

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

$$\dim\,(\mathrm{Ker}\, f) + \dim\,(\mathrm{Im}\, f) = \dim\,(E),$$

and that for every subspace W of E,

$$\dim\,(W) + \dim\,(W^{\perp}) = \dim\,(E).$$

Proposition 16.2. *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.

We will now prove that every square matrix has an SVD.

16.2 Singular Value Decomposition for Square Matrices

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.

The early history of the singular value decomposition is described in a fascinating paper by Stewart [30].

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 16.3. *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 $A\,A^\top$.

Theorem 16.3 suggests the following definition.

Definition 16.3. A triple (U, D, V) such that $A = VDU^\top$, 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*.

The proof of Theorem 16.3 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 2, 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

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

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 [32] 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 [17], Demmel [11], and Trefethen and Bau [34], 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 $U\Sigma V^\top$ of A with $U = V$.

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

$$A = U\Sigma U^\top.$$

(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 $\sigma_1, \ldots, \sigma_r$ are the singular values of A, i.e., the positive square roots of the nonzero eigenvalues of A^*A and $A\,A^*$, and $\sigma_{r+1} = \ldots = \sigma_n = 0$.

16.3 Polar Form for Square Matrices

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

Definition 16.4. A pair (R, S) such that $A = RS$ with R orthogonal and S symmetric positive semidefinite is called a *polar decomposition of A*.

Theorem 16.3 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

$$A = RS.$$

Furthermore, R, S are unique if A is invertible, but this is harder to prove.

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

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

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.

Theorem 16.3 can be easily extended to rectangular $m \times n$ matrices (see Strang [32] or Golub and Van Loan [17], Demmel [11], Trefethen and Bau [34]).

16.4 Singular Value Decomposition for Rectangular Matrices

Theorem 16.4. *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 $A = V D U^\top$, where D is of the form*

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

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 $A A^\top$.

A triple (U, D, V) such that $A = VDU^\top$ is called a *singular value decomposition (SVD) of A*.

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.

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 16.4 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 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

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

More generally, Hermann Weyl proved the following remarkable theorem:

Theorem 16.5. *(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 16.5 can be found in Horn and Johnson [20], Chapter 3, Section 3.3, where more inequalities relating the eigenvalues and the singular values of a matrix are given.

The SVD of matrices can be used to define the pseudo-inverse of a rectangular matrix.

Computing the SVD of a matrix A is quite involved. Most methods begin by finding orthogonal matrices U and V and a *bidiagonal* matrix B such that $A = VBU^{\top}$.

16.5 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 [20] (Section 3.4) we can make the following definitions:

Definition 16.5. 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 7, 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 [20] (Section 3.4, Corollary 3.4.4 and Problem 3).