Appendix C

Linear Algebra

C.1 BASIC DEFINITIONS

In this chapter we only deal with linear algebra over finite dimensional Euclidean spaces. We refer to vectors as column vectors.

Given two d dimensional vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, their inner product is

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{d} u_i v_i .$$

The Euclidean norm (a.k.a. the ℓ_2 norm) is $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$. We also use the ℓ_1 norm, $\|\mathbf{u}\|_1 = \sum_{i=1}^{d} |u_i|$ and the ℓ_∞ norm $\|\mathbf{u}\|_\infty = \max_i |u_i|$.

A subspace of $\mathbb{R}^d$ is a subset of $\mathbb{R}^d$ which is closed under addition and scalar multiplication. The span of a set of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is the subspace containing all vectors of the form

$$\sum_{i=1}^{k} \alpha_i \mathbf{u}_i$$

where for all i, $\alpha_i \in \mathbb{R}$.

A set of vectors $U = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is independent if for every i, $\mathbf{u}_i$ is not in the span of $\mathbf{u}_1, \ldots, \mathbf{u}_{i-1}, \mathbf{u}_{i+1}, \ldots, \mathbf{u}_k$. We say that U spans a subspace V if V is the span of the vectors in U. We say that U is a *basis* of V if it is both independent and spans V. The dimension of V is the size of a basis of V (and it can be verified that all bases of V have the same size). We say that U is an orthogonal set if for all $i \neq j$, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$. We say that U is an orthonormal set if it is orthogonal and if for every i, $\|\mathbf{u}_i\| = 1$.

Given a matrix $A \in \mathbb{R}^{n,d}$, the range of A is the span of its columns and the null space of A is the subspace of all vectors that satisfy $A\mathbf{u} = \mathbf{0}$. The rank of A is the dimension of its range.

The transpose of a matrix A, denoted $A^\top$, is the matrix whose (i, j) entry equals the (j, i) entry of A. We say that A is symmetric if $A = A^\top$.

C.2 EIGENVALUES AND EIGENVECTORS

Let $A \in \mathbb{R}^{d,d}$ be a matrix. A nonzero vector $\mathbf{u}$ is an eigenvector of A with a corresponding eigenvalue λ if

$$A\mathbf{u} = \lambda \mathbf{u}.$$

Theorem C.1 (Spectral Decomposition). *If $A \in \mathbb{R}^{d,d}$ is a symmetric matrix of rank k, then there exists an orthonormal basis of $\mathbb{R}^d$, $\mathbf{u}_1, \ldots, \mathbf{u}_d$, such that each $\mathbf{u}_i$ is an eigenvector of A. Furthermore, A can be written as $A = \sum_{i=1}^{d} \lambda_i \mathbf{u}_i \mathbf{u}_i^\top$, where each λ_i is the eigenvalue corresponding to the eigenvector $\mathbf{u}_i$. This can be written equivalently as $A = UDU^\top$, where the columns of U are the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_d$, and D is a diagonal matrix with $D_{i,i} = \lambda_i$ and for $i \neq j$, $D_{i,j} = 0$. Finally, the number of λ_i which are nonzero is the rank of the matrix, the eigenvectors which correspond to the nonzero eigenvalues span the range of A, and the eigenvectors which correspond to zero eigenvalues span the null space of A.*

C.3 POSITIVE DEFINITE MATRICES

A symmetric matrix $A \in \mathbb{R}^{d,d}$ is positive definite if all its eigenvalues are positive. A is positive semidefinite if all its eigenvalues are nonnegative.

Theorem C.2. *Let $A \in \mathbb{R}^{d,d}$ be a symmetric matrix. Then, the following are equivalent definitions of positive semidefiniteness of A:*

- *All the eigenvalues of A are nonnegative.*
- *For every vector $\mathbf{u}$, $\langle \mathbf{u}, A\mathbf{u} \rangle \geq 0$.*
- *There exists a matrix B such that $A = BB^\top$.*

C.4 SINGULAR VALUE DECOMPOSITION (SVD)

Let $A \in \mathbb{R}^{m,n}$ be a matrix of rank r. When $m \neq n$, the eigenvalue decomposition given in Theorem C.1 cannot be applied. We will describe another decomposition of A, which is called Singular Value Decomposition, or SVD for short.

Unit vectors $\mathbf{v} \in \mathbb{R}^n$ and $\mathbf{u} \in \mathbb{R}^m$ are called right and left *singular vectors* of A with corresponding *singular value* $\sigma > 0$ if

$$A\mathbf{v} = \sigma \mathbf{u} \quad \text{and} \quad A^\top \mathbf{u} = \sigma \mathbf{v}.$$

We first show that if we can find r orthonormal singular vectors with positive singular values, then we can decompose $A = UDV^\top$, with the columns of U and V containing the left and right singular vectors, and D being a diagonal $r \times r$ matrix with the singular values on its diagonal.

Lemma C.3. *Let $A \in \mathbb{R}^{m,n}$ be a matrix of rank r. Assume that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is an orthonormal set of right singular vectors of A, $\mathbf{u}_1, \ldots, \mathbf{u}_r$ is an orthonormal set of corresponding left singular vectors of A, and $\sigma_1, \ldots, \sigma_r$ are the corresponding singular* *values. Then,*

$$A = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top .$$

It follows that if U is a matrix whose columns are the $\mathbf{u}_i$'s, V is a matrix whose columns are the $\mathbf{v}_i$'s, and D is a diagonal matrix with $D_{i,i} = \sigma_i$, then

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

Proof. Any right singular vector of A must be in the range of $A^\top$ (otherwise, the singular value will have to be zero). Therefore, $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is an orthonormal basis of the range of A. Let us complete it to an orthonormal basis of $\mathbb{R}^n$ by adding the vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$. Define $B = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$. It suffices to prove that for all i, $A\mathbf{v}_i = B\mathbf{v}_i$. Clearly, if $i > r$ then $A\mathbf{v}_i = 0$ and $B\mathbf{v}_i = 0$ as well. For $i \le r$ we have

$$B\mathbf{v}_i = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^\top \mathbf{v}_i = \sigma_i \mathbf{u}_i = A\mathbf{v}_i ,$$

where the last equality follows from the definition. □

The next lemma relates the singular values of A to the eigenvalues of $A^\top A$ and $AA^\top$.

Lemma C.4. *$\mathbf{v}, \mathbf{u}$ are right and left singular vectors of A with singular value σ iff $\mathbf{v}$ is an eigenvector of $A^\top A$ with corresponding eigenvalue σ^2 and $\mathbf{u} = \sigma^{-1} A\mathbf{v}$ is an eigenvector of $AA^\top$ with corresponding eigenvalue σ^2.*

Proof. Suppose that σ is a singular value of A with $\mathbf{v} \in \mathbb{R}^n$ being the corresponding right singular vector. Then,

$$A^\top A\mathbf{v} = \sigma A^\top \mathbf{u} = \sigma^2 \mathbf{v}.$$

Similarly,

$$AA^\top u = \sigma A\mathbf{v} = \sigma^2 \mathbf{u}.$$

For the other direction, if $\lambda \neq 0$ is an eigenvalue of $A^\top A$, with $\mathbf{v}$ being the corresponding eigenvector, then $\lambda > 0$ because $A^\top A$ is positive semidefinite. Let $\sigma = \sqrt{\lambda}, \mathbf{u} = \sigma^{-1} A\mathbf{v}$. Then,

$$\sigma \mathbf{u} = \sqrt{\lambda} \frac{A\mathbf{v}}{\sqrt{\lambda}} = A\mathbf{v},$$

and

$$A^\top \mathbf{u} = \frac{1}{\sigma} A^\top A\mathbf{v} = \frac{\lambda}{\sigma} \mathbf{v} = \sigma \mathbf{v}.$$

□

Finally, we show that if A has rank r then it has r orthonormal singular vectors.

Lemma C.5. *Let $A \in \mathbb{R}^{m,n}$ with rank r. Define the following vectors:*

$$\mathbf{v}_1 = \operatorname*{argmax}_{\mathbf{v}\in\mathbb{R}^n:\|\mathbf{v}\|=1} \|A\mathbf{v}\|$$

$$\mathbf{v}_2 = \operatorname*{argmax}_{\substack{\mathbf{v}\in\mathbb{R}^n:\|\mathbf{v}\|=1\\ \langle\mathbf{v},\mathbf{v}_1\rangle=0}} \|A\mathbf{v}\|$$

$$\vdots$$

$$\mathbf{v}_r = \operatorname*{argmax}_{\substack{\mathbf{v}\in\mathbb{R}^n:\|\mathbf{v}\|=1\\ \forall i<r,\ \langle\mathbf{v},\mathbf{v}_i\rangle=0}} \|A\mathbf{v}\|$$

Then, $\mathbf{v}_1,\ldots,\mathbf{v}_r$ is an orthonormal set of right singular vectors of A.

Proof. First note that since the rank of A is r, the range of A is a subspace of dimension r, and therefore it is easy to verify that for all $i = 1,\ldots,r$, $\|A\mathbf{v}_i\| > 0$. Let $W \in \mathbb{R}^{n,n}$ be an orthonormal matrix obtained by the eigenvalue decomposition of $A^\top A$, namely, $A^\top A = WDW^\top$, with D being a diagonal matrix with $D_{1,1} \geq D_{2,2} \geq \cdots \geq 0$. We will show that $\mathbf{v}_1,\ldots,\mathbf{v}_r$ are eigenvectors of $A^\top A$ that correspond to nonzero eigenvalues, and, hence, using Lemma C.4 it follows that these are also right singular vectors of A. The proof is by induction. For the basis of the induction, note that any unit vector $\mathbf{v}$ can be written as $\mathbf{v} = W\mathbf{x}$, for $\mathbf{x} = W^\top\mathbf{v}$, and note that $\|\mathbf{x}\| = 1$. Therefore,

$$\|A\mathbf{v}\|^2 = \|AW\mathbf{x}\|^2 = \|WDW^\top W\mathbf{x}\|^2 = \|WD\mathbf{x}\|^2 = \|D\mathbf{x}\|^2 = \sum_{i=1}^n D_{i,i}^2 {x_i}^2.$$

Therefore,

$$\max_{\mathbf{v}:\|\mathbf{v}\|=1} \|A\mathbf{v}\|^2 = \max_{\mathbf{x}:\|\mathbf{x}\|=1} \sum_{i=1}^n D_{i,i}^2 {x_i}^2.$$

The solution of the right-hand side is to set $\mathbf{x} = (1,0,\ldots,0)$, which implies that $\mathbf{v}_1$ is the first eigenvector of $A^\top A$. Since $\|A\mathbf{v}_1\| > 0$ it follows that $D_{1,1} > 0$ as required. For the induction step, assume that the claim holds for some $1 \leq t \leq r-1$. Then, any $\mathbf{v}$ which is orthogonal to $\mathbf{v}_1,\ldots,\mathbf{v}_t$ can be written as $\mathbf{v} = W\mathbf{x}$ with all the first t elements of $\mathbf{x}$ being zero. It follows that

$$\max_{\mathbf{v}:\|\mathbf{v}\|=1,\forall i\leq t,\mathbf{v}^\top\mathbf{v}_i=0} \|A\mathbf{v}\|^2 = \max_{\mathbf{x}:\|\mathbf{x}\|=1} \sum_{i=t+1}^n D_{i,i}^2 {x_i}^2.$$

The solution of the right-hand side is the all zeros vector except $x_{t+1} = 1$. This implies that $\mathbf{v}_{t+1}$ is the $(t+1)$th column of W. Finally, since $\|A\mathbf{v}_{t+1}\| > 0$ it follows that $D_{t+1,t+1} > 0$ as required. This concludes our proof. □

Corollary C.6 (The SVD Theorem). *Let $A \in \mathbb{R}^{m,n}$ with rank r. Then $A = UDV^\top$ where D is an $r \times r$ matrix with nonzero singular values of A and the columns of U,V are orthonormal left and right singular vectors of A. Furthermore, for all i, $D_{i,i}^2$ is an eigenvalue of $A^\top A$, the ith column of V is the corresponding eigenvector of $A^\top A$ and the ith column of U is the corresponding eigenvector of $AA^\top$.*