Lecturer: Dabeen Lee

1 Outline

In this lecture, we review some mathematical backgrounds that constantly appear throughout the course. We mainly discuss topics in linear algebra. We should be comfortable with every concept mentioned in this lecture.

2 Inner product, outer product, and norms

A d-dimensional vector $v \in \mathbb{R}^d$ is a $d \times 1$ matrix. v is also called a "column" vector, and it can also be written as $(v_1, \ldots, v_d)^\top$ to make its coordinates explicit.

The *inner product* of two vectors $u, v \in \mathbb{R}^d$ is defined as $\sum_{i=1}^d u_i v_i$, and we often denote it by $u \cdot v$, $\langle u, v\rangle$, and $u^\top v$. The *outer product* of u and v is the $d \times d$ matrix given by $uv^\top$ whose entry at row i and column j is $u_i v_j$.

Given a vector $v \in \mathbb{R}^d$, its ℓ_p*-norm* for $p \geq 1$ is defined as $\|v\|_p = (\sum_{i=1}^d |v_i|^p)^{1/p}$. As a special case, the ℓ_2 norm is called the *Euclidean norm*. Note that the ℓ_2 norm $\|v\|_2$ is the square root of the inner product $v \cdot v = v^\top v$. Moreover, the ℓ_∞*-norm* is defined as $\|v\|_\infty = \max\{|v_i| : i = 1, \ldots, d\}$, and the ℓ_0*-norm*, denoted $\|v\|_0$, is defined as the number of nonzero coordinates of v.

In general, a norm is defined as follows.

Definition 1.1. A norm $\|\cdot\|$ on $\mathbb{R}^d$ is a real-valued function with the following properties.

1. Subadditivity/Triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$ for any $u, v \in \mathbb{R}^d$.
2. Absolute homogeneity: $\|\alpha v\| = |\alpha| \cdot \|v\|$ for any $v \in \mathbb{R}^d$ and any scalar $\alpha \in \mathbb{R}$.
3. Positive definiteness: if $\|v\| = 0$, then $v = 0$.

Definition 1.2. Given a norm $\|\cdot\|$ on $\mathbb{R}^d$, the associated *dual norm*, denoted $\|\cdot\|_*$, is defined as

$$\|u\|_* = \sup\left\{u^\top v : \ \|v\| \leq 1\right\}$$

for any $u \in \mathbb{R}^d$.

It can be shown that the dual norm is also a norm and that the dual of the dual norm is the original norm, i.e., $\|x\|_{**} = \|x\|$ for any $x \in \mathbb{R}^d$. For example, the dual of the Euclidean norm is the Euclidean norm itself. The dual of the ℓ_∞ norm is the ℓ_1 norm, while the dual of the ℓ_p norm for $p \geq 1$ is the ℓ_q norm where q satisfies $1/p + 1/q = 1$. The dual of the ℓ_1 norm is the ℓ_∞ norm.

Theorem 1.3. *For any $u, v \in \mathbb{R}^d$, we have $u^\top v \leq \|u\|_* \cdot \|v\|$.*

Proof. Note that

$$u^\top v = u^\top \frac{v}{\|v\|} \cdot \|v\| \leq \|u\|_* \cdot \|v\|$$

where the equality follows from the absolute homogeneity of the norm $\|\cdot\|$ and the inequality follows from the definition of dual norm as $\|v/\|v\|\| = 1$. □

3 Linear independence, subspace, span, dimension, and basis

Let $v^1, \ldots, v^k \in \mathbb{R}^d$ be d-dimensional vectors. A *linear combination* of the vectors is $\sum_{i=1}^k \alpha_i v^i$ for some $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$. We say that vectors $v^1, \ldots, v^k$ are *linearly independent* if there is no way we can write a vector as a linear combination of the others. Equivalently, $v^1, \ldots, v^k$ are linearly independent if the following is satisfied: if $\sum_{i=1}^k \alpha_i v^i = 0$, then we must have $\alpha_1 = \cdots = \alpha_k = 0$. Otherwise, we say that the vectors are *linearly dependent*.

We call $V \subseteq \mathbb{R}^d$ a *(linear) subspace* if V is closed under addition and scaling, i.e., $u + v \in V$ for any $u, v \in V$ and $\alpha \cdot v \in V$ for any $v \in V$ and $\alpha \in \mathbb{R}$. The *span* of vectors $v^1, \ldots, v^k$ is the set of all linear combinations of the vectors, i.e., $\{\sum_{i=1}^k \alpha_i v^i : \alpha_1, \ldots, \alpha_k \in \mathbb{R}\}$. Note that the span is a subspace. In fact, any subspace is the span of some vectors, and here, we say that the subspace is *spanned* by the vectors.

The *dimension* of subspace V is defined as the maximum number of linearly independent vectors in V. When the dimension of subspace V is r, any set of r linearly independent vectors in V is called a *basis*.

4 Projection to a subspace

If two vectors u and v are *orthogonal* (perpendicular), then $u^\top v = 0$. The angle θ between two vectors u and v can be measured by

$$\cos\theta = \frac{u^\top v}{\|u\|_2 \cdot \|v\|_2}.$$

The projection of vector u onto the line spanned by v is

$$\frac{u^\top v}{\|v\|_2^2} \cdot v.$$

Note that the ℓ_2 norm of the projection is precisely $\|u\|_2 \cdot |\cos\theta|$ where θ is the angle between u and v.

Given a subspace in $\mathbb{R}^d$, vectors $v^1, \ldots, v^r$ form an *orthonormal* basis if

- r is the dimension of the subspace (basis),
- $v^1, \ldots, v^r$ are pairwise orthogonal, and
- $\|v^1\|_2 = \cdots = \|v^r\|_2 = 1$ (normalized).

Any subspace in $\mathbb{R}^d$ admits an orthonormal basis, and we can obtain one using the *Gram-Schmidt method*.

Given a subspace V in $\mathbb{R}^d$ with an orthonormal basis $v^1, \ldots, v^r$, the projection of vector u to subspace V can be computed by

$$\sum_{i=1}^r u^\top v^i \cdot v^i.$$

5 Matrix rank, null space, column space, and orthogonal complement

Given an $n \times d$ matrix A, the *rank* of A is defined as the maximum number of linearly independent columns of A[1]. The *null space* of A, denoted $\text{null}(A)$, is defined as $\{x \in \mathbb{R}^d : Ax = 0\}$, and the *column space* of A, denoted $\text{col}(A)$, is defined as $\{y \in \mathbb{R}^n : y = Ax \text{ for some } x \in \mathbb{R}^d\}$. The null space is the collection of vectors that are orthogonal to the rows of A, and the column space is the subspace spanned by the columns of A.

When A is an $n \times n$ square matrix, the following statements are equivalent.

- A is invertible,
- $\det(A) \neq 0$,
- $\text{null}(A) = \{0\}$,
- $\text{col}(A) = \mathbb{R}^n$.

Given a subspace V in $\mathbb{R}^d$, the *orthogonal complement* of V is defined as

$$V^\perp = \{u \in \mathbb{R}^d : \ u^\top v = 0 \text{ for all } v \in V\}.$$

In particular, we have $(V^\perp)^\perp = V$. Note that $\text{null}(A)$, $\text{null}(A^\top)$, $\text{col}(A)$, and $\text{col}(A^\top)$ are all subspaces, and moreover, we have

$$(\text{null}(A))^\perp = \text{col}(A^\top), \quad (\text{col}(A))^\perp = \text{null}(A^\top).$$

6 Symmetric matrices, eigenvalues, eigenvectors, and positive semidefinite matrices

Let M be a $d \times d$ square matrix. Then we say that M is *symmetric* if $M = M^\top$, i.e., $M_{ij} = M_{ji}$ for any $i, j \in [d]$[2]. We say that (λ, v) is an *eigenvalue-eigenvector* pair for M if $Mv = \lambda v$. In fact, any symmetric matrix M satisfies the following properties.

- All the eigenvalues of M are real.
- The eigenvectors corresponding to distinct eigenvalues are orthogonal.

Theorem 1.4. *Let M be a symmetric matrix. Then M can be written as $M = Q\Lambda Q^\top$ where*

1. *Q is an orthonormal matrix, i.e. $Q^\top Q = QQ^\top = I$, whose columns are the eigenvectors of M,*
2. *Λ is a diagonal matrix whose diagonal entries are the eigenvalues of M.*

Here, $Q\Lambda Q^\top$ is called an *eigen decomposition* of M. Therefore, any $d \times d$ symmetric matrix M can be expressed as

$$M = \sum_{i=1}^{d} \lambda_i v^i v^{i\top}$$

where each (λ_i, v^i) is an eigenpair.

[1] Exercise: Check that we can replace columns by rows in the definition.

[2] Here, $[d]$ simply denotes $\{1, \dots, d\}$.

Definition 1.5. We say that a symmetric matrix is *positive semidefinite (PSD)* if all its eigenvalues are nonnegative.

Theorem 1.6. *Let M be a $d \times d$ symmetric matrix. Then M is PSD if and only if $x^\top Mx \geq 0$ for all $x \in \mathbb{R}^d$.*

Proof. Since M is symmetric, we can write M as $M = \sum_{i=1}^d \lambda_i v^i v^{i\top}$. In particular, $v^1, \ldots, v^d$ form an orthonormal basis of $\mathbb{R}^d$. Then for any $x \in \mathbb{R}^d$, there exist scalars $\alpha_1, \ldots, \alpha_d$ so that $x = \sum_{i=1}^d \alpha_i v^i$. Then $x^\top Mx \geq 0$ if and only if

$$\left(\sum_{i=1}^d \alpha_i v^i\right)^\top \left(\sum_{i=1}^d \lambda_i v^i v^{i\top}\right) \left(\sum_{i=1}^d \alpha_i v^i\right) = \sum_{i=1}^d \alpha_i^2 \lambda_i \geq 0.$$

Hence, $x^\top Mx \geq 0$ for all $x \in \mathbb{R}^d$ if and only if λ_i's are all nonnegative, as required. □