1 Outline

In this lecture, we study

- Linear algebra review (symmetric matrices, eigenvalue decomposition, positive semidefinite matices)
- Multivariate calculus review,
- Convex sets

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

Let M be an $d \times d$ square matrix. Then M is *symmetric* if $M = M^\top$, i.e., $M_{ij} = M_{ji}$ for any $i, j \in [d]$[1]. 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 2.1. *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, $QMQ^\top$ 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.

Definition 2.2. We say that a symmetric matrix is *positive semidefinite (PSD)* if all its eigenvalues are nonnegative.

Theorem 2.3. *Let M be an $d \times d$ symmetric matrix. Then M is PSD if and only if $x^\top Mx \geq 0$ for all $x \in \mathbb{R}^d$.*

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

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

3 Lipschitz continuity, gradient, and Hessian

We say that a function $f : \mathbb{R}^d \to \mathbb{R}$ is *Lipschitz continuous* with respect to norm $\|\cdot\|$ if there exists some nonnegative constant $L \geq 0$ such that

$$|f(x) - f(y)| \leq L\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n.$$

Here, we say that f is L-Lipschitz with respect to $\|\cdot\|$.

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function. Let e^i denote the ith unit vector. For example, $e^1 = (1, 0, \ldots, 0)^\top$ and $e^d = (0, \ldots, 0, 1)^\top$. Then the ith *partial derivative* of f is defined as

$$\frac{\partial f}{\partial x_i}(x) = \lim_{t \to 0} \frac{f(x + te^i) - f(x)}{t}.$$

Thus, the ith partial derivative is the directional derivative of f along the ith unit direction e^i. If all the partial derivatives of f exist at $x \in \mathbb{R}^d$, then we may define the *gradient* of f at x, given by

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_d}(x)\right)^\top.$$

If a function is Lipschitz continuous, then it is continuous and differentiable almost everywhere. However, there is a function that is Lipschitz continuous but not differentiable. For example, $f(x) = |x|$ for $x \in \mathbb{R}$.

Next,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \quad \text{for } i, j \in [d]$$

are the second partial derivatives of f. If all the second partial derivatives exist, then we may define the *Hessian* of f as follows

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(x) & \frac{\partial^2 f}{\partial x_d \partial x_2}(x) & \cdots & \frac{\partial^2 f}{\partial x_d^2}(x) \end{pmatrix}.$$

Moreover, if the second partial derivatives are continuous, then Schwarz's theorem implies that the Hessian is symmetric, i.e.,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \quad \text{for every } i, j \in [d].$$

4 A little bit of matrix calculus

Let A be an $n \times d$ matrix and $b \in \mathbb{R}^n$. Let $f(x) = g(Ax - b)$. Then by the chain rule,

$$\nabla f(x) = A^\top \nabla g(Ax - b), \quad \nabla^2 f(x) = A^\top \nabla^2 g(Ax - b) A.$$

Example 2.4. Consider $f(x) = Ax - b$. Then $\nabla f(x) = A^\top$.

Consider a quadratic function $f(x) = x^\top Q x + p^\top x$. Then the gradient of f is given by

$$\nabla f(x) = (Q + Q^\top)x + p, \quad \nabla^2 f(x) = Q + Q^\top.$$

Example 2.5. Consider $f(x) = \|Ax - b\|_2^2/2$. Then

$$\nabla f(x) = A^\top (Ax - b), \quad \nabla^2 f(x) = A^\top A.$$

5 Convex sets

5.1 Definition

Definition 2.6. A set $X \subseteq \mathbb{R}^d$ is *convex* if for any $u, v \in X$ and any $\lambda \in [0, 1]$,

$$\lambda u + (1 - \lambda)v \in X.$$

In words, the line segment joining any two points is entirely contained the set. In Figure 2.1, we have a convex set and a non-convex set.

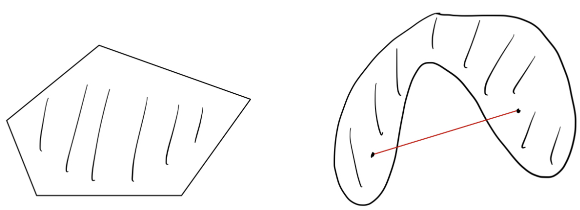

Figure 2.1: A convex set and a nonconvex set

Definition 2.7. Given $v^1, \ldots, v^k \in \mathbb{R}^d$, any linear combination $\lambda_1 v^1 + \cdots + \lambda_k v^k$ is a *convex combination* of $v^1, \ldots, v^k$ if

$$\sum_{i=1}^{k} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0 \text{ for } i = 1, \ldots, k.$$

The convex combination of two distinct points u, v is the line segment $\{\lambda u + (1 - \lambda)v : \ 0 \leq \lambda \leq 1\}$ connecting them.

Definition 2.8. The *convex hull* of a set X, denoted $\mathrm{conv}(X)$, is the set of all convex combinations of points in X. By definition,

$$\mathrm{conv}(X) = \left\{ \sum_{i=1}^{n} \lambda_i v^i : \begin{array}{c} n \in \mathbb{N}, \ v^1, \ldots, v^n \in X, \\ \sum_{i=1}^{n} \lambda_i = 1, \ \lambda_1, \ldots, \lambda_n \geq 0 \end{array} \right\}.$$

$\mathrm{conv}(X)$ is always convex regardless of X. Figure 2.2 shows some examples of taking the convex hull of a (nonconvex) set.

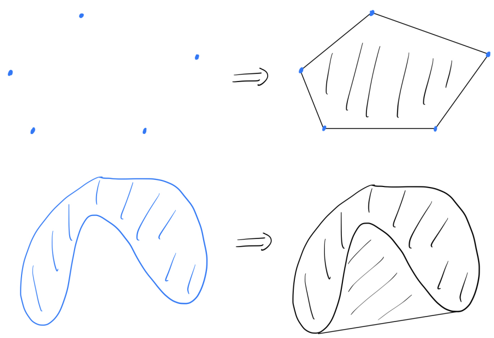

Figure 2.2: A convex set and a nonconvex set

5.2 Cones and affine subspaces

Definition 2.9. A set $C \subseteq \mathbb{R}^d$ is a *cone* if for any $v \in C$ and $\alpha > 0$, we have $\alpha v \in C$. Furthermore, if C is convex, then it is called a *convex cone.*

Note that not all cones are convex. Recall that we discussed linear and convex combinations of vectors.

Definition 2.10. Given $v^1, \dots, v^k \in \mathbb{R}^d$, any point of the form $\alpha_1 v^1 + \cdots + \alpha v_k v^k$ is a *conic combination* of $v^1, \dots, v^k$ if $\alpha_1, \dots, \alpha_k \geq 0$.

In other words, any nonnegative linear combination is a conic combination.

Definition 2.11. The *conic hull* of a set X, denoted $\text{cone}(X)$, is the set of all conic combinations of points in X. By definition,

$$\text{conv}(X) = \left\{ \sum_{i=1}^{n} \lambda_i v^i : \begin{array}{l} n \in \mathbb{N},\ v^1, \dots, v^n \in X, \\ \lambda_1, \dots, \lambda_n \geq 0 \end{array} \right\}.$$

As $\text{conv}(X)$, $\text{cone}(X)$ is always convex, Figure 2.3 shows an example taking the conic hull of a set in $\mathbb{R}^2$.

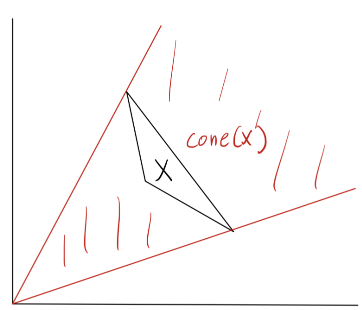

Figure 2.3: Taking the conic hull of a triangle in $\mathbb{R}^2$

Lastly, we define the notion of *affine subspaces.*

Definition 2.12. Given $v^1, \dots, v^k \in \mathbb{R}^d$, any point of the form $\theta_1 v^1 + \cdots + \theta v_k v^k$ is a *affine combination* of $v^1, \dots, v^k$ if $\theta_1 + \cdots + \theta_k = 1$.

In contrast to covex combinations, affine combinations allow negative multipliers.

Definition 2.13. The *affine hull* of a set X is the set of all affine combinations of points in X.

The affine hull of X is also referred to as the *affine subspace* spanned by X. In the previous lecture, we defined the linear subspace spanned by a finite set of vectors, but we can extend the definition to an arbitrary set. The *linear hull* of a set X is equivalent to the linear subspace spanned by X.

In Figure 2.4, we have a set S of two points in $\mathbb{R}^2$. The red line segment is $\mathrm{conv}(S)$, the green line through the two points is the affine subspace spanned by S, the blue cone depicts $\mathrm{cone}(S)$, and lastly, the orange regin (in fact, $\mathbb{R}^2$) is the linear subspace spanned by S.

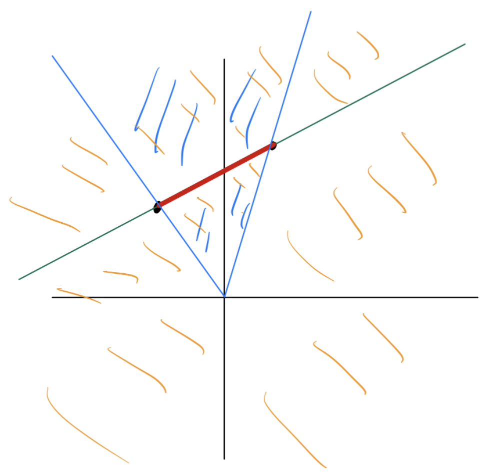

Figure 2.4: Comparing the linear subspace, the affine subspace, the convex hull, and the conic hull

Theorem 2.14. *An affine subspace is a translation of a linear subspace. For an affine subspace $V \subseteq \mathbb{R}^d$, there exist matrices A and b such that $V = \{x \in \mathbb{R}^d : Ax = b\}$.*

5.3 Examples

We saw that the convex hull and conic hull of a set are convex and that the linear subspace and affine subspace spanned by a set are convex. There are many more examples.

1. Empty set, singletons (sets of the form $\{v\}$),
2. Norm ball: $\{x \in \mathbb{R}^d : \|x - c\| \leq r\}$ where c is the center.
3. Ellipsoid: $\{x \in \mathbb{R}^d : (x - c)^\top P(x - c) \leq 1\}$ where P is positive definitie and c is the center.
4. Hyperplane: $\{x \in \mathbb{R}^d : a^\top x = b\}$ where $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$.
5. Half-space: $\{x \in \mathbb{R}^d : a^\top x \leq b\}$ where $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$.
6. Polyhedron: A *polyhedron* is a finite intersection of half spaces, $\{x \in \mathbb{R}^d : Ax \leq b\}$ where $A \in \mathbb{R}^{m \times d}$ and $b \in \mathbb{R}^m$. Here, $Ax \leq b$ is a short-hand notation for system $a_k^\top x \leq b_k$ for $k \in [m]$.
7. Polytope: A *polytope* is a polyhedron that is bounded. Equivalently, a polytope is the convex hull of a finite set of vectors.

8. Simplex: A set of the form $\{x \in \mathbb{R}^d : 1^\top x = 1,\ x \geq 0\}$, which is equal to the convex hull of $e^1, \ldots, e^d$, the d-dimensional unit vectors.
9. Nonnegative orthant: $\mathbb{R}^d_+ = \{x \in \mathbb{R}^d : x \geq 0\}$.
10. Positive orthant: $\mathbb{R}^d_{++} = \{x \in \mathbb{R}^d : x > 0\}$.

There are examples of convex cones, which are convex as well.

1. Norm cone: $\{(x, t) \in \mathbb{R}^d \times \mathbb{R} : \|x\| \leq t\}$. When $\|\cdot\|$ is the Euclidean norm, the cone is called the *second-order cone.*
2. Positive semidefinite cone: The set of all positive semidefinite matrices of a fixed dimension.