21-690: Methods of Optimization

Krish Matta

Contents

1   Introduction

Below are my notes for the course 21-690: Methods of Optimization taught in the Spring 2025 semester by Professor Nicholas Boffi at Carnegie Mellon University.

2   Convex Sets

2.1   Affine and Convex Sets

Definition (Affine). A set $C \subseteq \mathbb{R}^n$ is said to be *affine* if for all $x, y \in C$, we have that $\theta x + (1 - \theta)y \in C$ for all $\theta \in \mathbb{R}$.

Geometrically, affine sets are sets in which the line formed by any two points in the set is entirely contained in the set as well.

Definition (Affine Combination). An *affine combination* of $\{x_i\}_{i=1}^k \subseteq \mathbb{R}^n$ is a linear combination

$$\sum_{i=1}^k \theta_i x_i$$

where $\theta_i \in \mathbb{R}$ for all $i \in [k]$ and $\sum_{i=1}^k \theta_i = 1$.

Proposition. Let $C \subseteq \mathbb{R}^n$ be an affine set. Then, any affine combination of $\{x_i\}_{i=1}^k \subseteq C$ is contained in C.

Proof. By induction on k.

Definition (Affine Hull). The *affine hull* of a set $C \subseteq \mathbb{R}^n$ is the set

$$\mathrm{aff}(C) = \{\theta x + (1 - \theta)y : x, y \in C, \theta \in \mathbb{R}\}.$$

Exercise. Prove that the affine hull of $C \subseteq \mathbb{R}^n$ is the smallest affine set containing C.

Affine sets are certainly unbounded, as lines are unbounded. We can consider bounded sets by considering only line segments rather than entire lines. Thus lies the idea behind convexity.

Definition (Convex). A set $C \subseteq \mathbb{R}^n$ is said to be *convex* if for all $x, y \in C$, we have that $\theta x + (1 - \theta)y \in C$ for all $\theta \in [0, 1]$.

Geometrically, convex sets are sets in which the line segment formed by any two points in the set is entirely contained in the set as well. Note that we only consider line segments by restricting our coefficients to $[0, 1]$.

We can similarly extend the idea of affine combinations and affine hulls to convexity.

Definition (Convex Combination). A *convex combination* of $\{x_i\}_{i=1}^{k} \subseteq \mathbb{R}^n$ is a linear combination

$$\sum_{i=1}^{k} \theta_i x_i$$

where $\theta_i \in [0, 1]$ for all $i \in [k]$ and $\sum_{i=1}^{k} \theta_i = 1$.

Proposition. Let $C \subseteq \mathbb{R}^n$ be a convex set. Then, any convex combination of $\{x_i\}_{i=1}^{k} \subseteq C$ is contained in C.

Proof. By induction on k.

Definition (Convex Hull). The *convex hull* of a set $C \subseteq \mathbb{R}^n$ is the set

$$\text{conv}(C) = \{\theta x + (1 - \theta)y : x, y \in C, \theta \in [0, 1]\}.$$

Exercise. Prove that the convex hull of $C \subseteq \mathbb{R}^n$ is the smallest convex set containing C.

2.2   Examples

2.2.1   Cones

Definition (Cone). A set C is a *cone* if for all $x \in C$, $\lambda x \in C$ for all $\lambda > 0$.

The traditional image of a "cone" is itself a cone, if extended to infinity.

Not all cones are convex: the union of two different lines is a cone, but not convex.

2.2.2   Hyperplanes and Halfspaces

Definition (Hyperplane). Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The set

$$\{x \in \mathbb{R}^n : a^T x = b\}$$

is a *hyperplane*.

Geometrically, hyperplanes are lines in $\mathbb{R}^2$ and planes in $\mathbb{R}^3$.

Proposition. All hyperplanes are affine.

Proof. Consider the hyperplane

$$S = \{x \in \mathbb{R}^n : a^T x = b\}.$$

Consider $x, y \in S$ and $\theta \in \mathbb{R}$. Then,

$$a^T \left(\theta x + (1 - \theta)y\right) = \theta(a^T x) + (1 - \theta)(a^T y) = \theta b + (1 - \theta)b = b.$$

Hence, $\theta x + (1 - \theta)y \in S$ and so S is affine.

Definition (Halfspace). Let $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. The set

$$\{x \in \mathbb{R}^n : a^T x \le b\}$$

is a *halfspace*.

Geometrically, halfspaces are one side of a hyperplane.

Proposition. All halfspaces are convex.

2.2.3   Balls and Ellipsoids

Definition (Ball). A ball is a set

$$B(x_c, r) = \{x \in \mathbb{R}^n : ||x - x_c|| \le r\},$$

where $r > 0$.

Proposition. All balls are convex.

Proof. Consider some ball $B(x_c, r)$. Take $x, y \in B(x_c, r)$ and $\theta \in [0, 1]$. Let $z = \theta x + (1 - \theta)y$. We wish to show that $z \in B(x_c, r)$. Observe that by triangle inequality,

$$||z - x_c|| = ||\theta x + (1 - \theta)y - x_c|| \leq ||\theta x - \theta x_c|| + ||(1 - \theta)y - (1 - \theta x_c)|| \leq \theta r + (1 - \theta)r = r,$$

hence $z \in B(x_c, r)$.

Definition (Ellipsoid). An ellipsoid is a set

$$\{x \in \mathbb{R}^n : (x - x_c)^T P^{-1}(x - x_c) \leq 1\}$$

where $x_c \in \mathbb{R}^n$ and $P \in S_{++}^n$.

Proposition. All ellipsoids are convex.

Proof. Consider some ellipsoid

$$C = \{x \in \mathbb{R}^n : (x - x_c)^T P^{-1}(x - x_c) \leq 1\}.$$

Take $x, y \in C$, and $\theta \in [0, 1]$. Let $z = \theta x + (1 - \theta)y$. We wish to show that $z \in C$. Observe that

$$(z - x_c)^T P^{-1}(z - x_c)$$
$$= (\theta x + (1 - \theta)y - x_c)^T P^{-1}(\theta x + (1 - \theta)y - x_c)$$
$$= (\theta x + (1 - \theta)y - \theta x_c - (1 - \theta)x_c)^T P^{-1}(\theta x + (1 - \theta)y - \theta x_c - (1 - \theta)x_c)$$
$$= (\theta(x - x_c) + (1 - \theta)(y - x_c))^T P^{-1}(\theta(x - x_c) + (1 - \theta)(y - x_c))$$
$$= \theta^2(x - x_c)^T P^{-1}(x - x_c) + \theta(1 - \theta)((x - x_c)^T P^{-1}(y - x_c) + (y - x_c)^T P^{-1}(x - x_c))$$
$$+ (1 - \theta)^2(y - x_c)^T P^{-1}(y - x_c)$$
$$= \theta^2(x - x_c)^T P^{-1}(x - x_c) + 2\theta(1 - \theta)(x - x_c)^T P^{-1}(y - x_c) + (1 - \theta)^2(y - x_c)^T P^{-1}(y - x_c)$$
$$\leq \theta^2(x - x_c)^T P^{-1}(x - x_c) + 2\theta(1 - \theta)\sqrt{(x - x_c)^T P^{-1}(x - x_c)}\sqrt{(y - x_c)^T P^{-1}(y - x_c)}$$
$$+ (1 - \theta)^2(y - x_c)^T P^{-1}(y - x_c)$$
$$= \left(\theta\sqrt{(x - x_c)^T P^{-1}(x - x_c)} + (1 - \theta)\sqrt{(y - x_c)^T P^{-1}(y - x_c)}\right)^2$$
$$\leq (\theta + (1 - \theta))^2$$
$$= 1.$$

Hence, $z \in C$, and so all ellipsoids are convex.

2.2.4   Polyhedra

Definition (Polyhedra). A polyhedra is a set

$$P = \{x \in \mathbb{R}^n : \left(\forall i \in [m], a_i^T x \leq b_i\right) \wedge \left(\forall i \in [p], c_i^T x = d_i\right)\}.$$

Polyhedra are typically presented with the notation

$$P = \{x \in \mathbb{R}^n : Ax \preceq b, Cx = d\}$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, C \in \mathbb{R}^{p \times n}, d \in \mathbb{R}^p$.

Geometrically, polyhedra are the intersection of a finite number of hyperplanes and halfspaces.

2.3   Operations that Preserve Convexity

2.3.1   Intersection

Proposition. Intersection preserves convexity.

Proof. Let I be an index set such that for all i, $C_i \subseteq \mathbb{R}^n$ is a convex set. Let $C = \bigcap_{i \in I} C_i$. We claim that C is convex.

Consider $x, y \in C$, and $\theta \in [0, 1]$. Set $z = \theta x + (1 - \theta)y$. It suffices to show that $z \in C$.

By definition of C, $x, y \in C_i$ for all $i \in I$. By convexity of C_i, we have that $z \in C_i$ for all $i \in I$. Hence, $z \in C$, as desired.

Exercise. The positive semidefinite cone, S_+^n, is convex.

Solution. Observe that

$$S_+^n = \bigcap_{z \in \mathbb{R}^n} \{X \in \mathbb{R}^{n \times n} : z^T X z \geq 0\}.$$

Every set on the right hand side is convex. The intersection of convex sets is convex, hence the positive semidefinite cone is convex.

2.3.2   Affine Functions

Definition (Affine Function). A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *affine* if it is of the form

$$f(x) = Ax + b.$$

Proposition. The image of a convex set over an affine function is convex.

Proof. Let $C \subseteq \mathbb{R}^n$ be a convex set and $f : \mathbb{R}^n \to \mathbb{R}^m$ an affine function via $f(x) = Ax + b$. We wish to show that $f[C]$ is convex.

Consider $x, y \in f[C]$, and $\theta \in [0, 1]$. Define $z = \theta x + (1 - \theta)y$. It suffices to show that $z \in f[C]$.

As $x, y \in f[C]$, there exists $s, t \in C$ such that $f(s) = x$ and $f(t) = y$.

Let $u = \theta s + (1 - \theta)t$. As C is convex, we have that $u \in C$.

Thus,

$$\begin{aligned}
z &= \theta x + (1 - \theta)y \\
&= \theta f(s) + (1 - \theta)f(t) \\
&= \theta As + \theta b + (1 - \theta)At + (1 - \theta)b \\
&= A(\theta s + (1 - \theta)t) + b \\
&= Au + b \\
&= f(u) \in f[C],
\end{aligned}$$

as desired.

The above proposition implies that scaling, translation, and projection preserve convexity.

Proposition. The pre-image of a convex set over an affine function is convex.

Proof. Let $C \subseteq \mathbb{R}^m$ be a convex set and $f : \mathbb{R}^n \to \mathbb{R}^m$ an affine function via $f(x) = Ax + b$. We wish to show that $f^{-1}[C]$ is convex.

Consider $x, y \in f^{-1}[C]$, and $\theta \in [0, 1]$. It suffices to show that $z = \theta x + (1 - \theta)y \in f^{-1}[C]$, which in turn we must show that $f(z) \in C$.

Observe that

$$\begin{aligned}
f(z) &= f\left(\theta x + (1 - \theta)y\right)) \\
&= \theta Ax + \theta b + (1 - \theta)Ay + (1 - \theta)b \\
&= \theta f(x) + (1 - \theta)f(y) \in C
\end{aligned}$$

by convexity of C.

2.3.3 Perspective Function

Definition (Perspective Function). The perspective function $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ is defined via

$$P(s, t) = \frac{s}{t}.$$

Proposition. The image of a convex set over the perspective function is convex.

Proof. Let $C \subseteq \mathbb{R}^n \times \mathbb{R}_{++}$ be a convex set. We wish to show that $P[C]$ is convex.

Consider $x, y \in P[C]$. We wish to show that for any $\theta \in [0, 1]$, $\theta x + (1 - \theta)y \in P[C]$. The proof is difficult if done directly, hence we will take a slightly different approach.

Fix $\theta \in [0, 1]$. Since $x, y \in P[C]$, there exists $(a, s), (b, t) \in C$ such that $P(a, c) = x$ and $P(b, t) = y$.

By convexity of C, we have that

$$(\theta a + (1 - \theta)b, \theta s + (1 - \theta)t) = \theta(a, s) + (1 - \theta)(b, t) \in C.$$

Hence,

$$\frac{\theta s P(x)}{\theta s + (1 - \theta)t} + \frac{(1 - \theta)t P(y)}{\theta s + (1 - \theta)t} = \frac{\theta a + (1 - \theta)b}{\theta s + (1 - \theta)t} = P(\theta a + (1 - \theta)b, \theta s + (1 - \theta)t) \in P[C]$$

Let

$$\mu = \frac{\theta s}{\theta s + (1 - \theta)t}.$$

Through substitution, we have that

$$\mu P(x) + (1 - \mu)P(y) \in P[C].$$

As we vary $\theta \in [0, 1]$, μ varies from 0 to 1, implying that $P[C]$ is convex.

Proposition. The pre-image of a convex set over the perspective function is convex.

Proof. Let $C \subseteq \mathbb{R}^n$ be a convex set. We wish to show that $P^{-1}[C]$ is convex.

Consider $(x, s), (y, t) \in P^{-1}[C]$, and $\theta \in [0, 1]$. It suffices to show that $z = \theta(x, s) + (1 - \theta)(y, t) \in P^{-1}[C]$. Thus, we wish to show that $P(z) \in C$.

Observe that

$$P(z) = P\left(\theta(x, s) + (1 - \theta)(y, t)\right)$$
$$= P\left(\theta x + (1 - \theta)y, \theta s + (1 - \theta)t\right)$$
$$= \frac{\theta x + (1 - \theta)y}{\theta s + (1 - \theta)t}$$
$$= \frac{\theta s}{\theta s + (1 - \theta)t}P(x, s) + \frac{(1 - \theta)t}{\theta s + (1 - \theta)t}P(y, t).$$

Then, note that $P(x, s), P(y, t) \in C$, and

$$\frac{\theta s}{\theta s + (1 - \theta)t}, \frac{(1 - \theta)t}{\theta s + (1 - \theta)t} \in [0, 1].$$

Hence, by convexity of C,

$$P(z) = \frac{\theta s}{\theta s + (1 - \theta)t}P(x, s) + \frac{(1 - \theta)t}{\theta s + (1 - \theta)t}P(y, t) \in C,$$

implying that $z \in P^{-1}[C]$.

2.4 Separating Hyperplanes

Theorem (Separating Hyperplane). Two non-empty and disjoint convex sets can be separated by a hyperplane.

Formally, if there are non-empty and disjoint convex sets $C, D \subseteq \mathbb{R}^n$, there exists $a \in \mathbb{R}^n$, $b \in \mathbb{R}^n$ such that

$$\forall x \in C, a^T x \le b \quad \forall x \in D, a^T x \ge b.$$

Geometrically, there exists a hyperplane between C and D.

Proof. We only prove the theorem for the special case in which C, D are closed and bounded, hence compact.

Note that $C \times D$ is then compact. We may define the distance function $D : C \times D \to \mathbb{R}$ via $D(x, y) = ||x - y||$. By compactness, there exists $u, v \in C, D$ such that $D(u, v) = ||u - v||$ is minimized.

Define $a = v - u$ and $b = (v - u)^T(v + u)/2$. We claim that the hyperplane $\{a^T x = b\}$ separates C, D.

Assume for the sake of contradiction not. Then, without loss of generality, there exists $x \in D$ such that $a^T x < b$. Implying that

$$
\begin{aligned}
& a^T x < b \\
\iff & (v - u)^T x - \frac{(v - u)^T (v + u)}{2} < 0 \\
\iff & (v - u)^T \left(x - \frac{v + u}{2} \right) < 0 \\
\iff & (v - u)^T \left(x + \frac{-v - u}{2} \right) < 0 \\
\iff & (v - u)^T \left(x + \frac{v - u}{2} - v \right) < 0 \\
\iff & (v - u)^T (x - v) + ||v - u||/2 < 0.
\end{aligned}
$$

Clearly, $||v - u|| > 0$, meaning that $(v - u)^T (x - v) < 0$.

Intuitively, we can move v towards the direction $x - v$ and minimize the distance from u.

Formally, we can take the derivative

$$
\frac{d}{dt} ||(v + t(x - v)) - u||^2 \Big|_{t=0} = 2(v - u)^T (x - v) < 0
$$

per the above.

Thus, for some small $t > 0$, we have that

$$
||(v + t(x - v)) - u|| < ||v - u||.
$$

By convexity of D, $v + t(x - v) = (1 - t)v + tx \in D$, hence the above is a contradiction by definition of u, v.

Definition. We say that two convex sets $C, D \subseteq \mathbb{R}^n$ are *strictly separated* if there exists $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that

$$
\forall x \in C, a^T x < b \quad \forall x \in D, a^T x > b.
$$

Example. Let $C \subseteq \mathbb{R}^n$ be a closed convex set and $x_0 \in \mathbb{R}^n$ a point not in C. Then, C and $\{x_0\}$ are strictly separated.

To see why, note that as C is closed, $\mathbb{R}^n \setminus C$ is open. Hence, we can find $r > 0$ such that $B(x_0, r) \cap C = \emptyset$. Clearly, $B(x_0, r)$ is convex. By the separating hyperplane theorem, we may find $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that $a^T x \leq b$ for all $x \in C$, and $a^t x \geq b$ for all $x \in B(x_0, r)$. In particular, the last statement means that for any $u \in \mathbb{R}^n$ where $||u|| \leq r$,

$$a^T x_0 + a^T u = a^T (x_0 + u) \geq b.$$

The left hand side is minimized when $u = -\frac{a}{r||a||}$, hence

$$a^T x_0 - r \geq b \implies a^T x_0 \geq b + r > b.$$

Thus, the hyperplane strictly separates C and $\{x_0\}$.

We can use this result to show the following.

Proposition. Let $C \subseteq \mathbb{R}^n$ be a convex set, and $\mathcal{H}$ be the set of all halfspaces that contain C entirely. Then,

$$C = \bigcap \mathcal{H}.$$

Proof.

$C \subseteq \bigcap \mathcal{H}$

Trivial by definition of $\mathcal{H}$.

$\bigcap \mathcal{H} \subseteq C$

Take $x \in \bigcap \mathcal{H}$. Assume for the sake of contradiction that $x \notin C$. Then, we may find a strictly separating hyperplane between $\{x\}$ and C. Implying that $x \notin \bigcap \mathcal{H}$, a contradiction.

2.5   Supporting Hyperplanes

Definition. For a convex set $C \subseteq \mathbb{R}^n$, we say that a hyperplane

$$\{x \in \mathbb{R}^n : a^T x = a^T x\}$$

is a *supporting hyperplane* if $x_0 \in \partial C$ and $a^T x \leq a^T x_0$ for all $x \in C$. Geometrically, the hyperplane is tangent to a point on the boundary of C, and its halfspace contains the entirety of C.

3 Convex Functions

3.1 Basic Properties and Definitions

Definition. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if for all $x, y \in f$, $\theta \in [0, 1]$, we have that

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y).$$

If the inequality is strict, i.e.

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y),$$

for $x \neq y$, then the function is said to be *strictly convex*.

Intuitively, convex functions are those in which the epigraph of the function (the area above the function) is a convex set.

Remark. A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if it is convex when restricted to any line in its domain, i.e. for all $v \in \mathbb{R}^n$,

$$g(t) = f(x + tv)$$

is convex.

Theorem (First Order Characterization). A function $f : \mathbb{R}^n \to \mathbb{R}$ in C^1 is convex if and only if

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

for all $x, y \in \mathbb{R}^n$.

Proof.

Case: $n = 1$

First assume that f is convex. Then for any $x, y \in \mathbb{R}$, we have that

$$f(x + \theta(y - x)) \leq (1 - \theta)f(x) + \theta f(y)$$
$$\implies f(x + \theta(y - x)) \leq f(x) - \theta f(x) + \theta f(y)$$
$$\implies f(x) + \frac{f(x + \theta(y - x)) - f(x)}{\theta} \leq f(y)$$
$$\implies \lim_{\theta \to 0} f(x) + \frac{f(x + \theta(y - x)) - f(x)}{\theta} \leq f(y)$$
$$\implies f(x) + \lim_{\theta \to 0} \frac{f(x + \theta(y - x)) - f(x)}{\theta} \leq f(y)$$
$$\implies f(x) + (y - x) \lim_{\theta \to 0} \frac{f(x + \theta(y - x)) - f(x)}{\theta(y - x)} \leq f(y)$$
$$\implies f(x) + f'(x)(y - x) \leq f(y).$$

as desired.

Now instead assume that for all $x, y \in \mathbb{R}$,

$$f(x) + f'(x)(y - x) \leq f(y).$$

We wish to show that f is convex. Fix $x, y \in \mathbb{R}$ and $\theta \in [0, 1]$. Let $z = \theta x + (1 - \theta)y$. Then,

$$f(z) + f'(z)(x - z) \leq f(x)$$
$$\implies f(z) + (1 - \theta)f'(z)(x - y) \leq f(x).$$

Similarly, we can see that

$$f(z) + \theta f'(z)(y - x) \leq f(y).$$

Combining these,

$$f(z) = \theta f(z) + (1 - \theta)f(z) \leq \theta f(x) + (1 - \theta)f(y),$$

as desired.

Case: $n > 1$

We can study the one-dimensional function which varies x in the direction of $y - x$, i.e. $g(t) = f(x + t(y - x))$. The result then follows by the $n = 1$ case.

Remark. The above inequality is strict if and only if the function is strictly convex.

Corollary. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and $x \in \mathbb{R}^n$ a point such that $\nabla f(x) = 0$. Then, x is a global minimizer of f.

Proof. Consider some $y \in \mathbb{R}^n$. Per the first order characterization of convex functions,

$$f(x) = f(x) + \nabla f(x)^T (y - x) \leq f(y).$$

Theorem (Second Order Characterization). A function $f : \mathbf{dom}(f) \to \mathbb{R}$ in C^2 is convex if and only if

$$\nabla^2 f(x) \succeq 0$$

for all $x \in \mathbf{dom}(f)$.

Proof.

First assume that f is convex. Assume for the sake of contradiction that there exists some $x \in \mathbf{dom}(f)$ such that $\nabla^2 f(x)$ is not positive semi-definite. By definition, there exists some eigenvector $v \in \mathbb{R}^n$ such that $\nabla^2 f(x)v = \lambda v$ where $\lambda < 0$.

Define $g(t) = f(x + tv)$. Note that

$$g''(t) = v^T \nabla^2 f(x + tv)v.$$

Hence,

$$g''(0) = v^T \nabla^2 f(x)v = v^T \lambda v = \lambda ||v||^2 < 0.$$

By definition of the second derivative, for some small $\epsilon > 0$, $g'(c) < g'(0)$ for all $c \in (0, \epsilon)$.

Then,

15

$$g(\epsilon) = g(0) + (g(\epsilon) - g(0)) = g(0) + g'(c)\epsilon < g(0) + g'(0)\epsilon,$$

by the mean value theorem. Hence, a contradiction since g inherits convexity from f and the above violates the first order characterization of convexity. Thus, $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}(f)$.

Now assume that $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}(f)$. We claim that f is convex.

Fix $x, y \in \mathbf{dom}(f)$. Define

$$g(t) = f(x + t(y - x)).$$

Then,

$$f(y) = g(1) = g(0) + (g(1) - g(0)) = g(0) + g'(c) \geq g(0) + g'(0) = f(x) + \nabla f(x)^T (y - x)$$

by mean value theorem. The inequality follows from the fact that g'' is always non-negative, implying that g' is non-decreasing.

Since x, y are arbitrary, we have that f is convex by the first order characterization of convexity.

Remark. The above inequality is strict if and only if the function is strictly convex.

3.2 Operations that Preserve Convexity

3.2.1 Nonnegative Weighted Sum

Proposition. Let $\{f_i\}_{i=1}^n$ be a sequence of convex functions. Then,

$$\sum_{i=1}^n \omega_i f_i, \quad \omega_i \geq 0$$

is convex.

Proof. Fix $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$. Then,

$$\left(\sum_{i=1}^{n} \omega_i f_i\right)(\theta x + (1 - \theta y)) = \sum_{i=1}^{n} \omega_i f_i(\theta x + (1 - \theta)y)$$

$$\leq \sum_{i=1}^{n} \omega_i\left(\theta f_i(x) + (1 - \theta)f_i(y)\right)$$

$$= \theta\left(\sum_{i=1}^{n} \omega_i f_i(x)\right) + (1 - \theta)\left(\sum_{i=1}^{n} \omega_i f_i(y)\right)$$

$$= \theta\left(\sum_{i=1}^{n} \omega_i f_i\right)(x) + (1 - \theta)\left(\sum_{i=1}^{n} \omega_i f_i\right)(y)$$

as desired.

Remark. The above proposition generalizes to infinite sums (if they converge) as well as integrals. Specifically, if a function $f(x, y)$ is convex in x for all $y \in A$, then

$$g(x) = \int_{A} \omega(y)f(x, y)dy, \quad \omega(y) \geq 0$$

is convex.

3.2.2 Affine Composition

Proposition. Let $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{n \times m}$, and $b \in \mathbb{R}^n$. Then,

$$g(x) = f(Ax + b)$$

is convex if f is convex.

Proof. Fix $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$. Then,

$$\begin{aligned}
g(\theta x + (1 - \theta)y) &= f(A(\theta x + (1 - \theta)y) + b) \\
&= f(\theta(Ax + b) + (1 - \theta)(Ay + b)) \\
&\leq \theta f(Ax + b) + (1 - \theta)f(Ay + b) \\
&= \theta g(x) + (1 - \theta)g(y).
\end{aligned}$$

3.2.3 Maximum and Supremum

Proposition. Let $\{f_i\}_{i=1}^n$ be a sequence of convex functions. Then,

$$g(x) = \max_i f_i(x)$$

is convex.

Proof. Fix $x, y \in \mathbb{R}^n$ and $\theta \in [0, 1]$. Then,

$$
\begin{aligned}
g\left(\theta x + (1-\theta)y\right) &= \max_i f_i\left(\theta x + (1-\theta)y\right) \\
&\leq \max_i \left(\theta f_i(x) + (1-\theta)f_i(y)\right) \\
&\leq \theta \max_i f_i(x) + (1-\theta)\max_i f_i(y) \\
&= \theta g(x) + (1-\theta)g(y).
\end{aligned}
$$

Remark. The above proposition generalizes to the supremum. Specifically, if $f(x, y)$ is convex in x for all y, then

$$g(x) = \sup_y f(x, y)$$

is convex.

3.2.4 Representation as Supremum of Affine Functions

Proposition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then,

$$f(x) = \sup\{g(x) : g \text{ affine}, g(y) \leq f(y) \quad \forall y \in \mathbb{R}^n\}.$$

Proof.

$f(x) \leq \sup\{g(x) : g \text{ affine}, g(y) \leq f(y) \quad \forall y \in \mathbb{R}^n\}$

Define

$$\mathbf{epi}(f) = \{(x, y) : f(x) \leq y\}.$$

We claim that $\mathbf{epi}(f)$ is convex. Fix $(x_1, y_1), (x_2, y_2) \in \mathbf{epi}(f)$, and $\theta \in [0, 1]$. Then let $z = \theta(x_1, y_1) + (1-\theta)(x_2, y_2)$. So,

$$z = (\theta x_1 + (1 - \theta)x_2, \theta y_1 + (1 - \theta)y_2).$$

By convexity of f,

$$f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2) \leq \theta y_1 + (1 - \theta)y_2$$

implying that $z \in \mathbf{epi}(f)$, as desired. Hence, $\mathbf{epi}(f)$ is convex.

Now fix some $x \in \mathbb{R}^n$. We shall show that indeed,

$$f(x) \leq \sup\{g(x) : g \text{ affine}, g(y) \leq f(y) \quad \forall y \in \mathbb{R}^n\}.$$

Observe that $(x, f(x)) \in \partial \mathbf{epi}(f(x))$, hence we may find $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ such that

$$\begin{bmatrix} a \\ b \end{bmatrix} \begin{bmatrix} x - z \\ f(x) - t \end{bmatrix} \leq 0$$

for all $(z, t) \in \mathbf{epi}(f)$.

Then,

$$a^T(x - z) + b(f(x) - z) \leq 0$$
$$\implies a^T(x - z) + b(f(x) - f(z) - s) \leq 0$$

for all s. Implying that $b > 0$.

We can then write the above as

$$g(z) = \frac{a^T}{b}(x - z) + f(x) \leq f(z)$$

for all $z \in \mathbb{R}^n$.

Hence, g is an affine function that underestimates f over all z, and achieves $g(x) = f(x)$. We thus have the result.

$$f(x) \geq \sup\{g(x) : g \text{ affine}, g(y) \leq f(y) \quad \forall y \in \mathbb{R}^n\}$$

Follows by definition.

3.2.5   Perspective of a Function

Definition. Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. We define the perspective of f as $g_f : \mathbb{R}^n \times \mathbb{R}_{>0} \to \mathbb{R}$ via

$$g(x, t) = tf(x/t).$$

Proposition. If $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then g_f is convex.

Proof. It suffices to show that the epigraph of g_f is convex. Note that the epigraph of g_f is the preimage of the epigraph of f over the perspective function, hence is convex.

3.3   Convex Conjugates

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$. The *conjugate* of f, $f^\star : \mathbb{R}^n \to \mathbb{R}$ is defined via

$$f^\star(y) = \sup_{x \in \mathbf{dom}(f)} \{y^T x - f(x)\}.$$

Geometrically, the conjugate of a function f is the greatest distance between f and the hyperplane $y^T x$.

Proposition. For any function $f : \mathbb{R}^n \to \mathbb{R}$, the conjugate $f^\star$ is always convex.

Proof. Observe that $y^T x - f(x)$ is convex in y, hence $f^\star(y) = \sup_{x \in \mathbf{dom}(f)} \{y^T x - f(x)\}$ is convex in y.

3.3.1   Basic Properties

Proposition (Fenchel's Inequality). Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Then,

$$f(x) + f^\star(y) \geq x^T y$$

for all x, y.

Proof. Note that

$$f^\star(y) \geq x^T y - f(x)$$

for all x, y. The result is then immediate.

4 Convex Optimization Problems

4.1 Optimization Problems

Definition. An *optimization problem* is of the form

$$
\begin{aligned}
\min_{x} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p.
\end{aligned}
$$

Its domain is the intersection of domains for each function, i.e.

$$
D = \bigcap_{i=0}^{m} \mathbf{dom}(f_i) \cap \bigcap_{i=1}^{p} \mathbf{dom}(h_i).
$$

Definition. The *feasible set* of an optimization problem is the set

$$
\Omega = \{x \in D : f_i(x) \leq 0, h_i(x) = 0\}.
$$

We say that an optimization problem is feasible if its feasible set is not empty.

Definition. We define the optimal value of an optimization problem as the value

$$
p^{\star} = \inf_{x \in \Omega} f_0(x).
$$

If $\Omega = \emptyset$, then $p^{\star} = \infty$.

Definition. If there exists a sequence $\{x_i\}_{i=1}^{\infty} \subseteq \Omega$ such that $f_0(x_k) \to -\infty$ as $k \to \infty$, we say that the optimization problem is *unbounded below* and $p^{\star} = -\infty$.

Definition. We say that $x \in \Omega$ is ϵ - *suboptimal* if $f_0(x) \leq p^{\star} + \epsilon$.

Definition. We say that $x \in \Omega$ is *locally optimal* if there exists some $R > 0$ such that $f_0(z) \geq f_0(x)$ for all $z \in B(x, R)$.

Definition. If $f_i(x) = 0$ for some $i \in [m]$ and $x \in \Omega$, we say that constraint i is *active* at x.

Definition. We call an optimization problem of the form

$$
\begin{aligned}
\min_{x} \quad & 0 \\
\text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p.
\end{aligned}
$$

a *feasibility problem.*

Remark. Note that maximization problem can be formulated as optimization problems by taking $-f_0$.

4.1.1 Slack Variables

Slack variables allow us to express inequalities as equalities.

In particular, note that

$$
f_i(x) \leq 0 \iff f_i(x) + \xi = 0
$$

for some $\xi \geq 0$ (in particular, $\xi = -f_i(x)$). Here, ξ is a *slack variable.*

More generally, we can reformulate the optimization problem

$$
\begin{aligned}
\min_{x} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p.
\end{aligned}
$$

as

$$
\begin{aligned}
\min_{x, \xi} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) + \xi_i = 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p \\
& \xi_i \geq 0, \quad i = 1, \ldots, m.
\end{aligned}
$$

4.2 Convex Problems

Definition. A *convex problem* is an optimization problem of the form

$$
\begin{aligned}
\min_{x} \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& a_i^T x = b, \quad i = 1, \ldots, p.
\end{aligned}
$$

where f_i is convex for all i.

Remark. The feasible set of a convex optimization problem is convex:

$$
\Omega = \bigcap_{i=1}^{m} \{x : f_i(x) \leq 0\} \cap \mathbf{dom}(f_0) \cap \{x : Ax = b\},
$$

i.e. it is the intersection of convex sets.

Proposition. Any local solution to a convex optimization problem is a global solution.

Proof. Let $x \in \Omega$ be a local solution to the typical convex optimization problem. Then, there exists $R > 0$ such that x is optimal in the R ball around it.

Assume for the sake of contradiction that x is not locally optimal. Then, there exists some $y \in \Omega$ such that $f_0(y) < f_0(x)$.

We can find some $\theta \in [0, 1]$ such that $x + \theta(y - x) \in B(x, R)$. Then,

$$
f_0\left(x + \theta(y - x)\right) \leq (1 - \theta)f_0(x) + \theta f_0(y) < f_0(x),
$$

a contradiction.

The above proposition provides some intuition as to why convex optimization problems are particularly nice to work with.

Proposition. Let $f_0 \in C^1(\Omega)$ be a convex function. Then, $x \in \Omega$ is optimal if and only if

$$
\nabla f_0(x)^T (y - x) \geq 0
$$

for all $y \in \Omega$.

Proof. First assume that $x \in \Omega$ is optimal. Assume for the sake of contradiction that $\nabla f_0(x)^T (y - x) < 0$ for some y. Define

$$g(t) = f(x + t(y - x)).$$

Observe then that

$$g'(0) = \nabla f_0(x)^T (y - x) < 0.$$

Thus, for some $\epsilon > 0$, for all $c \in (0, \epsilon)$, we have that $g(c) < g(0)$, implying that

$$f(x + c(y - x)) < f(x),$$

a contradiction.

Now assume that

$$\nabla f_0(x)^T (y - x) \geq 0$$

holds for all x. We claim that x is optimal. Observe that

$$f(x) \leq f(x) + \nabla f(x)^T (y - x) \leq f(y)$$

as desired.

Corollary. Consider some convex optimization problem where $f_0 \in C^1(\Omega)$, $\Omega = \mathbb{R}^n$ or $\Omega = \mathbf{dom}(f_0)$, and there is an optimal point x. Then,

$$\nabla f(x)^T (y - x) = 0.$$

Proof.

Case: $\Omega = \mathbb{R}^n$

For any optimal point x, we must have that

$$\nabla f(x)^T (y - x) \geq 0.$$

If the gradient were not zero, we can find y such that $\nabla f(x)^T(y - x) > 0$, then see that for vector $-y$, $\nabla f(x)^T(y - x) < 0$, a contradiction. Hence,

$$\nabla f(x) = 0.$$

Case: $\Omega = \mathbf{dom}(f_0)$

As f_0 is continuously differentiable, its domain must be open. Hence, we may find $\theta > 0$ small enough such that $y = x - \theta\nabla f_0(x) \in \Omega$.

Then,

$$-\theta||\nabla f_0(x)^T|| = \nabla f_0(x)^T(-\theta\nabla f_0(x)) = \nabla f_0(x)^T(y - x) \geq 0$$

which is only true if the gradient were zero.

Remark. Consider some convex optimization problem in which $f_0 \in C^1(\Omega)$ and the only constraint is $Ax = b$.

Then, the optimality condition for a point x states that

$$\nabla f_0(x)^T(y - x) \geq 0$$

for all $y \in \Omega$, i.e. for all y such that $Ay = b$. In particular, we can express $y = x + v$ where $Av = 0$. Then by substitution,

$$\nabla f_0(x)^T v \geq 0$$

for all $v \in \mathcal{N}(A)$.

Moreover, $\mathcal{N}(A)$ is a subspace, and $\nabla f_0(x)^T v$ is a non-negative linear function over $v \in \mathcal{N}(A)$. Hence,

$$\nabla f_0(x)^T v = 0.$$