2.1 Review from last time

A convex optimization problem is of the form

$$\min_{x \in D} f(x) \quad \text{subject to} \quad g_i(x) \leq 0, \ i = 1, \ldots, m \quad h_j(x) = 0, \ j = 1, \ldots, r$$

where the criterion $f(x)$ is convex, the inequality constraint functions $g_i(x)$ are convex, and the equality constraint functions $h_j(x)$ are affine. It has a nice property that any local minimizer is a global minimizer.

Nonconvex problems are mostly treated on a case by case basis.

2.2 Convex sets

2.2.1 Definitions of convex sets

A convex set is defined as $C \in \mathbb{R}^n$ such that $x, y \in C \implies tx + (1-t)y \in C$ for all $0 \leq t \leq 1$. In other words, a line segment joining any two elements lies entirely in the set.

Figure 2.1: Convex set v.s. nonconvex set
A convex combination of $x_1, \ldots, x_k \in \mathbb{R}^n$ is any linear combination:

$$\sum_{i=1}^{k} \theta_i x_i = \theta_1 x_1 + \ldots + \theta_k x_k$$

with $\theta_i \geq 0, i = 1, \ldots, k$ and $\sum_{i=1}^{k} \theta_i = 1$

A convex hull of a set $\mathcal{C}$ is the set of all convex combination of its elements, which is always convex. Any convex combination of points in $\text{conv}(\mathcal{C})$ is also

$$\text{conv}(\mathcal{C}) = \{ \sum_{i=1}^{k} \theta_i x_i : k \geq 1, x_i \in \mathcal{C}, \theta_i \geq 0, \sum_{i=1}^{k} \theta_i = 1 \}$$

2.2.2 Examples of convex sets

- Norm ball: $\{ x : \|x\| \leq r \}$, for given norm $\|\cdot\|$, radius $r$
- Hyperplane: $\{ x : a^T x = b \}$, for given $a, b$
- Halfspace: $\{ x : a^T x \leq b \}$, for given $a, b$
- Affine Space: $\{ x : Ax = b \}$, for given $A, b$
- Polyhedron: $\{ x : Ax \leq b \}$, for given $A, b$. You can visualize every row of $A$ as a normal vector for each hyperplane involved. Also, $\{ x : Ax \leq b, Cx = d \}$ is also a polyhedron because the equality $Cx = d$ can be made into two inequalities $Cx \leq d$ and $Cx \leq d$.

![Figure 2.2: Polyhedron with rows of $A$ equal to $a_1, \ldots, a_k$](image)

- Simplex: is a special case of polyhedra, given by the convex hull of a set of affinely independent points $x_0, \ldots, x_k$ (i.e. $\text{conv}\{x_0, \ldots, x_k\}$). Affinely independent means that $x_1 - x_0, \ldots, x_k - x_0$ are linearly independent. A canonical example is the probability simplex

$$\text{conv}\{e_1, \ldots, e_n\} = \{ \omega : \omega \leq 0, 1^T \omega = 1 \}$$
2.2.3 Definitions of convex cones

- A cone is $C \in \mathbb{R}^n$ such that
  \[ x \in C \implies tx \in C \text{ for all } t \geq 0 \]

- A convex cone is a cone that is also convex:
  \[ x_1, x_2 \in C \implies t_1 x_1 + t_2 x_2 \in C \text{ for all } t_1, t_2 \geq 0 \]

![Convex cone](image)

Figure 2.3: Convex cone

- A conic combination of points $x_1, \ldots, x_k \in \mathbb{R}^n$ is, for any $\theta_i \geq 0, i = 1, \ldots, k$, any linear combination
  \[ \theta_1 x_1 + \ldots + \theta_k x_k \]

- A conic hull collects all conic combinations of $x_1, \ldots, x_k$ (or a general set $C$)
  \[ \text{conic}(\{x_1, \ldots, x_k\}) = \{\theta_1 x_1 + \ldots + \theta_k x_k, \theta_i \geq 0, i = 1, \ldots, k\} \]

2.2.4 Examples of convex cones

- **Norm cone**: A norm cone is $\{(x, t) : \|x\| \leq t\}$ Under the $\ell_2$ norm, this is called a second-order cone.

- **Normal cone** Given set $C$ and point $x \in C$, a normal cone is
  \[ N_C(x) = \{g : g^T x \geq g^T y, \text{ for all } y \in C\} \]
  In other words, it’s the set of all vectors whose inner product is maximized at $x$. So the normal cone is always a convex set regardless of what $C$ is.

![Normal cone](image)

Figure 2.4: Normal cone

- **PSD cone** A positive semidefinite cone is the set of positive definite symmetric matrices. ($\mathbb{S}^n$ is the set of $n \times n$ symmetric matrices)
  \[ \mathbb{S}_+^n = \{X \in \mathbb{S}^n : X \succeq 0\} \]
2.2.5 Properties of convex sets

- **Separating hyperplane theorem**: two disjoint convex sets have a separating hyperplane between them as shown in figure:2.5

![Figure 2.5: Illustration of separating hyperplane](image)

Formally, if \( C \), \( D \) are nonempty disjoint convex sets, then there exists \( a, b \) such that

\[
C \in \{ x : a^T x \leq b \} \\
D \in \{ x : a^T x \geq b \}
\]

- **Supporting hyperplane theorem**: any boundary point of a convex set has a supporting hyperplane passing through it. Formally, given an nonempty convex set \( C \), for every point \( x_0 \in bd(C) \), there exists \( a \) such that

\[
C \in \{ x : a^T x \leq a^T x_0 \}
\]

2.2.6 Operations preserving convexity

- **Intersection**: The intersection of convex sets is also a convex sets.

- **Scaling and Translation**: If \( C \) is a convex set, then the following is convex for any \( a, b \).

\[
a C + b = \{ a x + b : x \in C \}
\]

- **Affine images and preimages**: If \( f(x) = A x + b \) and \( C \) is convex then

\[
f(C) = \{ f(X) : x \in C \}
\]

is convex, and if \( D \) is convex then

\[
f^{-1}(D) = \{ x : f(x) \in D \}
\]

is convex.

- **Perspective images and preimages**: For Function \( P : \mathbb{R}^n \times \mathbb{R}_{++} \rightarrow \mathbb{R}^n \) (where \( \mathbb{R}_{++} \) denotes positive reals),

$P(x, z) = x/z$

for $z > 0$ is a perspective function. If $C \subseteq \text{dom}(P)$ is convex, then so is $P(C)$, and if $D$ is convex then so is $P^{-1}(D)$.

- **Linear-fractal images and preimages**: A linear fractal function is a perspective map composed with an affine function, defined on $C^T x + d > 0$:

  $$f(x) = \frac{Ax + b}{c^T x + d}$$

The image and preimage of $f(x)$ are both convex.

### 2.2.7 Example of Operations on Convex Sets

#### 2.2.7.1 Linear matrix inequality solution set

Given $A_1, ..., A_k, B \in \mathbb{S}^n$, a **linear matrix inequality** is of the form

$$x_1A_1 + x_2A_2 + ... + x_kA_k \preceq B$$

for a variable $x \in \mathbb{R}^k$. The set $C$ of points $x$ that satisfy the above inequality is convex.

There are 2 approaches to prove that $C$ is convex.

**Approach 1**: We could directly verify that for $x, y \in C \Rightarrow tx + (1 - t)y \in C$. This follows by checking that, for any $v$, we have,

$$v^T (B - \sum_{i=1}^k (tx_i + (1 - t)y_i)A_i)v \geq 0$$

$$tv^T (B - \sum_{i=1}^k x_i)v + (1 - t)v^T (B - \sum_{i=1}^k y_i)v \geq 0.$$  

The above is true because $x, y \in C$.

**Approach 2**: Let $f : \mathbb{R}^k \rightarrow \mathbb{S}^n, f(x) = B - \sum_{i=1}^k x_iA_i$ and note that this is the affine preimage of a convex set, $C = f^{-1}(\mathbb{S}^n_+)$

#### 2.2.7.2 Fantope

Given some integer $k \geq 0$, the **fantope** of order $k$ is $F = \{Z \in \mathbb{S}^n : 0 \preceq Z \preceq I, \text{tr}(Z) = k\}$. We could prove that $F$ is convex in 2 ways.

**Approach 1**: We could prove that $F$ is convex by taking two matrices $0 \preceq Z, W \preceq I$ and $\text{tr}(Z) = \text{tr}(W) = k$ which implies the same for $tZ + (1 - t)W$.

**Approach 2**: We recognize the fact that the fantope is:

$$F = \{Z \in \mathbb{S}^n : Z \succeq 0\} \cap \{Z \in \mathbb{S}^n : Z \preceq I\} \cap \{Z \in \mathbb{S}^n : \text{tr}(Z) = k\}$$

which is an intersection of linear inequality and equality constraints like a polyhedron but for matrices.
2.2.7.3 Conditional probability set

Let $U, V$ be random variables over $\{1, \ldots, n\}$ and $\{1, \ldots, m\}$. Let $C \subseteq \mathbb{R}^{nm}$ be a set of joint distributions for $U, V$, i.e., each $p \in C$ defines joint probabilities

$$p_{ij} = \mathbb{P}(U = i, V = j)$$

Let $D \subseteq \mathbb{R}^{nm}$ contain corresponding conditional distributions, i.e., each $q \in D$ defines

$$q_{ij} = \mathbb{P}(U = i | V = j)$$

Assume $C$ is convex, let’s prove that $D$ is convex. The set $D$ can be rewritten as an image of a linear fractional function:

$$D = \left\{ q \in \mathbb{R}^{nm} : q_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n}} \text{ for some } p \in C \right\} = f(C)$$

Hence $D$ is convex.

2.3 Convex Functions

2.3.1 Definitions

A **convex function** is a function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\text{dom}(f) \subseteq \mathbb{R}^n$ is convex, and

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

The value of the function lies below the line segment joining $f(x), f(y)$.

![Figure 4: Graph of a convex function](image)

A **concave function** has a similar function definition as a convex function but with an opposite inequality.

$$f \text{ concave } \iff -f \text{ convex}$$

**Some important modifiers:**

**Strictly convex:** A function $f$ is strictly convex if $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$. In words, $f$ is convex and has greater curvature than a linear function.
Strongly convex: A function $f$ is strongly convex with parameter $m > 0$ if $f - \frac{m}{2} \|X\|_2^2$ is convex. In words, $f$ is at least as convex as a quadratic function.

From the above definition,

Strong Convexity $\Rightarrow$ Strict Convexity $\Rightarrow$ convexity.

It is similarly defined for concave function.

2.3.2 Examples of convex functions

- **Univariate functions:**
  - **Exponential function:** The exponential function $e^{ax}$ is convex for any $a$.
  - **Power function:** The power function $x^a$ is convex for $a \geq 1$ or $a \leq 0$ and is concave for $0 \leq a \leq 1$.
  - **Logarithmic function:** The logarithmic function $\log(x)$ is always concave.

- **Affine function:** The affine function $a^T x + b$ is both convex and concave.

- **Quadratic function:** The quadratic function $\frac{1}{2} x^T Q x + b^T x + c$ is convex provided that $Q \succeq 0$ (positive semidefinite).

- **Least square loss:** $\|y - Ax\|^2$ is always convex since $A^T A$ is always a PSD matrix.

- **Norm:** $\|X\|$ is convex for any norm; e.g., $l_p$ norms.

  \[
  \|X\|_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \text{ for } p \geq 1, \quad \|X\|_\infty = \max_{i=1,..,n} |x_i|
  \]

  where $\sigma_1(X) \geq \sigma_2(X) \geq \sigma_r(X) \geq 0$ are the singular value of the matrix $X$.

- **Indicator function:** If $C$ is convex, then its indicator function is also convex. Its indicator function is given by

  \[
  I_C(x) = \begin{cases} 
  0, & x \in C \\
  \infty, & x \notin C
  \end{cases}
  \]

- **Support function:** For any set $C$ (convex or not), its support function defined by $I_C^*(y)$ is convex.

  \[
  I_C^*(y) = \max_{y \in C} x^T y
  \]

- **Max function:** The maximum function, $f(x) = \max \{x_1, x_2, x_3, ..., x_n\}$ is a convex function.
2.4 Key properties of convex functions

• A function is convex if and only if its restriction to any line is convex. For example, let \( f : \mathbb{R}^n \to \mathbb{R} \) be a function and \( x_0, a \in \mathbb{R}^n \) be a point in the domain of \( f \). Let \( g(t) = f(x_0 + ta) \). Then \( f \) is convex if and only if \( g \) is convex for every choice of \( x_0 \) and \( a \). This property is useful for proving the convexity of certain functions.

• **Epigraph characterization:** A function \( f \) is convex if and only if its epigraph is a convex set, where the epigraph is defined as:

\[
\text{epi}(f) = \{(x,t) \in \text{dom}(f) \times \mathbb{R} : f(x) \leq t\}
\]

Intuitively, the epigraph is the set of points that lie above the graph of the function.

• **Convex sublevel sets:** If \( f \) is convex, then every sublevel set of \( f \) is convex, where a sublevel set is defined as

\[
\{x \in \text{dom}(f) : f(x) \leq t\}
\]

for some parameter \( t \in \mathbb{R} \). Unfortunately, the converse of this statement is not true. For example, \( f(x) = \sqrt{|x|} \) is not a convex function but each of its sublevel sets are convex sets.

• **First-order characterization:** If \( f \) is differentiable, then \( f \) is convex if and only if \( \text{dom}(f) \) is convex, and

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

for all \( x,y \in \text{dom}(f) \). Intuitively, the graph of \( f \) must completely lie above each of its tangent hyperplanes. This characterization shows that for a differentiable \( f \), \( x \) minimizes \( f \) if and only if \( \nabla f(x) = 0 \).

• **Second-order characterization:** If \( f \) is twice differentiable, then \( f \) is convex if and only if \( \text{dom}(f) \) is convex, and the Hessian matrix \( \nabla^2 f(x) \) is positive semi-definite for all \( x \in \text{dom}(f) \).

• **Jensen’s inequality:** If \( f \) is convex, and \( X \) is a random variable supported on \( \text{dom}(f) \), then \( f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)] \). A good way to remember the direction of the inequality is to try the function \( f(x) = x^2 \); \( \mathbb{E}[x^2] - \mathbb{E}[x]^2 \) is the variance, which must be non-negative.

2.5 Operations preserving convexity

Like for convex sets, there are some common operations that preserve convexity. They are useful for proving the convexity of functions without resorting to the definition.

• **Non-negative linear combination:** If \( f_1, \ldots, f_m \) are convex, then \( a_1 f_1 + \ldots + a_m f_m \) is convex for any \( a_1, \ldots, a_m \geq 0 \).

• **Pointwise maximization:** If \( f_s \) is convex for any \( s \in S \), then \( f(x) = \max_{s \in S} f_s(x) \) is convex. The set \( S \) does not need to be finite.

• **Partial minimization:** If \( g(x,y) \) is convex in \( x,y \) and \( C \) is a convex set, then \( f(x) = \min_{y \in C} g(x,y) \) is convex.

Pointwise maximization and partial minimization are similar. However, the set \( C \) in partial minimization needs to be convex while the set \( S \) in pointwise maximization does not.
• **Affine composition**: If \( f \) is convex, then \( g(x) = f(Ax + b) \) is convex.

• **General composition**: Suppose \( f = h \circ g \), where \( g : \mathbb{R}^n \to \mathbb{R} \), \( h : \mathbb{R} \to \mathbb{R} \), \( f : \mathbb{R}^n \to \mathbb{R} \). Then:
  - \( f \) is convex if \( h \) is convex and non-decreasing, \( g \) is convex.
  - \( f \) is convex if \( h \) is convex and non-increasing, \( g \) is concave.
  - \( f \) is concave if \( h \) is concave and non-decreasing, \( g \) is concave.
  - \( f \) is concave if \( h \) is concave and non-increasing, \( g \) is convex.

A good way to remember these is to consider \( n = 1 \) and twice-differentiable \( h \) and \( g \), taking the derivative using the chain rule,

\[
f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)
\]

- Suppose \( h \) is convex and non-decreasing, \( g \) is convex. Then \( h''(g(x)) \geq 0 \), \( h'(g(x)) \geq 0 \), and \( g''(x) \geq 0 \), so \( f''(x) \geq 0 \) and \( f \) is convex.
- Suppose \( h \) is convex and non-increasing, \( g \) is concave. Then \( h''(g(x)) \geq 0 \), \( h'(g(x)) \leq 0 \), and \( g''(x) \leq 0 \), so \( f''(x) \leq 0 \) and \( f \) is convex.
- Suppose \( h \) is concave and non-decreasing, \( g \) is concave. Then \( h''(g(x)) \leq 0 \), \( h'(g(x)) \geq 0 \), and \( g''(x) \leq 0 \), so \( f''(x) \leq 0 \) and \( f \) is concave.
- Suppose \( h \) is concave and non-increasing, \( g \) is convex. Then \( h''(g(x)) \leq 0 \), \( h'(g(x)) \leq 0 \), and \( g''(x) \geq 0 \), so \( f''(x) \leq 0 \) and \( f \) is concave.

• **Vector composition**: Suppose \( f(x) = h(g(x)) = h(g_1(x), \ldots, g_k(x)) \) where \( g : \mathbb{R}^n \to \mathbb{R}^k \), \( h : \mathbb{R}^k \to \mathbb{R} \), \( f : \mathbb{R}^n \to \mathbb{R} \). Then
  - \( f \) is convex if \( h \) is convex and non-decreasing in each argument, \( g \) is convex.
  - \( f \) is convex if \( h \) is convex and non-increasing in each argument, \( g \) is concave.
  - \( f \) is concave if \( h \) is concave and non-decreasing in each argument, \( g \) is concave.
  - \( f \) is concave if \( h \) is concave and non-increasing in each argument, \( g \) is convex.

### 2.5.1 Example: Distances to a set

Let \( C \) be an arbitrary set, and let \( f(x) \) be the maximum distance from \( x \) to any point in \( C \), under an arbitrary norm:

\[
f(x) = \max_{y \in C} ||x - y||
\]

\( f_y(x) \) is convex for any fixed \( y \) since it is an affine composition with a norm. Directly applying pointwise maximization, we see that \( f \) is convex.

Now consider a convex \( C \) and let \( f(x) \) be the minimum distance from \( x \) to any point in \( C \), under an arbitrary norm:

\[
f(x) = \min_{y \in C} ||x - y||
\]

\( g(x, y) \) is jointly convex in \( x \), \( y \). Directly applying partial minimization, we see that \( f \) is convex.
2.5.2 Example: log-sum-exp function

This function is also known as “soft max” because it smoothly approximates $\max_{i=1,...,k}(a_i^T x + b_i)$:

$$g(x) = \log \left( \sum_{i=1}^{k} e^{a_i^T x + b_i} \right)$$

To show that $g$ is convex, we only need to show that $f(x) = \log \left( \sum_{i=1}^{n} e^{x_i} \right)$ is convex since $g$ is an affine composition involving $f$. We can show this using the second-order characterization:

$$\frac{\partial}{\partial x_i} f(x) = \frac{e^{x_i}}{\sum_{\ell=1}^{n} e^{x_{\ell}}}$$

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) = \frac{e^{x_i} e^{x_j}}{\sum_{\ell=1}^{n} e^{x_{\ell}}} 1\{i = j\} - \frac{e^{x_i} e^{x_j}}{\left(\sum_{\ell=1}^{n} e^{x_{\ell}}\right)^2}$$

Now, the Hessian matrix can be written as

$$\nabla^2 f(x) = \text{diag}(z) - zz^T$$

where $z_i = e^{x_i} / \left(\sum_{\ell=1}^{n} e^{x_{\ell}}\right)$. This matrix is diagonally dominant, thus positive semi-definite and $f$ is convex.