10-725/36-725: Convex Optimization **Fall 2019**

Lecture 2: August 28

Lecturer: Ryan Tibshirani *Scribes: Shuyang Yang, Xingyu Liu, Bo Lei*

2.1 Convex Optimization Problem

A convex optimization problem is of the form:

$$\min_{x \in D} f(x)$$

subject to

$$g_i(x) \leq 0,\ i = 1, ..., m$$
$$h_j(x) = 0,\ j = 1, ..., r$$

where f and g_i are all convex , and h_j are affine. Any local minimizer of a convex optimization problem is a global minimizer.

2.2 Convex Sets

2.2.1 Definitions

Definition 2.1 *Convex set: a set $C \subseteq \mathbb{R}^n$ is a convex set if for any $x, y \in C$, we have*

$$tx + (1-t)y \in C,\ for\ all\ 0 \leq t \leq 1$$

Definition 2.2 *Convex combination of $x_1, ..., x_k \in \mathbb{R}^n$: any linear combination*

$$\theta_1 x_1 + ... + \theta_k x_k,\ with\ \theta_i \geq 0,\ and\ \sum_{i=1}^k \theta_i = 1$$

Definition 2.3 *Convex hull of set C: all convex combinations of elements in C. The convex hull is always convex.*

Definition 2.4 *Cone: a set $C \subseteq \mathbb{R}^n$ is a cone if for any $x \in C$, we have $tx \in C$ for all $t \geq 0$*

Definition 2.5 *Convex cone: a cone that is also convex, i.e.,*

$$x_1, x_2 \in C \implies t_1 x_1 + t_2 x_2 \in C\ for\ all\ t_1, t_2 \geq 0$$

Definition 2.6 *Conic combination of $x_1, ..., x_k \in \mathbb{R}$: any linear combination*

$$\theta_1 x_1 + ... + \theta_k x_k, \text{ with } \theta_i \geq 0$$

Definition 2.7 *Conic hull of set C: all conic combinations of elements in C.*

2.2.2 Examples of convex sets

- Empty set, point, line.
- Norm ball: $\{x : \|x\| \leq r\}$, for given norm $\|\cdot\|$, radius r.
- Hyperplane: $\{x : a^T x = b\}$, for given a, b.
- Halfspace: $\{x : a^T x \leq b\}$.
- Affine space: $\{x : Ax = b\}$, for given A, b.
- Polyhedron: $\{x : Ax \leq b\}$, where $\leq$ is interpreted componentwise. The set $\{x : Ax \leq b, Cx = d\}$ is also a polyhedron.
- Simplex: special case of polyhedra, given by $\text{conv}\{x_0, ..., x_k\}$, where these points are affinely independent. The canonical example is the probability simplex,

$$\text{conv}\{e_1, ..., e_n\} = \{w : w \geq 0, 1^T w = 1\}$$

2.2.3 Examples of convex cones

- Norm cone: $\{(x, t) : \|x\| \leq t\}$, for given norm $\|\cdot\|$. It is called second-order cone under the l_2 norm $\|\cdot\|_2$.
- Normal cone: given any set C and point $x \in C$, the normal cone is

$$\mathcal{N}_C(x) = \{g : g^T x \geq g^T y, \text{ for all } y \in C\}$$

  This is always a convex cone, regardless of C.
- Positive semidefinite cone:

$$\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : X \succeq 0\}$$

  where $X \succeq 0$ means that X is positive semidefinite ($\mathbb{S}^n$ is the set of $n \times n$ symmetric matrices).

2.2.4 Key properties of convex sets

- **Separating hyperplane theorem**: two disjoint convex sets have a separating between hyperplane them. Formally, if C, D are nonempty convex sets with $C \cap D = \emptyset$, then there exists a, b such that

$$C \subseteq \{x : a^T x \leq b\}, \ D \subseteq \{x : a^T x \geq b\}$$

- **Supporting hyperplane theorem**: a boundary point of a convex set has a supporting hyperplane passing through it. Formally, if C is a nonempty convex set, and $x_0 \in \text{bd}(C)$, then there exists a such that

$$C \subseteq \{x : a^T x \leq a^T x_0\}$$

2.2.5 Operations preserving convexity

2.2.5.1 Operations

- Intersection: the intersection of convex sets is convex.
- Scaling and translation: if C is convex, then $aC + b = \{ax + b : x \in C\}$ is convex for any a, b.
- Affine images and preimages: if $f(x) = Ax + b$ 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. Compared to scaling and translation, this operation also has rotation and dimension reduction.
- Perspective images and preimages: the perspective function is $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ (where $\mathbb{R}_{++}$ denotes positive reals),
$$P(x, z) = x/z$$
for $z > 0$. 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-fractional images and preimages: the perspective map composed with an affine function,
$$f(x) = \frac{Ax + b}{c^T x + d}$$
is called a linear-fractional function, defined on $c^T x + d > 0$. If $C \subseteq \text{dom}(f)$ is convex then so is $f(C)$, and if D is convex then so is $f^{-1}(D)$.

2.2.5.2 Example: linear matrix inequality solution set

Given $A_1, ..., A_k, B \in \mathbb{S}^n$, a **linear matrix inequality** is of the form
$$x_1 A_1 + x_2 A_2 + ... + x_k A_k \preceq B$$
for a variable $x \in \mathbb{R}^k$. Let's prove the set C of points x that satisfy the above inequality is convex.

Approach 1: directly verify that $x, y \in C \Rightarrow tx + (1 - t)y \in C$. This follows by checking that, for any v,
$$v^T \left(B - \sum_{i=1}^{k} (tx_i + (1 - t)y_i) A_i \right) v \geq 0$$
Approach 2: let $f : \mathbb{R}^k \to \mathbb{S}^n$, $f(x) = B - \sum_{i=1}^{k} x_i A_i$. Note that $C = f^{-1}(\mathbb{S}^n_+)$, affine preimage of convex set.

2.2.5.3 Example: conditional probability set

Let U, V be random variables over $\{1, ..., n\}$, $\{1, ..., m\}$. 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. Write
$$D = \left\{ q \in \mathbb{R}^{nm} : q_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}, \text{ for some } p \in C \right\} = f(C)$$
where f is a linear-fractional function, hence D is convex.

2.3 Convex Functions

2.3.1 Definitions

Definition 2.8 *Convex function: f: $\mathbb{R}^n \to \mathbb{R}$ such that the domain of function f $dom(f) \subseteq \mathbb{R}^n$ is convex.*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y),\ \ for\ 0 \leq t \leq 1$$

And all $x, y \in dom(f)$

In other words, the function lies below the line segment joining $f(x)$ and $f(y)$

Definition 2.9 *Concave function: opposite inequality of the definition above, so that*

$$f\ concave \Leftrightarrow -f\ convex$$

which is to say, f being concave implies -f being convex.

Important modifiers:

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

Note: strongly convex implies strictly convex, which subsequently implies convex. In equation format:

$$strongly\ convex \Rightarrow strictly\ convex \Rightarrow convex$$

2.3.2 Examples of convex and concave functions

- Univariate functions
  (1) Exponential function: e^{ax} is convex for any a over $\mathbb{R}$
  (2) Power function: x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ (nonnegative reals); x^a is concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$
  (3) Logarithmic function: $log(x)$ is concave over R_{++}
- Affine function: $a^T x + b$ is both convex and concave.
- Quadratic function: $\frac{1}{2}x^T Qx + b^T x + c$ is convex provided that $Q \geq 0$ (positive semidefinite)
- Least squares loss: $||y - Ax||_2^2$ is always convex (since $A^T A$ is always positive semidefinite)
- $||x||$ is convex for any norm, for example: l_p norms

  $$||x||_p = (\sum_{i=1}^{n} x_p^i)^{1/p}\ for\ p \geq 1, ||x||_\infty = \max_{i=1,...n} |x_i|$$

  as well as operator (spectral) and trace (nuclear) norms

  $$||X||_{op} = \sigma_1(X), ||X||_{tr} = \sum_{i=1}^{r} \sigma_r(X)$$

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

- Indicator function: if C is convex, then its indicator function

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

  is convex

- Support function: for any set C (convex or not), its support function

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

  is convex

- Max function: $f(x) = max\{x_1, ...x_n\}$ is convex.

2.3.3 Key properties of convex functions

- A function is convex if and only if its restriction to any line is convex
- Epigraph characterization: a function f is convex if and only if its epigraph

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

  is a convex set.

- Convex sublevel sets: if f is convex, then its sublevel sets

$$x \in dom(f) : f(x) \leq t$$

  are convex, for all $t \in \mathbb{R}$. The converse is not true.

- First-order characterization: if f is differentiable, then f is convex if and only if dom(f) is convex, and

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

  for all $x, y \in dom(f)$. Therefore for a differentiable convex function $\nabla f(x) = 0 \Leftrightarrow x$ minimizes f.

- Second-order characterization: if f is twice differentiable, then f is convex if and only if $dom(f)$ is convex, and $\nabla^2 f(x) \geq 0$ for all $x \in dom(f)$.
- Jensen's inequality: if f is convex, and X is a random variable supported on $dom(f)$, then $f(\mathbb{E}[\mathbb{X}]) \leq \mathbb{E}[f(x)]$.
- Long-sum-exp function: $g(x) = log(\sum_{i=1}^{k} e^{a_i^T x + b_i})$ for fixed a_i, b_i. This is often called the soft max, since it smoothly approximates $\max_{i=1,...,k}(a_i^T x + b_i)$.

2.3.4 Operations preserving convexity

- Nonnegative linear combination: $f_1, ...f_m$ convex implies $a_1 f_1 + ... + a_m f_m$ is also convex for any $a_1, ...a_m \geq 0$.
- Pointwise maximization: if f_s is convex for any $s \in S$, then $f(x) = max_{s \in S}$ is also convex.
  Note: the set S is the number of functions f_x, which can be infinite.

- Partial minimization: if $g(x, y)$ is convex in x, y, and C is convex, then $f(x) = min_{y \in C} g(x, y)$ is convex.
- Affine composition: if f is convex, then $g(x) = f(Ax + b)$ is convex.
- General composition: suppose $f = hg$, where $g : \mathbb{R}^n \to \mathbb{R}, h : \mathbb{R} \to \mathbb{R}, f : \mathbb{R}^n \to \mathbb{R}$. Then:
  (1) f is convex if h is convex and nondecreasing, g is convex
  (2) f is convex if h is convex and nonincreasing, g is concave
  (3) f is concave if h is concave and nondecreasing, g is concave
  (4) f is convex if h is convex and nonincreasing, g is convex
  Note: To memorize this, think of the chain rule when $n = 1$:

  $$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

- Vector composition: suppose that:

  $$f(x) = h(g(x)) = h(g_1(x), ..., g_k(x))$$

  where $g : \mathbb{R}^n \to \mathbb{R}^k, h : \mathbb{R}^k \to \mathbb{R}, f : \mathbb{R}^n \to \mathbb{R}$. Then:
  (1) f is convex if h is convex and nondecreasing in each argument, g is convex
  (2) f is convex if h is convex and nonincreasing in each argument, g is concave
  (3) f is concave if h is concave and nondecreasing in each argument, g is concave
  (4) f is concave if h is concave and nonincreasing in each argument, g is convex