CS292A Convex Optimization: Gradient Methods and Online Learning **Spring 2019**

Lecture 1: April 2

Lecturer: Yu-Xiang Wang *Scribes: Alon, Dheeraj*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

1.1 Convex optimization problems

Definition 1.1 (Convex optimization problem) *The optimization problem:*

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

subject to

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

is a convex optimization problem when the functions f and g_i are convex, and h_j are affine.
See definition 1.9 for convex function

Important Note: For convex optimization problems, local minima are global minima.

1.2 Convex sets

Definition 1.2 (Convex sets) $C \subseteq \mathbb{R}^n$ *is a convex set* iff*:*

$$x, y \in C \;\Rightarrow\; tx + (1-t)y \in C, \quad \forall\, 0 \leq t \leq 1$$

Definition 1.3 *A* ***Convex Combination*** *of $x_1, ..., x_k \in \mathbb{R}^n$ is any linear combination $\theta_1 x_1 + ... + \theta_k x_k$ where $\theta_i \geq 0$ for all i, and $\sum_{i=1}^n \theta_i = 1$*

The **Convex Hull** of set C is denoted $conv(C)$, and is the set of all convex combinations of the elements in C. The convex hull is always convex.

1.3 Examples of convex sets

Norm Ball: $\{x : \|x\| \leq r\}$
Hyperplane: $\{x : a^T x = b\}$
Half-space: $\{x : a^T x \leq b\}$
Polyhedron: $\{x : Ax \leq b\}$

1.4 Cones

Definition 1.4 (Cone) *$C \subseteq \mathbb{R}^n$ is a cone if*

$$x \in C \;\Rightarrow\; tx \in C, \quad \forall t \geq 0$$

Definition 1.5 (Convex Cone:) *If a cone is convex, then we call it a convex cone, i.e.*

$$x_1, x_2 \in C \;\Rightarrow\; t_1 x_1 + t_2 x_2 \in C, \quad \forall t_1, t_2 \geq 0$$

Definition 1.6 *A **Conic Combination** of $x_1, ..., x_k \in \mathbb{R}^n$ is any linear combination $\theta_1 x_1 + ... + \theta_k x_k$ where $\theta_i \geq 0$ for all i.*

The **Conic Hull** of a set C is denoted $coni(C)$, and is the intersection of all convex cones containing C, plus the origin.

1.5 Examples of convex cones

Norm Cone: $\{(x,t) | \|x\| \leq t\}$, for any norm $\|\cdot\|$.
Normal Cone: Given any set C and a point $x \in C$, the normal cone is defined as

$$N_C(x) = \{g | g^T x \geq g^T y, \forall y \in C\}$$

Positive Semidefinite Cone: $\S^n_+ = \{X \in \S^n | X \succeq 0\}$ is the positive semidefinite cone made up of all nxn positive semidefinite matrices X.

1.6 Properties of convex sets

Definition 1.7 (Separating Hyperplane Theorem) *Any two disjoint convex sets have a hyperplane separating them. So, if C, D are nonempty, disjoint convex sets, then there exists a, b such that*

$$C \subseteq \{x \mid a^T x \leq b\}$$

$$D \subseteq \{x \mid a^T x \geq b\}$$

In plain terms: there is a hyperplane which splits the space into two half-spaces. The set C is fully contained in one half-space, and D in the other half-space.

Definition 1.8 (Supporting Hyperplane Theorem) *If C is a nonempty convex set, and x_0 is in the boundary of C, then $\exists a$ such that*

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

In other words: For any convex set C, every boundary point has a hyperplane passing through it such that C is entirely contained in one of the closed half-spaces bounded by the hyperplane.

1.7 Operations preserving convexity

Intersection: The intersection of convex sets is a convex set.
Affine images and preimages: Any affine transformation of a convex set is also a convex set. Formally, if $f(x) = Ax + b$ and C is convex, then the image of C under f

$$f(C) = \{f(x) \mid x \in C\}$$

is convex. Also, if D is convex, then its preimage

$$f^{-1}(D) = \{x \mid f(x) \in D\}$$

is convex.

Perspective images and preimages: The perspective function over a convex set is a convex set. Additionally, the preimage of the perspective function over a convex set is also a convex set. The perspective function is $P : \mathbb{R}^n \times \mathbb{R}_{++} \longmapsto \mathbb{R}^n$

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

for $z > 0$.

Linear-fractional images and preimages: The perspective map composed with an affine function is called a linear-fractional function and has the form:

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

where $c^T x + d > 0$. If $C \subseteq dom(f)$ is convex, then $f(C)$ is also convex. Also, if D is convex, then so is $f^{-1}(D)$.

1.8 Convex functions

Definition 1.9 (Convex functions) *$f : \mathbb{R}^n \to \mathbb{R}$ is a convex function if $dom(f) \subseteq \mathbb{R}^n$ and $f(tx+(1-t)y) \leq tf(x) + (1-t)f(y)$ for all $0 \leq t \leq 1$ and all $x, y \in dom(f)$.*

*f is **strictly convex** if $f(tx + (1-t)y) < tf(x) + (1-t)f(y)$ for $x \neq y$ and $0 < t < 1$.*

*f is **strongly convex** with parameter $m > 0$ if $f - \frac{m}{2}\|x\|_2^2$ is convex.*

Note that strongly convex ⇒ strictly convex ⇒ convex.

Definition 1.10 (Concave functions) *f is a concave function iff $-f$ is convex.*

1.9 Key properties of convex functions

Epigraph characterization: a function f is convex *iff* 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 $\forall t \in \mathbb{R}$.

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

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

for all $x, y \in dom(f)$. In other words, the tangent at x is a global under-estimator of the function. We see that $\nabla f(x) = 0 \Leftrightarrow x$ minimizes f.

Second-order characterization: if f is twice differentiable, then f is convex *iff* $dom(f)$ is convex and its Hessian $\nabla^2 f(x) \succeq 0$ (positive semidefinite).

Jensen's inequality: if f is convex and X is a random variable supported on $dom(f)$, the $f(\mathbb{E}[X] \leq \mathbb{E}[f(X)]$.

1.10 Examples of Convex functions

Univariate functions:

- Exponential function: e^{ax} is convex for any a on $\mathbb{R}$.
- Power function: x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ (non-negative reals). It is concave for $0 \leq a \leq 1$.
- Logarithmic function: $\log(x)$ is concave over $\mathbb{R}_{++}$ (set of positive reals)

Affine functions: $f(x) = a^T x + b$ is both convex and concave

Quadratic functions: $f(x) = \frac{1}{2} x^T Q x + b^T x + c$ is convex provided that $Q \succeq 0$. This follows directly by observing that Hessian of f is Q.

Least squares loss: $f(x) = \|y - Ax\|_2^2$ is always convex in x because f can be represented as $x^T A^T A x - 2y^T A x + y^T y$. Since $A^T A$ is positive semidefinite for any A, f is a quadratic convex function.

Norms:

- The l_p norm of x, denoted by $\|x\|_p$, is convex for any $p \geq 1$ where,

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^n x_i^p\right)^{1/p} & \text{for } 1 \leq p < \infty \\ \max_i |x_i| & \text{p} = \infty \end{cases}$$

- The operator (spectral) and trace (nuclear) norms of a matrix defined by $\|\boldsymbol{X}\|_{op} = \sigma_1(\boldsymbol{X})$ and $\|\boldsymbol{X}\|_{tr} = \sum_{i=1}^r \sigma_r(\boldsymbol{X})$ is convex in $\boldsymbol{X}$. Here $\sigma_1(\boldsymbol{X}) \geq ... \geq \sigma_r(\boldsymbol{X})$ are singular values of $\boldsymbol{X}$

Indicator function: If C is a convex set, then its indicator function $I_C(x)$ is convex where,

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

Support function: for any set C (convex or not), its support function $I_C^*(x) = \max_{y \in C} x^T y$ is convex.

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

1.11 Operations preserving convexity

Non-negative linear combination: $f_1, ..., f_m$ convex implies $a_1 f_1 + ... + a_m f_m$ is convex for any $a_1, ..., a_m \geq 0$.

Pointwise maximization: if f_i is convex for any $i \in I$, where I is a possibly infinite set, then $f(x) = \max_{i \in I} f_i(x)$ is convex.

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. This result trivially extends to partial minimization over a subset of the function's arguments.

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 nondecreasing, g is convex
- f is convex if h is convex and nonincreasing, g is concave
- f is concave if h is concave and nondecreasing, g is concave
- f is concave if h is concave and nonincreasing, g is convex

To remember these, think of chain rule for $n = 1$ and see how we can make f'' positive.

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

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:

- f is convex if h is convex and nondecreasing in each argument, g is convex
- f is convex if h is convex and nonincreasing in each argument, g is concave
- f is concave if h is concave and nondecreasing in each argument, g is concave
- f is concave if h is concave and nonincreasing in each argument, g is convex

1.11.1 Examples

Proposition 1.11 *Let C be an arbitrary set and x be an arbitrary point. The maximum distance function to C under an arbitrary norm $\|\cdot\|$ $f(x) = \max_{y \in C} \|x - y\|$ is convex.*

Proof: Consider $f_y(x) = \|x - y\|$, the distance from x to a fixed $y \in C$. Then f is a pointwise maximum of convex functions represented as $f(x) = \max_{y \in C} f_y(x)$ ■

Proposition 1.12 *The distance between a point x and its projection to a convex set C given by $d(x) = \min_{y \in C} \|x - y\|$ is convex*

Proof: Let $h(x, y) = \|x - y\|$. Then $d(x) = \min_{y \in C} h(x, y)$ which is a partial minimization of a convex function over a convex set. ■

Proposition 1.13 *The soft max function $g(x) = \log(\sum_{i=1}^{k} e^{a_i^T x + b_i})$, for fixed a_i, b_i is convex.*

Proof: Due to affine composition rule, it is sufficient to show that $f(x) = \log(\sum_{i=1}^k e^{x_i})$ is convex. We make use of Hölder's inequality which states that $x^T y \leq \|x\|_p \|y\|_q$ where $\frac{1}{p} + \frac{1}{q} = 1$. For $\lambda \in (0,1)$, We have,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \log(\sum_{i=1}^k e^{\lambda x_i + (1-\lambda)y_i}) \\ &\leq \log\left(\left(\sum_{i=1}^k \left(e^{\lambda x_i}\right)^{\frac{1}{\lambda}})\right)^{\lambda} \cdot \left(\sum_{i=1}^k \left(e^{(1-\lambda)y_i}\right)^{\frac{1}{1-\lambda}}\right)^{1-\lambda}\right) \\ &= \lambda f(x) + (1-\lambda) f(y), \end{aligned}$$

where in the second step, we applied Hölder's inequality with $p = \frac{1}{\lambda}$ and $q = \frac{1}{1-\lambda}$. ■

Remark: The function $f(x) = \log(\sum_{i=1}^k e^{x_i})$ smoothly approximates $\max_i x_i$. This can be seen by noting that,

$$\max_i x_i \leq f(x) \leq \log(k) + \max_i x_i$$