MATH 680: Computation Intensive Statistics **Winter 2018**

Lecture 1. Convex Sets and functions: Wednesday January 24

Lecturer: *Scribes:*

1.1 Convex sets

- The domain of f

$$\text{dom}(f) = \{x : \ f(x) \text{ is defined and finite}\}$$

- Affine function

$$h_j(x) = 0 \Leftrightarrow h_j(x) \leq 0 \ h_j(x) \geq 0$$

Definition 1. Convex set: $C \subset \mathbb{R}^n$ such that if $x, y \in C$ then $tx + (1-t)y \in C$ for all $0 \leq t \leq 1$.

Definition 2. Convex function: $f : \mathbb{R}^n \to \mathbb{R}$ such that $\text{dom}(f) \subset \mathbb{R}^n$ convex, and

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

for $0 \leq t \leq 1$ and all $x, y \in \text{dom}(f)$.

Definition 3. Optimization problem

$$\begin{aligned} \min_{x \in D} & f(x) \\ \text{subject to } & g_i(x) \leq 0, \ \ i = 1, ..., m \\ & h_j(x) = 0, \ \ j = 1, ..., p \end{aligned}$$

Here $D = \text{dom}(f) \cap \bigcap_{i=1}^m \text{dom}(g_i) \cap \bigcap_{i=1}^p \text{dom}(h_j)$, common domain of all the functions.

Definition 4. Convex optimization problem: optimization problem set-up above provided that the functions f and g_i, $i = 1, ..., m$ are convex, and h_j, $j = 1, ..., p$ are affine:

$$h_j(x) = a_j^T x + b_j, \ \ j = 1, ..., p$$

Comments: Note we can represent the constraints as follow:

1. $g(x) \geq 0$ and $-g(x) \leq 0$.
2. $h(x) \leq 0$ and $h(x) \geq 0 \iff h(x) = 0$.
3. Domain of convex optimization problem is always convex (intersection of convex sets is also convex set).
4. $\min_x f(x) \iff \max_x -f(x)$

Motivation for convex problems: local minima = global minima!

Proof. Use contradiction. If x is not a global minima, then there must exist some feasible $z \in D$ such that

$$f(z) < f(x)$$

then

$$||z - x||_2 > \rho$$

Now we choose

$$y = tx + (1-t)z$$

for some $0 \leq t \leq 1$, then $\square$

- $y \in D$
- y satisfies the constraints

$$\begin{aligned} h_j(y) &= a_j^T(tx + (1-t)z) + b_j \\ &= 0 \end{aligned}$$

$$\begin{aligned} g_i(y) &\leq tg_i(x) + (1-t)g_i(z) \\ &\leq 0 \end{aligned}$$

- Now take a very large value of t such that $||y - x||_2 \leq \rho$. By the convexity of f, we have

$$\begin{aligned} f(y) &= f(tx + (1-t)z) \\ &\leq tf(x) + (1-t)f(z) \\ &< tf(x) + (1-t)f(x) \\ &= f(x). \end{aligned}$$

  Therefore we have found y in the neighborhood of x and $y < x$. This contradicts with the fact that x is the local minimum.

Definition 5. Convex combination of $x_1, ..., x_k \in \mathbb{R}^n$ is any linear combination

$$\theta_1 x_1 + ... + \theta_k x_k$$

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

Definition 6. Convex hull of a set C, denoted $\text{conv}(C)$, is the set of all the convex combinations of elements. It is the smallest convex set that contains set C (note C is not necessarily convex).

Examples of convex sets:

- Norm ball: $\{x : ||x|| \leq r\}$. It is a convex set by using the triangular inequality of the norm

$$\begin{aligned} \|tx + (1-t)y\| &\leq t\|x\| + (1-t)\|y\| \\ &\leq r \end{aligned}$$

- Affine space:

$$Cx = d \Longleftrightarrow Cx \leq d \ \ -Cx \leq -d$$

- Polyhedron:
$$\{x : Ax \leq b\} \iff \{x : \ a_i^T x \leq b_i \ i = 1, \dots, m\}$$
a_i is the i-th row of A (note: this is an intersection of m halfspaces);
  - This definition generalizes to $\{x : Ax \leq b, Cx = d\}$ since we can rewrite the $Cx = d$ constraint using inequality constraints.
- 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
$$w^T e = \text{conv}\{e_1, ..., e_n\} = \{w : w \geq 0, 1^T w = 1\}$$
where
$$e = [e_1, e_2, \dots, e_n]$$
where
$$e_1 = (1, 0, \dots, 0)^T$$
where $e_1, ..., e_n$ are the standard basis vectors in $\mathbb{R}^n$ and $w \in \mathbb{R}^n$.
- **Note:** $x_0, \dots, x_k$ are affine independent $\iff$ $x_1 - x_0, ..., x_k - x_0$ are linearly independent.

Definition 7. Cone $C \subset \mathbb{R}^n$ such that $x \in C \implies tx \in C$ for all $t \geq 0$.

- **Note:** 0 must be lies in the cone.

Definition 8. Normal cone: $N_c(x)$ is a normal cone to set C at the point $x \in C$ which satisfies
$$N_c(x) = \{g : \langle g, y - x\rangle \leq 0 \text{ for all } y \in C\}$$

Proposition 1. *Normal cone is convex cone.*

Proof. We show $N_c(x)$ is a cone and convex □

1. To show $N_c(x)$ is a cone
   (a) Fix any $g \in N_c(x)$ and $t \geq 0$
   (b) By definition
   $$\langle g, y - x\rangle \leq 0 \text{ for all } y \in C$$
   then
   $$\langle tg, y - x\rangle = t\,\langle g, y - x\rangle \leq 0 \text{ for all } y \in C$$
   Thus $tg \in N_c(x)$. Therefore $N_c(x)$ is a cone
2. To show $N_c(x)$ is a convex set $g_1, g_2 \in N_c(x)$, we want to show $tg_1 + (1-t)g_2 \in N_c(x)$
   (a) Fix $g_1, g_2 \in N_c(x)$ and $t \in [0, 1]$
   $$\langle g_1, y - x\rangle \leq 0 \text{ for all } y \in C$$
   $$\langle g_2, y - x\rangle \leq 0 \text{ for all } y \in C$$

Thus

$$\begin{aligned}\langle tg_1 + (1-t)g_2, y-x\rangle &= t\,\langle g_1, y-x\rangle + (1-t)\,\langle g_2, y-x\rangle \\ &\le 0 \text{ for all } y \in C\end{aligned}$$

Therefore

$$tg_1 + (1-t)g_2 \in N_c(x)$$

Thus $N_c(x)$ is a convex set.

Note:

- For $N_c(x)$, set C can be any set, not necessarily convex
- $N_c(x) = \{0\}$ for any x inside C

Basic linear algebra facts:

- $X \in \mathbb{S}^n \Longrightarrow \lambda(X) \in \mathbb{R}^n$
- $X \in \mathbb{S}^n_+ \Longrightarrow \lambda(X) \in \mathbb{R}^n_+$
- $X \in \mathbb{S}^n_{++} \Longrightarrow \lambda(X) \in \mathbb{R}^n_{++}$

We can define an inner product over $\mathbb{S}^n$: given $X, Y \in \mathbb{S}^n$,

$$\langle X, Y\rangle = \mathrm{tr}(XY)$$

We can define a partial ordering over $\mathbb{S}^n$: given $X, Y \in \mathbb{S}^n$,

$$X \succeq Y \Longleftrightarrow X - Y \in \mathbb{S}^n_+$$

Proposition 2. *The set of all positive semidefinite matrices $x \in \mathbb{S}^n_+$ is a convex cone*

1. If $x \in \mathbb{S}^n_+$ then $tx \in \mathbb{S}^n_+$ for $t \ge 0$
2. $x, y \in \mathbb{S}^n_+$ then $tx + (1-t)y \in \mathbb{S}^n_+$ for $t \ge 0$

Proposition 3. *The set of all the points x that satisfies*

$$\{x : \ x_1A_1 + \cdots x_kA_k \preceq B\}$$

is a convex set.

Proof. Method 1: Let $C = \{x : \ x_1A_1 + \cdots x_kA_k \preceq B\}$.
Assume $\forall x, \ y \in C$ and let

$$z = tx + (1-t)y \ , t \in [0,1]$$

We want to show $z \in C$

$$\begin{aligned}&B - (z_1A_1 + \cdots + z_kA_k) \\ =&B - ((tx_1 + (1-t)y_1))A_1 + \cdots + (tx_k + (1-t)y_k)A_k) \\ =&t(B - (x_1A_1 + \cdots + x_kA_k)) + (1-t)(B - (y_1A_1 + \cdots + y_kA_k)) \\ \succeq&0\end{aligned}$$

Method 2:

$$\{x:\ x_1A_1+\cdots x_kA_k \preceq B\} \Longleftrightarrow \{x:\ f(x)\succeq 0\},\ \mathbb{S}_+^n \text{ is convex so } f^{-1}(\mathbb{S}_+^n) \text{ is convex}$$

□

Question: We have two disjoint convex sets. Do we always have a hyperplane which strictly separate two sets.
Answer: No

Proof. For example let $C = \{x : a^Tx \leq b\}$, $D = \{x : a^Tx > b\}$. C and D can not be strictly separated by a hyperplane. One of the set is open. □

1.2 Convex functions

Example 1. Indicator function is a convex function

$$I_C(tx+(1-t)y) \leq tI_C(x)+(1-t)I_Cy$$

If $x, y \in C$, then

$$0 \leq 0$$

If $x \notin C$ and $y \notin C$

$$... \leq \infty$$

Proposition 4. *Note: strongly convex* $\Longrightarrow$ *strictly convex* $\Longrightarrow$ *convex*

Proposition 5. *If f is differentiable, and* $\forall x, y \in \operatorname{dom}(f)$,

- f is **convex** $\Longleftrightarrow$ $f(y) \geq f(x) + \nabla f(x)(y-x)$
- f is **strictly convex** $\Longleftrightarrow$ $f(y) > f(x) + \nabla f(x)(y-x)$
- f is **strongly convex**

$$\Longleftrightarrow\ f(y) \geq f(x) + \nabla f(x)(y-x) + \frac{m}{2}||y-x||_2^2$$

  ie,

- if $m = 0$, convex function
- if $m \to 0$, strictly convex
- if $m > 0$, strongly convex

Proposition 6. *If f is twice continuously differentiable,*

- f is **convex** $\Longleftrightarrow$

$$f''(x) \geq 0, \quad \forall x \in \operatorname{dom}(f)$$

$$\nabla^2 f(x) \succeq 0, \quad \text{(positive semidefinite)}$$

- f is **strictly convex** $\Longleftarrow$

$$f''(x) > 0, \quad \forall x \in \mathrm{dom}(f)$$
$$\nabla^2 f(x) \succ 0 \quad \text{(positive definite)}$$

- f is **strongly convex** $\Longleftrightarrow$

$$f''(x) \geq m > 0 \quad \forall x \in \mathrm{dom}(f)$$

$$\nabla^2 f(x) \succeq m \succ 0 \quad \text{(bounded)}$$

Example 2. If f is strictly convex with $f''(x_n) = \frac{1}{n}$, then it is not strongly convex since

$$\lim_{n\to\infty} f''(x_n) = \lim_{n\to\infty} \frac{1}{n} = 0$$

Example 3. Least squares loss.

$$\min_\beta f(\beta) \iff \min_\beta ||y - X\beta||_2^2$$

$$\nabla^2 f(\beta) = X^T X \succeq 0$$

1. $X^T X \succ 0, n \geq p$ full column rank;
2. $X^T X \succeq 0$, otherwise.

Proposition 7. *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 $x, y \in \mathrm{dom}(f)$. *So for a differentiable convex function,*

$$\nabla f(x) = 0 \implies x \textit{ minimizes } f$$

Example 2: Nonnegative linear combination - Logistic Regression
In logistic regression, we wish to solve

$$\min_\beta \sum_{i=1}^n log(1 + exp\{-y_i x_i^T \beta\})$$

for $x_i \in \mathbb{R}^p$ and $y_i \pm 1$. To verify that this function is convex, we need to verify if

$$f(t) = log(1 + exp(t))$$

is convex. We take the second derivative,

$$f''(t) = \frac{e^t}{(1 - e^t)^2} > 0$$

and conclude that it is convex.