10-725/36-725: Convex Optimization **Fall 2016**

Lecture 2: August 31

Lecturer: Lecturer: Ryan Tibshirani *Scribes: Scribes: Lidan Mu, Simon Du, Binxuan Huang*

2.1 Review

A convex optimization problem is of the form

$$\begin{aligned} min_{x\in D} \quad & f(x) \\ subjectto \quad & g_i(x) \leq 0, i = 1, ..., m \\ & h_j(x) = 0, j = 1, ..., r \end{aligned}$$

where f and g_i, $i = 1, ..., m$ are all convex, and $h_j, j = 1, ..., r$ are affine. A local minimizer for a convex optimization is a global minimizer.

2.2 Convex Sets

2.2.1 Definition

Convex set is a set $C \subseteq R^n$ such that

$$x, y \in C \Rightarrow tx + (1-t)y \in C \text{ for all } 0 \leq t \leq 1$$

In other words, line segment joining any two elements lies entirely in the set.

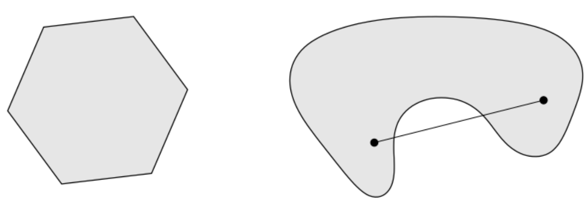

Figure 2.1: A convex set and a nonconvex set

Convex combination of $x_1, ..., x_k \in R^n$ is any linear combination

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

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

Convex hull of set C, $conv(C)$, is all convex combinations of elements. A convex hull is allways convex, but set C is not required to be convex.

2.2.2 Examples of Convex Sets

Here are some examples of convex sets:

Trivial ones: 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 inquality $\leq$ is interprated componentwise—for any vectors x, y $x \leq y$ means $x_i \leq y_i$ for all i. Note: the set $\{x : Ax \leq b, Cx = d\}$ is also a polyhedron, because it is equivalent to $\{x : Ax \leq b, Cx \leq d, -Cx \leq -d\}$ **Simplex**: it is a special case of polyhedra, given by $conv\{x_0, ..., x_k\}$, where

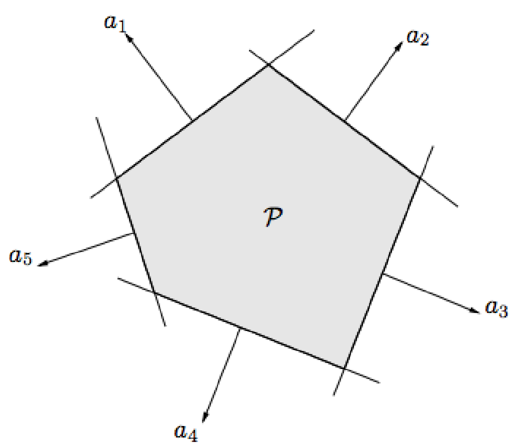

Figure 2.2: A polyhedron in two dimensional space, where $\{a_i\}$ is A's row.

these points are affinely independent. The canonical example is the probability simplex, $conv\{e_1, ..., e_n\} = \{w : w \geq 0, 1^T w = 1\}$.

Two related definition:
$x_0, ..., x_k$ are affinely independent means $x_1 - x_0, ..., x_k - x_0$ are linear independent.
$x_0, ...x_k$ are linear independent means $a_0 x_0 + ... + a_k x_k = 0 \Rightarrow a_0 = ... = a_k = 0$

2.3 Cones

2.3.1 definition

Cone is a set $C \subseteq R^n$ such that

$$x \in C \Rightarrow tx \in C \text{ for all } t \geq 0$$

Convex cone is a cone that is also convex, i.e.,

$$x_1, x_2 \in C \Rightarrow t_1 x_1 + t_2 x_2 \in C \text{ for all } t_1, t_2 \geq 0$$

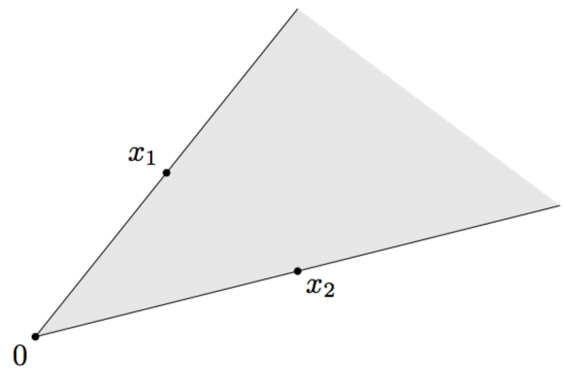

Figure 2.3: A convex cone in two dimensional space

Note there exist some non-convex cones. One example is two intersecting lines.

Conic combination of $x_1, ..., x_k \in R^n$ is any linear combination

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

with $\theta_i \geq 0$, $i = 1, ..., k$.

Conic hull of $\{x_1, ..., x_k\}$ is collection of all conic combinations $\{\sum_i \theta_i x_i : \theta \in R_+^k\}$.

2.3.2 Examples of Convex Cones

Norm cone: $\{(x, t_=(: ||x|| \leq t\}$, for a norm$||\cdot||$. Under l_2 norm $||\cdot||_2$, it is called second-order cone.

Normal cone: given any set C and point $x \in C$, we can define normal cone as

$$N_C(x) = \{g : g^T x \geq g^T y \text{ for all } y \in C\}$$

Normal cone is always a convex cone.

Proof: For $g_1, g_2 \in N_C(x)$, $(t_1 g_1 + t_2 g_2)^T x = t_1 g_1^T x + t_2 g_2^T x \geq t_1 g_1^T y + t_2 g_2^T y = (t_1 g_1 + t_2 g_2)^T y$ for all $t_1, t_2 \geq 0$

Positive semidefinite cone is $S_+^n = \{X \in S^n : X \succeq 0\}$, where $X \succeq 0$ means that X is positive semidefinite (and S^n is the set of $n \times n$ matrices)

Positive semidefinite: a matrix X is positive semidefinite if all the eigenvalues of X are larger or equal to $0 \iff a^T X a \geq 0$ for all $a \in R^n$.

2.4 Key properties of convex sets

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

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

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

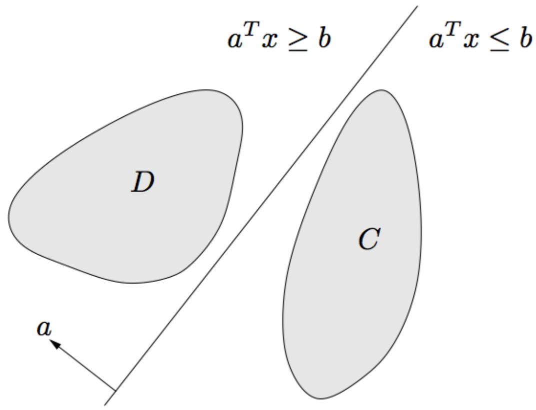

Figure 2.4: A line separates two disjoint convex sets in two dimensional space

Supporting hyperplane theorem: if C is a nonempty convex set, and $x_0 \in \mathit{boundary}(C)$, then there exists a such that

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

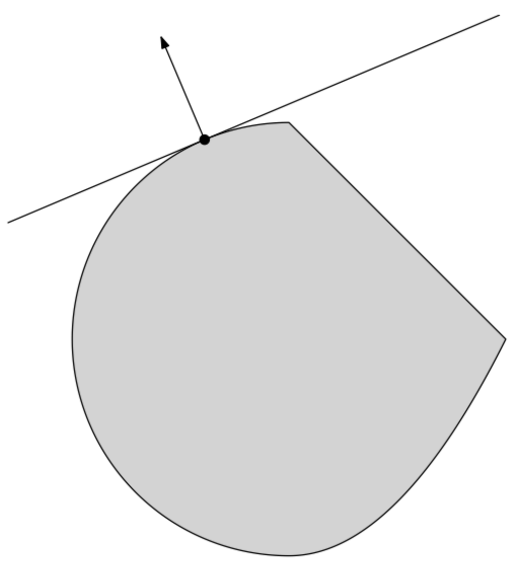

Figure 2.5: A supporting hyperplane that passing a boundary point of a convex set in two dimensional space

2.5 Operations Preserving Convexity of Convex Sets

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

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

is convex, and if D is convex then

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

is convex. Note here f^{-1} does not mean f must be inversible.

2.5.1 Example: linear matrix inequality solution set

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

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

for a variable $x \in R^k$.

Let's prove the set C of points x that satisfy the above inequality is convex.

Approach 1: directly verify that $x, y \in C \Rightarrow tx + (1 - t)y \in C$.

Then for any v,

$$\begin{aligned}
&v^T(B - \sum_{i=1}^{k}(tx_i + (1-t)y_i)A_i)v \\
=&v^T[t(B - \sum_i x_iA_i)]v + v^T[(1-t)(B - \sum_i y_iA_i)]v \\
\geq&0
\end{aligned}$$

Approach 2: let $f : R^k \to S^n$, $f(x) = B - \sum_{i=1}^{k} x_iA_i$. Note that $C = f^{-1}(S^n_+)$, affine preimage of convex set.

2.6 Convex Functions

2.6.1 Definitions

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

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \quad \text{for } 0 \leq t \leq 1$$

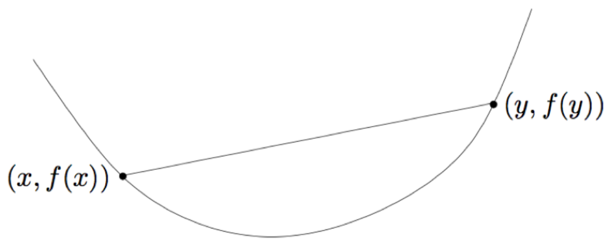

Figure 2.6: Convex function

and all $x, y \in \text{dom}(f)$. In other words, f lies below the line segment joining $f(x), f(y)$ as shown in the following figure.

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

$$f(tx + (1-t)y) \geq tf(x) + (1-t)f(y) \quad \text{for } 0 \leq t \leq 1$$

and all $x, y \in \text{dom}(f)$. So that we have

$$f \text{ concave } \Leftrightarrow -f \text{ convex.}$$

Important modifiers:

Strictly convex means that

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y) \quad \text{for } 0 \leq t \leq 1$$

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 $m > 0$ means that $f - \frac{m}{2}||x||_2^2$ is convex. In words, f is at least as convex as a quadratic function.

Note that strongly convex $\Rightarrow$ strictly convex $\Rightarrow$ convex. For example, function $f(x) = \frac{1}{x}$ is strictly convex but not strongly convex.

2.6.2 Examples of Convex Functions

Univariate functions

- **Exponential function** e^{ax} is convex for any a over $\mathbb{R}$
- **Power function** x^a is convex for $a \geq 1$ or $a \leq 0$ over $\mathbb{R}_+$ and concave for $0 \leq a \leq 1$ over $\mathbb{R}_+$
- **Logarithmic function** $\log x$ is concave over $\mathbb{R}_{++}$

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 $Q \succeq 0$

Least squares loss $||y - Ax||_2^2$ is always convex (since $A^T A$ is always positive semidefinite)

Norm $||x||$ is convex for any norm. For example, l_p norms,

$$||x||_p = \left(\sum_{i=1}^{n}\right)^{1/p} x_i^p \quad \text{for } p \geq 1, \quad ||x||_\infty = \max_{i=1,\ldots,n} |x_i|$$

and also operator (spectral) and trace (nuclear) norms,

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

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

2.7 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 is a convex set, where the epigraph is defined as

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

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

Convex sublevel sets If f is convex, then its sublevel sets

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

are convex, for all $t \in \mathbb{R}$. The converse 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 $f(y) \geq f(x) + \nabla f(x)^T (yx)$ for all $x, y \in \text{dom}(f)$. In other words, f must completely lie above each of its tangent hyperplanes. Therefore 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 $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom}(f)$.

Jensens inequality If f is convex, and X is a random variable supported on $\text{dom}(f)$, then $f(\text{E}[X]) \geq \text{E}[f(X)]$.

2.8 Operations Preserving Convexity of Convex Functions

Nonnegative linear combination $f_1, ..., f_m$ convex implies $a_1 f_1 + ... + a_m f_m$ convex for any $a_1, ..., a_m \geq 0$.

Pointwise maximization if f_s is convex for any $s \in S$, then $f(x) = \max_{s \in S} f_s(x)$ is convex. Note that the set s here 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.

2.8.1 Example: distances to a set

Let C be an arbitrary set, and consider the maximum distance to C under an arbitrary norm $||.||$:

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

Proof: $f_y(x) = ||x - y||$ is convex in x for any fixed y, so by pointwise maximization rule, f is convex. ■

Let C be convex, and consider the minimum distance to C:

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

Proof: $g(x, y) = ||x - y||$ is convex in x, y jointly, and C is assumed convex, so by applying partial minimization rule, f is convex.

■