EE 381V: Large Scale Optimization **Fall 2014**

Lecture 2 — September 2

Lecturer: Sanghavi *Scribe: Chang-sik Choi, Jeonghun Park, Kai Zhong*

2.1 Overview of the last Lecture

Last lecture, we investigate some property on the definition of convex set and convex equation. A convex optimization problem can be described as follows:

$$\begin{aligned} &\min f_0(x) \\ \text{s.t. } &f_i(x) \leq 0, \quad i = 1, 2, ..., m \\ &a_i^T x = b_i, \quad i = 1, 2, ..., p \end{aligned} \tag{2.1}$$

where $f_i, i = 0, 1, \ldots, m$ are all convex functions.
Note that for a convex function, g, the constraint

$$g(x) \geq 0 \tag{2.2}$$

will have a feasible set that could be non convex.

2.2 Lecture 2 overview

We introduce unconstrained problem and investigate its optimality. Unconstrained problem can be denoted as:

$$\min\{f_o(x)\}. \tag{2.3}$$

Necessary/sufficient conditions for the optimality are as follows:

$$x \text{ is global minimum} \Rightarrow \nabla f_o(x) = 0, \text{for all function} \tag{2.4}$$

$$x \text{ is global minimum} \Leftarrow \nabla f_o(x) = 0, \text{for convex function.} \tag{2.5}$$

Proof: Recall the cone structure at the definition of convexity. For all x and y,

$$f_o(y) \leq f_o(x) + (y - x)^T \nabla f_o(x).$$

For $\hat{x}$, $\nabla f_o(\hat{x}) = 0$. Furthermore, $f_o(y) \leq f_o(\hat{x})$ for all y. In today lecture, we also introduce examples of convex set and associated convex problems. Examples such as convex polar cone, normal cone will be covered.

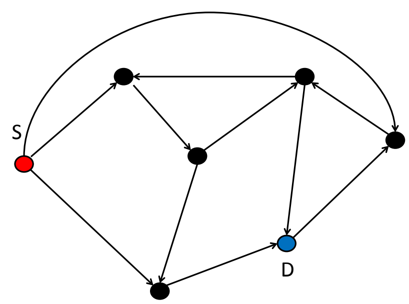

Figure 2.1. Max Flow Problem in a Directed Graph.

2.3 Convex Modeling

In this subsection, we covered convex modeling and its applications.

2.3.1 Max Flow and Min Cut

Suppose a directed graph $G = (V, E)$. one realization of directed graph is shown in figure 1. Let V denotes the set of the graph's vertices, and E denotes the set of edges. Each link has finite capacity with c_e. Source and drain vertexes are described as v_s and v_d. Variables are the flows on each edge and will be described as f_e. Obviously, $f_e \geq 0$ and $f_e \leq c_e$. One possible question one can ask is *what is the maximum flow between source and drain?*. Mathematically, it can be represented as maximizing the objective function:

$$\sum_{e \in S_{out}} f_e, \tag{2.6}$$

under the constraints:

$$\sum_{e \in v_{in}} f_e = \sum_{e \in v_{out}} f_e, \forall v \in V \setminus \{s, d\}, \tag{2.7}$$

$$0 \leq f_e \leq c_e, \forall e \in E, \tag{2.8}$$

where the first constraint describes the flow conservation law, and the second constraint denotes the natural condition of flow on edges.

2.3.2 Optimal Inequalities in Probability

Now consider a totally different problem. Assume X is a integer-valued random variable. Given the following moment constraints:

$$\mu_i = \mathbb{E}[X^i], \quad i = 1, 2, 3, 4, 5. \tag{2.9}$$

The goal is to find the lower bound and upper bound for $\mathbb{P}\{X \in [20, 30]\}$.
Let $P_j = P\{X = j\}$, the problem of finding the upper bound can be formulated by an optimization problem,

$$\begin{aligned} &\max \sum_{j=20}^{30} P_j \\ &\text{s.t. } P_j \geq 0, \text{ for any } j \\ &\qquad \sum_j P_j = 1 \\ &\qquad \sum_j P_j j^i = \mu_i, \quad i = 1, 2, 3, 4, 5. \end{aligned} \tag{2.10}$$

This is an optimization problem with infinitely many variables, and other than the non-negativity constraints, there are 6 other constraints.

2.3.3 Matrix Completion

In many practical problems, one wants to recover a matrix from some of its entries. For example, in the Netflix prize, given a few ratings between the users and the films, the goal is to predict all the other ratings. See Figure 2.2. Without any assumptions, this is an impossible task. However, noting that the rank of a matrix is an important measure for a matrix, we can model this matrix completion problem as an affine rank minimization problem,

$$\begin{aligned} &\min \operatorname{rank}(X) \\ &\text{s.t. } X_{ij} = M_{ij}, \text{ for observed } (i, j) \end{aligned}$$

	Users						
Films	1		2			4	
		3			2		
	4		5	3			3
		2			3		
				5		2	

Figure 2.2. Matrix Completion

However the rank minimization problem is generally a non-convex problem for which the running time of most known algorithms is exponential. One way to deal with it is to apply convex relaxation. The most popular convex relaxation of the rank is the nuclear norm, which is defined as,

$$\|X\|_* := \sum_{i=1}^{r} \sigma_i(X),$$

where r is the rank and σ_i is the singular values of X. Now the matrix completion problem can be formulated as,

$$\begin{aligned} &\min \|X\|_* \\ &\text{s.t. } X_{ij} = M_{ij}, \text{ for observed } (i,j) \end{aligned}$$

This optimization problem can be solved efficiently by semidefinite programing.

2.4 Characterizing Optimal Solutions of Constrained Convex Optimization

We now turn to obtaining a characterization of the optimal solution to a constrained convex optimization problem, with a smooth convex objective function.

$$\begin{aligned} &\min f_0(x) \\ &\text{s.t. } f_i(x) \leq 0, \quad i = 1, 2, ..., m \end{aligned} \tag{2.11}$$

where $f_0(x)$ is smooth convex objective function, and $\{f_i(x)\}_{i=1,2,...,m}$ are convex functions. We define the feasible set $\mathcal{X}$ as $\mathcal{X} := \{x | f_i(x) \leq 0, \ i = 1, 2, ..., m\}$.
We will begin with some further concepts related to convex sets. First we need the definition of distance in order to talk about closed and open sets, convergence, limits, etc.

Definition 1. *A* metric *is a mapping $d(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ that satisfies: $d(x, y) \geq 0$ with equality iff $x = y$; $d(x, y) = d(y, x)$, and $d(x, z) + d(z, y) \geq d(x, y)$.*

Now we can give a simple definition of open and closed sets.

Definition 2. (Open Sets) *A set $C \subseteq \mathbb{R}^n$ is called* **open** *if $\forall x \in C$, $\exists \varepsilon > 0$ such that $B_\varepsilon(x) \subseteq C$.*

Example 1. *The unit sphere without boundary in $\mathbb{R}^n$ is an open set: $B_1(0) = \{x : \|x\| < 1\}$.*

Example 2. *The set of all $n \times n$ symmetric matrices with strictly positive eigenvalues, S^n_+, is an open set.*

Definition 3. (Closed Sets) *A set $C \subseteq \mathbb{R}^n$ is called* **closed** *if $x_n \in C, x_n \to \bar{x} \ \Rightarrow \ \bar{x} \in C$.*

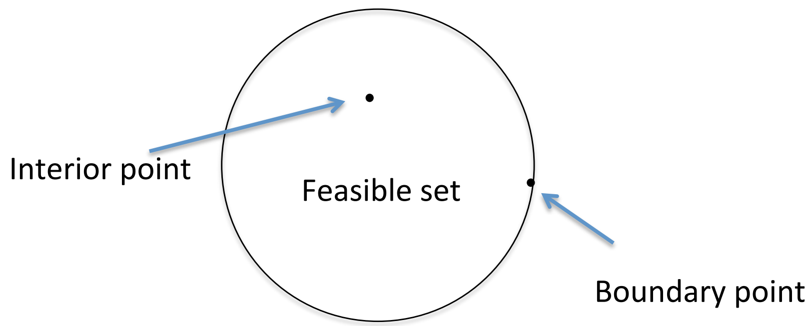

Figure 2.3. Interior points and boundary points.

Definition 4. (Interior) *A point $x \in C$ is an* **interior** *point of C, denoted by $x \in \text{Int}C$, if $\exists \varepsilon > 0$ such that $B_\varepsilon(x) \subseteq C$.*

The distinction between an interior point and a boundary point, is important for optimization, since no constraints are active at an interior point. The notion of *relative interior* is particularly important in convex analysis and optimization, because many feasible sets are often expressed using an intersection with an affine space that is not full dimensional. We do not go into the details here, but simply give the basic idea through a picture. For a point to be in the interior of a set, the set must contain a small ball around that point. Relative interior replaces this with the requirement that the set contain the intersection of a ball around the point and the affine hull of the set itself. Figure **??** illustrates this idea. The shaded circular shape lies in three dimensions, but its affine hull is only two dimensional, and hence it can have no interior. However, the notion of relative interior recovers the intuitive notion of "interior" of the shape, recognizing that it is really a two-dimensional object.

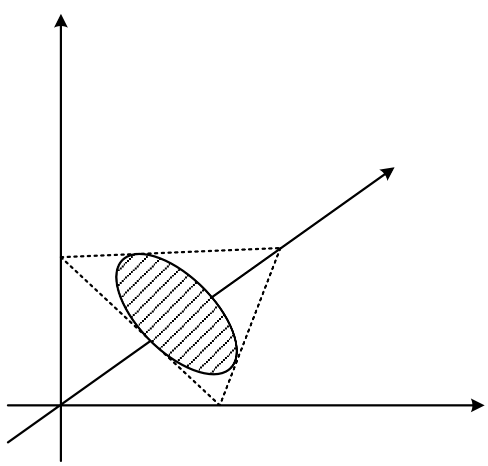

Figure 2.4. Interior and relative interior.

Theorem 2.1. *If there exists an interior point x of the feasible set $\mathcal{X}$ such that $\nabla f_0(x) = 0$, then x is the optimal of* (**??**)

What if there is no feasible point that satisfies the optimality condition, i.e., $\nabla f(x) \neq 0$, $\forall x \in \mathcal{X}$, where $\mathcal{X}$ is the feasible set. In this case, we can find the optimal point x^* by checking the condition of $(y - x^*)^T \nabla f(x) \leq 0$, $\forall y \in \mathcal{X}$. To formalize this case, a convex cone is introduced as in the following definition.

Definition 5 (Convex Cone). *K is a convex cone, if $x_1, x_2 \in K$, then $\lambda_1 x_1 + \lambda_2 x_2 \in K$, $\forall \lambda_1, \lambda_2 \geq 0$.*

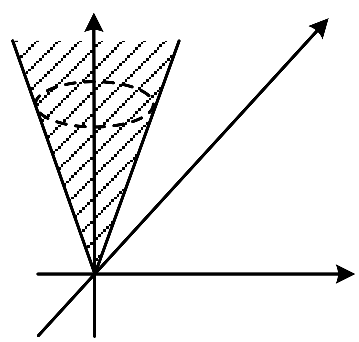

Figure 2.5. Example of a convex cone in 3D.

Definition 6 (Polar Cone). *Given a cone K, its polar cone, denoted by K°, is characterized by*

$$K^\circ = \{x : \langle x, v \rangle \leq 0,\ \forall v \in K\}. \tag{2.12}$$

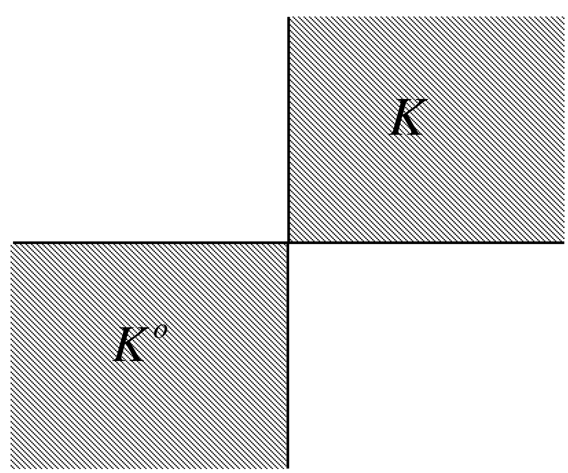

Figure 2.6. Example of a polar cone K° of a convex cone K

Example 3. $(\mathbb{R}^2_+)^\circ = \mathbb{R}^2_-$.

Exercise 1. *If S_n^+ is defined as*

$$S_n^+ = \{A \in \mathbb{R}^{n\times n} : A^T = A \text{ and all eigenvalues of } A \geq 0\},$$

then S_n^+ is a convex cone.

Proof:

$$A \in S_n^+ \Leftrightarrow x^T A x \geq 0,\ \forall x \in \mathbb{R}^n, \tag{2.13}$$

therefore if $A_1,\ A_2 \in S_n^+$,

$$\begin{aligned} x^T\left(\lambda_1 A_1 + \lambda_2 A_2\right)x &= \lambda_1 x^T A_1 x + \lambda_2 x^T A_2 x \\ &\geq 0, \forall x \end{aligned} \tag{2.14}$$

□

Exercise 2. *Prove the polar cone of S_n^+ is S_n^-, which is defined as*

$$S_n^- := \{A \in \mathbb{R}^{n\times n} : A^T = A \text{ and all eigenvalues of } A \leq 0\},$$

Hint: Prove the following two statements,

- $M \in S_n^- \Rightarrow \langle M, A\rangle \leq 0,\ \forall A \in S_n^+$
- $M \notin S_n^- \Rightarrow \exists A \in S_n^+$, s.t., $\langle M, A\rangle > 0$

Definition 7 (Tangent Cone). *For a non-empty convex set C and a point x, the set of all feasible directions is defined as*

$$F_C\left(x\right) = \{d : \exists \epsilon > 0 \text{ s.t. } x + \epsilon d \in C\}. \tag{2.15}$$

Then, the corresponding tangent cone of set C is defined as

$$T_C(x) = \text{closure}\left(F_C(x)\right). \tag{2.16}$$

Definition 8 (Normal Cone). *For a non-empty convex set C, the normal cone of C at point x is defined as*

$$N_C(x) = \left\{s :\ \langle s, y - x\rangle \leq 0,\ \forall y \in C\right\}. \tag{2.17}$$

Theorem 2.2. *Let $C \in \mathbb{R}^n$ be a nonempty, convex set, and let $x \in C$. Then the normal cone of C at x is the polar cone of the tangent cone of C at x. That is,*

$$N_C(x) = \left(T_C(x)\right)^{\circ}. \tag{2.18}$$

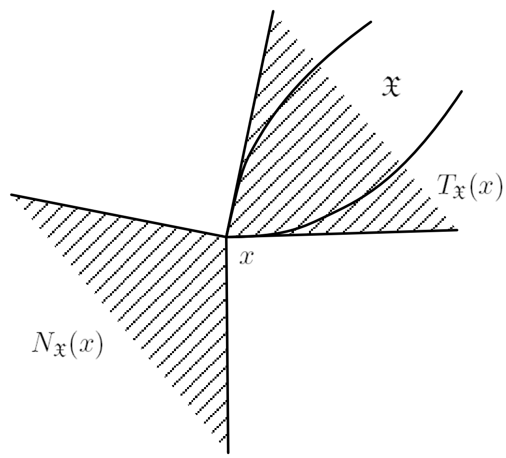

Figure 2.7. Example of tangent cone and normal cone.

Proof: Let $s \in N_C(x)$. Then for any $d \in F_C(x)$, there exists $\epsilon > 0$ such that $x + \epsilon d \in C$. Hence,

$$\langle s, d\rangle = \frac{1}{\epsilon}\langle s, x + \epsilon d - x\rangle \leq 0.$$

For any $\hat{d} \in T_C(x)$, there exists a sequence d_n such that $d_n \in F_C(x)$ and $d_n \to \hat{d}$. As $\langle s, d_n\rangle \leq 0$, we have $\langle s, \hat{d}\rangle \leq 0$, which means $s \in (T_C(x))^\circ$, i.e. $N_C(x) \subseteq (T_C(x))^\circ$.

On the other hand, let $s \in (T_C(x))^\circ$. Then $\langle s, d\rangle \leq 0$ for $\forall\ d \in T_C(x)$. For any $y \in C$, because C is convex, there exists $\hat{d} \in T_C(x)$ and $\alpha > 0$ such that $y = x + \alpha\hat{d}$. Hence, we have that

$$\langle s, y - x\rangle = \alpha\langle s, \hat{d}\rangle \leq 0.$$

Thus, $s \in N_c(x)$, i.e. $(T_C(x))^\circ \subseteq N_C(x)$. $\square$