Lecture 11: Convex Programming, Part III – February 27

Instructor: Niao He *Scribe: Shuanglong Wang*

Courtesy warning: These notes do not necessarily cover everything discussed in the class. Please email TA (swang157@illinois.edu) if you find any typos or mistakes.

In this lecture, we cover the following topics

- Some Remarks on Minimax Problem
- Some Remarks on Optimality Conditions
- Polynomial Solvability of Convex Programs

11.1 Minimax Problem

Recall in the last lecture, we have discussed

[Minimax Theorem] If X and Y are convex compact sets, $L(x,y), X \times Y \to \mathbf{R}$ is convex-concave and continuous, then $L(x,y)$ has a saddle point on $X \times Y$, and

$$\min_{x\in X}\max_{y\in Y} L(x,y) = \max_{y\in Y}\min_{x\in X} L(x,y)$$

Remark 1: Based on the proof analysis, we can immediate see that some of the assumptions can be relaxed. Here is a general result:

Theorem 11.1 *Let X and Y be convex and one of them is compact. Let $L(x,y)$ be lower-continuous (l.s.c.) and quasi-convex in $x \in X$. and upper semi-continuous(u.s.c) and quasi-concave in $y \in Y$. Then*

$$\min_{x\in X}\max_{y\in Y} L(x,y) = \max_{y\in Y}\min_{x\in X} L(x,y)$$

Remark 2: The compactness is sufficient bu not necessary, see for examples:

1. $\min_x \max_y (x+y) = \infty \neq -\infty = \max_y \min_x (x+y)$
2. $\min_x \max_{0\leq y\leq 1} (x+y) = -\infty = -\infty = \max_{0\leq y\leq 1} \min_x (x+y)$
3. $\min_x \max_{y\leq 1} (x+y) = -\infty = -\infty = \max_{y\leq 1} \min_x (x+y)$

Remark 3: Minimax problem could also arise from Fenchel dual:

Let $f(x)$ be closed and convex, then $f = f^{**}$. We can equivalently write f as

$$f(x) = \max_{y\in\mathbf{R}^n} \left\{ y^T x - f^*(x) \right\}$$

Hence

$$\min_{x\in X} f(x) = \min_{x\in X} \max_{y\in\mathbf{R}^n} \left\{ y^T x - f^*(x) \right\}$$

Remark 4: Alternative optimization does not necessarily converge to the saddle point.
Consider the saddle point problem

$$\min_{-1\le x\le 1} \max_{-1\le y\le 1} xy$$

The problem has a unique saddle point: $(0,0)$. Start with any (x_0, y_0), where $x_0 > 0$ and do alternative maximization over y and minimization over x:

$$(x_0, y_0) \Rightarrow (x_0, 1) \Rightarrow (-1, 1) \Rightarrow (-1, -1) \Rightarrow (1, -1) \Rightarrow (1, -1) \Rightarrow (1, 1) \Rightarrow ...$$

This will not converge to the saddle point.

11.2 Optimality Conditions for Convex Programs

- **General Constrained Differentiable Case:**

$$\begin{aligned} \min_{x\in X} \quad & f(x) \\ & g_i(x) \le 0, \quad i = 1, ..., m \qquad (P) \end{aligned}$$

We have already shown that for a feasible solution x^* to be optimal:

$$x_* \in X \text{ is optimal for (P)} \overset{\text{slater condition}}{\Longleftrightarrow} \exists \lambda^* \ge 0 \text{ s.t. } \begin{array}{ll} (a) & \nabla f(x_*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x_*) \in N_X(x_*) \\ (b) & \lambda_i^* g_i(x_*) = 0, \forall i = 1, \dots, m \end{array} \tag{11.1}$$

- **Simple Constrained Differentiable Case:**

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

Proposition 11.2 *Assume X is convex and $f(x)$ is convex and differentiable at x^*. Then*

$$x^* \in X \textit{ is optimal } \Leftrightarrow \nabla f(x^*) \in N_X(x^*), i.e.\ \nabla f(x^*)^T (x - x^*) \ge 0, \forall x \in X$$

Proof:

$(\Leftarrow)$ $\quad f(x) \ge f(x^*) + \nabla f(x^*)^T (x - x^*) \ge f(x^*), \forall x \in X$

$(\Rightarrow)$ $\quad \nabla f(x^*)(x - x^*) = \lim_{\epsilon\to 0} \frac{f(x^* + \epsilon(x - x^*)) - f(x^*)}{\epsilon} \ge 0$

■

Remark

- If $x^* \in \int(X)$, then x^* is optimal $\Leftrightarrow \nabla f(x^*) = 0$
- (Unconstrained case): If $X = \mathbf{R}^n$, then x^* is optimal $\Leftrightarrow \nabla f(x^*) = 0$
- For general (non-convex) optimization problems, $\nabla f(x^*) = 0$ is only a necessary but not sufficient condition for x^* to be optimal.

• <u>**Nondifferentiable Case:**</u>

Proposition 11.3 *Assume $f(x)$ is convex on X, and $x^* \in rint(dom(f))$, then*

$$x^* \in X \textit{ is optimal } \Leftrightarrow \exists g \in \partial f(x^*), \textit{ s.t. } g^T(x - x^*) \geq 0, \forall x \in X$$

Proof:

($\Leftarrow$) $f(x) \geq f(x^*) + g^T(x - x^*) \geq f(x^*), \forall x \in X$

($\Rightarrow$) $\max_{g \in \partial f(x^*)} g^T(x - x^*) = f'(x^*; x - x^*) = \lim_{\epsilon \to 0} \frac{f(x^* + \epsilon(x - x^*)) - f(x^*)}{\epsilon} \geq 0$

■

Remark In the unconstrained case when $X = \mathbf{R}^n$, x^* is optimal $\Leftrightarrow 0 \in \partial f(x^*)$

11.3 Solve Convex Programs

$$\begin{array}{lll} \min & f(x) & \\ \text{s.t.} & g_i(x) \leq 0, \quad i = 1, ..., m & (P) \\ & x \in X & \end{array}$$

In practice, to solve a problem means to find a ''approximate" solution to (P) with a small inaccuracy $\epsilon > 0$.

Measure of accuracy of an approximate solution $\hat{x}$: Some function $\epsilon(\hat{x})$ such that $\epsilon(\hat{x}) \geq 0$ and $\epsilon(\hat{x}) \to 0$ as $\hat{x} \to x^*$. For instance:

(i) $\epsilon(\hat{x}) = \inf_{x^* \in X_{opt}} \parallel \hat{x} - x^* \parallel$, where X_{opt} is the optimal set.

(ii) $\epsilon(\hat{x}) = f(\hat{x}) - Opt(P)$, where $Opt(P)$ is the optimal value.

(iii) $\epsilon(\hat{x}) = \max\left(f(\hat{x}) - Opt(P), \max_{1 \leq i \leq m} \left[g_i(\hat{x})\right]_+\right)$, where $u_+ = \max(u, 0)$

Black-box oriented numerical method: We assume the objective and constraints can be accessed through oracles.

- zero-order oracle: $0 = \big(f(x), g_1(x), ..., g_m(x)\big)$
- first-order oracle: $0 = \big(\partial f(x), \partial g_1(x), ..., \partial g_m(x)\big)$
- second-order oracle: $0 = \big(\nabla^2 f(x), \nabla^2 g_1(x), ..., \nabla^2 g_m(x)\big)$
- separation oracle: given x, either reports $x \in X$ or returns a separator, i.e. a vector $e \neq 0$, such that $e^T x \geq \sup_{y \in X} e^T y$. Note that when $x \notin \text{int}(X)$, a separator does exist.

Complexity of a numerical method M: Given an input $\epsilon > 0$, a problem instance P,

- *oracle complexity:* number of oracles required to solve the problem (P) up to accuracy $\epsilon > 0$
- *arithmetic complexity:* number of arithmetic operation (bit-wise operation) requirement to solve the problem (P) up to accuracy $\epsilon > 0$

A solution method M for a family P of problems is called *polynomial* if $\forall p \in P$, the arithmetic complexity

$$Compl_M(\epsilon, p) \leq O(1) \underbrace{\big[dim(P)\big]^\alpha}_{\text{polynomial of size}} \cdot \underbrace{\ln(V(P)/\epsilon)}_{\text{number of accuracy digits}}$$

where $V(P)$ is some data-dependent quantity.

We say the family P of problems *polynomially solvable*, i.e. *computationally tractable*, if it admits polynomial solution methods.

11.4 Convex Problems are Polynomially Solvable

Illustration: One-dimensional Case

$$\min_{x \in [a,b]} f(x)$$

- Zero-order line search: Assume there exists a unique minimizer
  - Initialize a localizer $G_1 = [a, b]$ for x^*
  - At each iteration, choose $a_t, b_t \in G_t$, update the localizer

$$G_{t+1} \leftarrow \begin{cases} [a, b_t] \cap G_t, & \text{if } f(a_t) \leq f(b_t) \\ [a_t, b] \cap G_t, & \text{if } f(a_t) > f(b_t) \end{cases}$$

  If we choose a_t, b_t that split $[a, b]$ into equal length, $|G_{t+1}| = \frac{2}{3}|G_t|$. We get linear convergence.

- First-order line search (bisection)

- Initialize a localizer $G_1 = [-R, R] \supset [a, b]$
- At each iteration, compute the midpoint c_t of $G_t = [a_t, b_t]$

  if $c_t \notin [a, b]$,

$$G_{t+1} = \begin{cases} [a_t, c_t], & \text{if } c_t > b \\ [c_t, b_t], & \text{if } c_t < a \end{cases}$$

  if $c_t \in [a, b]$ and $f'(c_t) \neq 0$

$$G_{t+1} = \begin{cases} [a_t, c_t], & \text{if } f'(c_t) > 0 \\ [c_t, b_t], & \text{if } f'(c_t) < 0 \end{cases}$$

  otherwise, this implies c_t is optimal

Note that $x^* \in G_t$ and $|G_{t+1}| = \frac{1}{2}|G_t|$, we get linear convergence.