IE 521 Convex Optimization

Lecture 10: Solving Convex Programs

Niao He

16th April 2019

Outline

Recap: Lagrange Duality

General convex program:

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

Lagrange dual program:

$$\begin{aligned} \max_{\lambda} \quad & \underline{L}(\lambda) := \inf_{x\in X} L(x,\lambda) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned} \qquad (D)$$

where $L(x,\lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$.

Recap: Optimality Conditions

KKT conditions:

$$x_* \in X \text{ is optimal for (P)}$$

$$\overset{\text{(slater)}}{\Longleftrightarrow} \exists \lambda^* \geq 0, \text{ s.t. } \begin{cases} \nabla f(x_*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x_*) \in N_X(x_*) \\ \lambda_i^* g_i(x_*) = 0, \forall i = 1, \ldots, m \end{cases}$$

Saddle point condition:

$$x_* \in X \text{ is optimal for (P)}$$

$$\overset{\text{(slater)}}{\Longleftrightarrow} \exists \lambda^* \geq 0, \text{ s.t. } (x^*, \lambda^*) \text{ is a saddle point of } L(x, \lambda)$$

Recap: Minimax Theorem

Theorem. (von Neumann, 1928) Assume

- X and Y be convex and compact,
- $L(x, y)$ is continuous, **convex-concave** on $X \times Y$.

Then $L(x, y)$ has a saddle point on $X \times Y$, and

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

Minimax Lemma

Lemma. Let $f_i(x), i = 1, ..., m$ be convex and continuous on a convex compact set X. Then

$$\min_{x\in X} \max_{1\leq i\leq m} f_i(x) = \min_{x\in X} \sum_{i=1}^{m} \lambda_i^* f_i(x)$$

for some $\lambda^* \in \Delta_m := \{\lambda \in \mathbb{R}^m : \lambda \geq 0, \sum_{i=1}^{m} \lambda_i = 1\}$.

Proof of Minimax Lemma

Consider the epigraph form of the problem $\min_{x\in X}\max_{1\le i\le m} f_i(x)$:

$$\begin{aligned}\min_{x,t}\quad & t\\ \text{s.t.}\quad & f_i(x)-t\le 0, i=1,...,m \qquad\qquad (P)\\ & (x,t)\in X_t = X\times\mathbb{R}.\end{aligned}$$

- The optimal value $t^*=\min_{x\in X}\max_{1\le i\le m} f_i(x)$ is finite.
- (P) satisfies Slater condition and is solvable.
- The Lagrange function is $L(x,t;\lambda)=t+\sum_{i=1}^m\lambda_i(f_i(x)-t)$.
- There exists $(x^*,t^*)\in X_t$ and $\lambda^*\ge 0$, such that

$$\begin{cases}\frac{\partial L}{\partial t}L(x^*,t^*;\lambda^*)=1-\sum_{i=1}^m\lambda_i^*=0\\ \sum_{i=1}^m\lambda_i^*(f_i(x^*)-t^*)=0\end{cases}\Rightarrow\begin{cases}\sum_{i=1}^m\lambda_i^*=1\\ \sum_{i=1}^m\lambda_i^* f_i(x^*)=t^*\end{cases}$$

Therefore, $\exists\lambda^*\in\Delta_m$ such that

$$\min_{x\in X}\max_{1\le i\le m} f_i(x)=t^*=\min_{(x,t)\in X_t} L(x,t;,\lambda^*)=\min_{x\in X}\sum_{i=1}^m\lambda_i^* f_i(x)$$

Proof of Minimax Theorem

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

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

Both (P) and (D) are solvable. Suffice to show $\mathrm{Opt}(D) \geq \mathrm{Opt}(P)$.

- Define the sets $X(y) = \{x \in X : L(x,y) \leq \mathrm{Opt}(D)\}$.
- If $\{X(y) : y \in Y\}$ intersect, then $\mathrm{Opt}(P) \leq \mathrm{Opt}(D)$.
- Note that for any $y \in Y$, $X(y)$ is <u>nonempty</u>, <u>compact</u> and <u>convex</u> (why?).
- By Helley's theorem, sufficient to show that every finite collection of these sets intersect.

Proof of Minimax Theorem (Continued)

$$X(y) = \{x \in X : L(x, y) \leq \mathrm{Opt}(D)\}.$$

- Suppose $\exists y_1, ..., y_m \in Y$, s.t. $X(y_1) \cap ... \cap X(y_m) = \emptyset$. Then

$$\begin{aligned}
\mathrm{Opt}(D) &< \min_{x \in X} \max_{i=1,\ldots,m} L(x, y_i) \\
&= \min_{x \in X} \sum_{i=1}^{m} \lambda_i^* L(x, y_i) && \text{(by Minimax Lemma)} \\
&\leq \min_{x \in X} L(x, \sum_{i=1}^{m} \lambda_i^* y_i) && \text{(by concavity of } L(x, \cdot)) \\
&= \underline{L}(\bar{y}) \leq \mathrm{Opt}(D)
\end{aligned}$$

Contradiction!

Sion's Minimax Theorem (1958)

Theorem. Assume

- X and Y are convex and one of them is compact,
- $L(x, y) : X \times Y \to \mathbb{R}$ is **lower semi-continuous** and **quasi-convex** on $x \in X$,
- $L(x, y) : X \times Y \to \mathbb{R}$ is **upper semi-continuous** and **quasi-concave** on $y \in Y$.

Then $L(x, y)$ has a saddle point on $X \times Y$, and

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

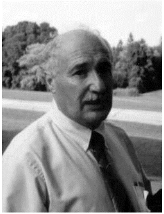
From Theory to Algorithm

Question: How to Solve Convex Programs?

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

History Note

- **1947:** Dantzig introduced the Simplex Method for LP
- **1950s-60s**: Simplex Method was successfully applied to many problems of large scale
- **1973:** Klee and Minty proved that Simplex Method is not a polynomial-time algorithm
- **1976-77:** Shor, Nemirovski and Yudin independently introduced the Ellipsoid method for convex programs
- **1979:** Khachiyan proved the poly-time solvability of LP



Naum Shor
(1937-2006)

Leonid Khachiyan
(1952-2005)

History Note (Continued)

- **1984:** Karmarkar introduced poly-time interior point method for LP
- **late-1980s:** Renegar & Gonzaga introduced path-following interior point method for LP
- **1988:** Nesterov and Nemirovski extended interior point method for convex programs
- **after 1990s:** many solvers for convex programs

mosek

CVX RESEARCH

Accuracy Measures

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

Goal: Find an ''approximate'' solution to (P) with a small inaccuracy $\epsilon > 0$.

Accuracy Measure: given $\hat{x}$, the accuracy measure $\epsilon(\hat{x})$ should satisfy:

- $\epsilon(\hat{x}) \geq 0$
- $\epsilon(\hat{x}) \to 0$ as $\hat{x} \to x^*$.

Examples of Accuracy Measure

- $\epsilon(\hat{x}) = \inf_{x^* \in X^*} \|\hat{x} - x^*\|_2$
- $\epsilon(\hat{x}) = f(\hat{x}) - \mathrm{Opt}(P)$, where $\hat{x}$ is feasible
- $\epsilon(\hat{x}) = \max\left(f(\hat{x}) - \mathrm{Opt}(P), \max_{1 \le i \le m} \left[g_i(\hat{x})\right]_+\right)$
- $\epsilon(\hat{x}) = f(\hat{x}) - \mathrm{Opt}(P) + \sum_{i=1}^m \rho_i [g_i(\hat{x})]_+$, where $\rho_i > 0$.

Black-box Oracles

Access the objective and constraints through oracles:

- Zero-order oracle:

$$\mathcal{O} = \big(f(x), g_1(x), ..., g_m(x)\big)$$

- First-order oracle:

$$\mathcal{O} = \big(\partial f(x), \partial g_1(x), ..., \partial g_m(x)\big)$$

- Second-order oracle:

$$\mathcal{O} = \big(\nabla^2 f(x), \nabla^2 g_1(x), ..., \nabla^2 g_m(x)\big)$$

- Separation oracle for X: given x, either reports $x \in X$ or returns a separator, i.e. a vector $a \neq 0$, such that

$$a^T x \geq \sup_{y \in X} a^T y.$$

Example

$$\min_x \quad f(x) := \max_{1\leq j\leq J} f_j(x)$$
$$\text{s.t.} \quad X := \{x : g_i(x) \leq 0, i = 1, ..., m\}$$

where $f_j(x)$ and $g_i(x)$ are convex and differentiable. Assume we can compute $f_j(x), \nabla f_j(x), \forall j$ and $g_i(x), \nabla g_i(x), \forall i$.

- First-order oracle for f:

$$\partial f(x) = \text{Conv}\left\{\nabla f_j(x) | f(x) = f_j(x)\right\}$$

- Separation oracle for X:

$$x \in X \Leftrightarrow g_i(x) \leq 0, \forall i = 1, ..., m$$
$$x \notin X \Leftrightarrow \exists i' \in \{1, \dots, m\}, \text{ s.t. } g_{i'}(x) > 0$$
$$\Rightarrow \nabla g_{i'}(x)^T(y - x) \leq g_{i'}(y) - g_{i'}(x) \leq 0, \forall y \in X$$
$$\Rightarrow \omega^T x \geq \sup_{y\in X} \omega^T y, \text{ for } \omega = \nabla g_{i'}(x)$$

Complexity

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 operations (bit-wise operations) required to solve the problem (P) up to accuracy $\epsilon > 0$

Polynomial Solvability

Definition. A solution method M for a family $\mathcal{P}$ of problems is called *polynomial* if $\forall P \in \mathcal{P}$, the arithmetic complexity

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

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

Definition. $\mathcal{P}$ is called *polynomially solvable* if it admits polynomial solution methods.

Illustration: Solving 1D Convex Problem

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

Zero-order line search:

- Initialize a localizer $G_1 = [a, b] \ni x^*$
- At iteration t, 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.

Illustration: Solving 1D Convex Problem

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

First-order line search (Bisection)

- Initialize a localizer $G_1 = [-R, R] \supset [a, b]$
- At iteration t, 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 $f'(c_t) \neq 0$

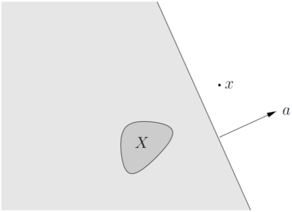
$$G_{t+1} = \begin{cases} [a_t, c_t], & \text{if } f'(c_t) > 0 \\ [c_t, b_t], & \text{if } f'(c_t) < 0 \end{cases}$$

  - otherwise, this implies c_t is optimal

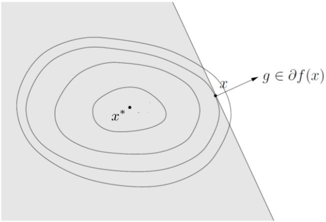
Note $x^* \in G_t$ and $|G_{t+1}| = \frac{1}{2}|G_t|$, we get linear convergence.

Cutting Plane Methods

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



(a) Separation oracle



(b) First-order oracle

(a) $X \subseteq \{y : a^T(y - x) \leq 0\}$ if $x \notin X$;
(b) $X^* \subseteq \{y : g^T(y - x) \leq 0\}$ if x is not optimal.

Figures from Boyd and Vandenberghe notes (2008)

Cutting Plane Methods

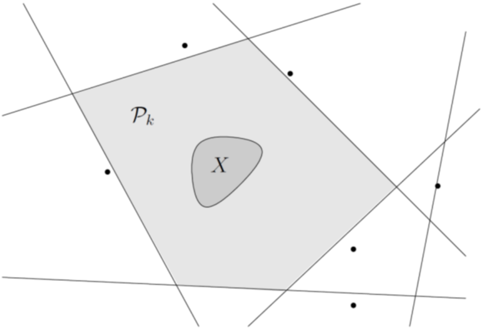


Figure: Localization Polyhedron

$$\mathcal{P}_1 \supseteq \cdots \supseteq \mathcal{P}_k \supseteq X$$

Q. How to choose the query point to cut the most off?

Figures from Boyd and Vandenberghe notes (2008)

Cutting Plane Methods

- *Center of gravity method*: choose the query to be the center of the gravity of $\mathcal{P}_k$.
- *Maximum volume ellipsoid cutting plane method*: choose the query to be the center of the maximum volume ellipsoid contained in $\mathcal{P}_k$.
- *Chebyshev center cutting-plane method*: choose the query point to be the Chebyshev center of $\mathcal{P}_k$, i.e., the center of the largest Euclidean ball that lies in $\mathcal{P}_k$.

References

- Ben-Tal & Nemirovski, Chapter 7