IE 521: Convex Optimization **Spring 2017, UIUC**

Lecture 12: Polynomial Solvability of Convex Programs – March 01

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

- Center of Gravity Method
- Ellipsoid Method

References: Bental & Nemirovski, Chapter 7

12.1 Convex Program

We aim to solve the convex optimization problem

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

where f is convex and $X \subseteq \mathbf{R}^n$ is closed, bounded convex with non-empty interior (also called a convex body).

Let r, R be such that $\{x : \| x - c \|_2 \le r\} \subseteq X \subseteq \{x : \| x \|_2 \le R\}$ for some c.

Setup: We assume that the objective and constraint set can be accessed through

- Separation Oracle for X: a routine that given an input x, either reports $x \in X$ or returns a vector $\omega \neq 0$, s.t. $\omega^T x \ge \sup_{y\in X} \omega^T y$
- First Order Oracle for f: a routine that given an input x, returns a subgradient $g \in \partial f(x)$, i.e. $f(y) \ge f(x) + g^T(y - x), \forall y$.
- Zero Order Oracle for f: a routine that given an input x, returns the function value $f(x)$.

Example:

$$\begin{aligned} \min_{x} \quad & \max_{1\le j\le J} f_j(x) \\ \text{s.t.} \quad & g_i(x) \le 0, \quad i = 1, ..., m \end{aligned}$$

where $f_j(x), 1 \le j \le J$ and $g_i(x), 1 \le i \le m$ are convex and differentiable. Here, we have

$$f(x) = \max_{1\le j\le J} f_j(x)$$

$$X = \{x \in \mathbf{R}^n : g_i(x) \le 0, i = 1, ..., m\}$$

1. We can build first order oracle for f if at a given x, we can compute $f_1(x), ..., f_J(x)$, and $\nabla f_1(x), ..., \nabla f_J(x)$. This is because:

$$\partial f(x) \supseteq \text{Conv}\{\partial f_j(x), j \text{ such that } f(x) = f_j(x)\}$$

2. We can build separation oracle for X if at given x, we can compute $g_1(x), ..., g_m(x)$ and $\nabla g_1(x), ..., \nabla g_m(x)$. This is because

$$x \in X \Leftrightarrow g_i(x) \leq 0, \forall i = 1, ..., m$$

and

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

12.2 Center of Gravity Method (Levin, 1965; Newman, 1965)

We consider the simple cutting plane scheme

- Initialize $G_0 = X$
- At iteration $t = 1, 2, ..., T$, do
  - Compute the center of gravity: $x_t = \frac{1}{\text{Vol}(G_{t-1})} \int_{x \in G_{t-1}} x dx$
  - Call the first order oracle and obtain $g_t \in \partial f(x_t)$
  - Set $G_t = G_{t-1} \cap \left\{ y : g_t^T(y - x_t) \leq 0 \right\}$
- Output $\hat{x}_T \in \arg\min_{x \in \{x_1, ..., x_T\}} f(x)$

Lemma 12.1 *Let C be a centered convex body in $\mathbf{R}^n$ with $\int_x x dx = 0$. Then $\forall \omega \neq 0$*

$$Vol(C \cap \left\{ x : \omega^T x \leq 0 \right\}) \leq \left(1 - (\frac{n}{n+1})^n\right) Vol(C) \leq (1 - \frac{1}{e}) Vol(C)$$

As an immediate result of the center of gravity method, we have

- $x^* \in G_t, \forall t \geq 1$ because $X_{opt} \subseteq \{y : f(y) \leq f(x_t)\} \subseteq \left\{ y : g_t^T(y - x_t) \leq 0 \right\}$
- $\text{Vol}(G_t) \leq (1 - \frac{1}{e})^t \text{Vol}(X)$

Theorem 12.2 *The approximate solution generated by the center of gravity methods satisfies:*

$$f(\hat{x}_T) - f^* \leq (1 - \frac{1}{e})^{\frac{T}{n}}\, Var_X(f)$$

where $Var_X(f) = \max_{x \in X} f(x) - \min_{x \in X} f(x)$*, and* f^* *is the optimal value.*

Proof: Let $\delta \in ((1 - \frac{1}{e})^{\frac{T}{n}}, 1)$, and consider the neighborhood of x^*

$$X_\delta = \{x^* + \delta(x - x^*) : x \in X\}$$

$$\mathrm{Vol}(X_\delta) = \delta^n \mathrm{Vol}(X) > (1 - \frac{1}{e})^T \mathrm{Vol}(X) \geq \mathrm{Vol}(G_T)$$

Hence $X_\delta / G_T \neq 0$. Let $y = x^* + \delta(z - x^*) \in X_\delta / G_T$ for some $z \in X$.
Thus, for certain $t^* \leq T$, we have $y \in G_{t^*-1}/G_{t^*}$. Since $y \notin G_{t^*}$, we have $g_{t^*}^T(y - x_{t^*}) > 0$, hence $f(y) > f(x_{t^*})$. Since $y = x^* + \delta(z - x^*)$, by convexity of f,

$$\begin{aligned} f(y) = f(\delta z + (1 - \delta)x^*) &\leq \delta f(z) + (1 - \delta) f(x^*) \\ &= f(x^*) + \delta[f(z) - f(x^*)] \\ &\leq f(x^*) + \delta \mathrm{Var}_X(f) \end{aligned}$$

Hence $f(\hat{x}_T) \leq f(x_{t^*}) \leq f^* + \delta \mathrm{Var}_X(f)$. Let $\delta \to (1 - \frac{1}{e})^{\frac{T}{n}}$, we got the desired result. ■

Remarks:

1. To obtain a solution with small inaccuracy $\epsilon > 0$, the number of oracles needed are polynomially dependent on the dimension.
   - separation oracle: $N(\epsilon) = n \log(\frac{\mathrm{Var}_X(f)}{\epsilon})$
   - first order oracle: $N(\epsilon) = n \log(\frac{\mathrm{Var}_X(f)}{\epsilon})$
   - zero order oracle: $N(\epsilon) = n \log(\frac{\mathrm{Var}_X(f)}{\epsilon})$
2. The rate of convergence of center of gravity methods is exponentially fast and this is usually called a linear rate.
3. The center of gravity is not necessarily polynomial and cannot be used as a computational tool because finding the center of gravity at each step can be extremely difficult, even for polytopes.
4. We can use ellipsoid as the localizer, so that is is easy to compute the center.

12.3 Ellipsoid Method (Shor, Nemirovsky, Yudin, 1970s)

Ellipsoid. Let Q be a symmetric positive definite matrix, and c be the center, an ellipsoid is uniquely characterized by (c, Q):

$$\begin{aligned} E(c, Q) &= \left\{x \in \mathbf{R}^n : (x - c)^T Q^{-1} (x - c) \leq 1\right\} \\ &= \left\{x = c + Q^{\frac{1}{2}} u : u^T u \leq 1\right\} \end{aligned}$$

The Volume of $E(c, Q)$ is

$$\mathrm{Vol}(E(c,Q)) = \mathrm{Det}(Q^{\frac{1}{2}})\mathrm{Vol}(B_n)$$

where B_n is a n-dimensional Euclidean ball with $\mathrm{Vol}(B_n) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$.

Proposition 12.3 *Let $E = E(c,Q) = \left\{x \in \mathbf{R}^n : (x-c)^T Q^{-1}(x-c) \leq 1\right\}$ be an ellipsoid. Let $H_+ = \left\{x : \omega^T x \leq \omega^T c\right\}$ be a half space with $\omega \neq 0$ that pass through the center c. Then the half ellipsoid $E \cap H_+$ can be contained by the ellipsoid $E^+ = E(c^+, Q^+)$ with*

$$c^+ = c - \frac{1}{n+1}q$$
$$Q^+ = \frac{n^2}{n^2+1}(Q - \frac{2}{n+1}qq^T)$$

where $q = \frac{Q\omega}{\sqrt{\omega^T Q \omega}}$ is the step from c to the boundary of E. Moreover,

$$Vol(E^+) \leq \exp\left\{-\frac{1}{2n}\right\} Vol(E)$$

The Ellipsoid Method. The algorithm works as follows:

- Initialize $E(c_0, Q_0)$ with $c_0 = 0, Q_0 = R^2 I$
- At iteration $t = 1, 2, ..., T$, do
  - Call separation oracle with the input c_{t-1}
  - If $c_{t-1} \notin X$, obtain a separator $u \neq 0$
  - If $c_{t-1} \in X$, call first order oracle and obtain a subgradient $\omega \in \partial f(c_t)$
  - Set the new ellipsoid $E(c_t, Q_t)$ with

$$c_t = c_{t-1} - \frac{1}{n+1}\frac{Q_{t-1}\omega}{\sqrt{\omega^T Q_{t-1}\omega}}$$

$$Q_t = \frac{n^2}{n^2-1}(Q_{t-1} - \frac{2}{n+1}\frac{Q_{t-1}\omega\omega^T Q_{t-1}}{\omega^T Q_{t-1}\omega})$$

- Output

$$\hat{x}_T = \arg\min_{c \in \{c_1,\ldots,c_T\}\cap X} f(c)$$

As a immediate result,

$$\mathrm{Vol}(E_t) \leq \exp\left\{-\frac{t}{2n}\right\}\mathrm{Vol}(E_0) = \exp\left\{-\frac{t}{2n}\right\}R^n \mathrm{Vol}(B_n)$$

Theorem 12.4 *The approximate solution generated by the Ellipsoid methods after T steps, for $T > 2n^2 \log(\frac{R}{r})$ satisfies:*

$$f(\hat{x}_T) - f^* \leq \frac{R}{r} Var_X(f) \exp\left\{-\frac{T}{2n^2}\right\}$$

Proof: Similar as the proof for the center of gravity method. Set $\delta \in (\frac{R}{r}\exp\left\{-\frac{T}{2n^2}\right\}, 1)$. Then $X_\delta = \{x^* + \delta(x - x^*) : x \in X\}$

$$\text{Vol}(X_\delta) = \delta^n \text{Vol}(X) \geq \delta^n \gamma^n \text{Vol}(B_n) > R^n \exp\left\{-\frac{T}{2n}\right\} \text{Vol}(B_n) \geq \text{Vol}(E_t)$$

Hence, $X_\delta / E_t \neq \emptyset$. The rest of the proof follows similarly as in Theorem 12.2. ■

Remark: The oracle complexity for Ellipsoid method is $O(n^2 \log(\frac{1}{\epsilon}))$, which is only slightly worse than the center of gravity method.

Remark: The Ellipsoid method works for any general convex problems as long as they admit separation and first order oracles. Moreover, the algorithm is polynomial if it takes only polynomial time to call those oracles. For instance, for linear programs with n variables and m constraints, it takes $O(nm)$ computation cost for the separation oracle.

In summary, here are some advantages and disadvantages of Ellipsoid method:

+ : universal
+ : simple to implement and steady for small size problems
+ : low order dependence on the number of functional constraints
− : quadratic growth on the size of problem, inefficient for large-scale problems.