IE 521 Convex Optimization

Lecture 11: Center of Gravity, Ellipsoid Method

Niao He

7th April 2019

Outline

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

Convergence

Given solutions $\{x_t\}$ and accuracy measure $\mathcal{E}(x_t)$

$$\lim_{t\to\infty} \frac{\mathcal{E}(x_{t+1})}{\mathcal{E}(x_t)^p} = q$$

- Linear convergence: $p = 1, q \in (0,1)$
  - E.g., $\mathcal{E}(x_t) = O(e^{-\alpha t})$, where $\alpha > 0$
- Sublinear convergence: $p = 1, q = 1$
  - E.g., $\mathcal{E}(x_t) = \frac{1}{t^\beta}$, where $\beta > 0$
- Superlinear convergence: $p = 1, q = 0$
  - E.g., $\mathcal{E}(x_t) = O(e^{-\alpha t^2})$, where $\alpha > 0$
- Convergence of order p: $p > 1, q > 0$
  - When $p = 2$, called quadratic convergence.
  - E.g., $\mathcal{E}(x_t) = O(e^{-\alpha p^t})$, where $\alpha > 0$

Illustration: Convergence

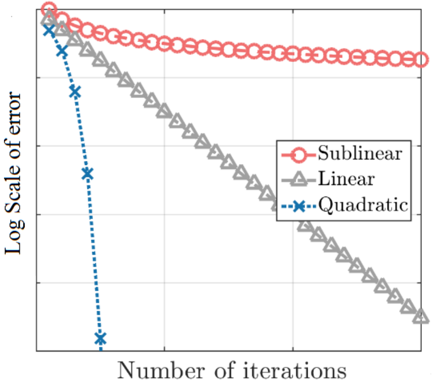


Figure: sublinear, linear, quadratic convergence

Solving Convex Program

We focus on the following general convex problem

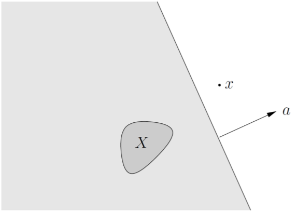
$$\min_{x \in X} \; f(x)$$

Problem Setting:

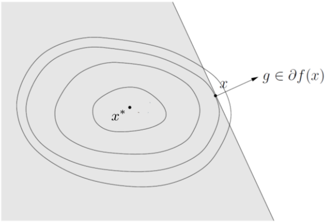
- f is convex and admits *zero- and first-order oracles*;
- $X \subset \mathbb{R}^n$ is a convex body (convex, compact, with nonempty interior) and admits *separation oracle*.

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

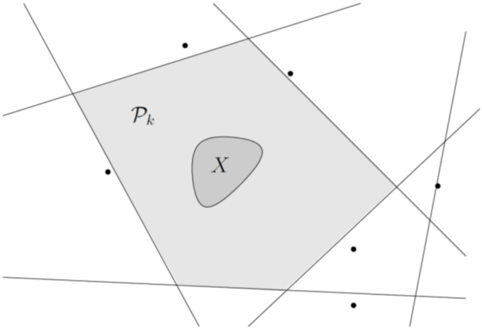


Figure: Localization Polyhedron

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

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 $\mathcal{P}_k$.

Center of Gravity Method

(Levin, 1965; Newman, 1965)

- Initialize $G_0 = X$
- At iteration $t = 1, 2, ..., T$, do
  - Compute the center of gravity: $x_t = \frac{1}{\mathrm{Vol}(G_{t-1})} \int_{x \in G_{t-1}} x dx$
  - 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 $\hat{x}_T \in \arg\min_{x \in \{x_1, ..., x_T\}} f(x)$

Center of Gravity Method

Lemma (Grünbaum [1960]). Let $C \subset \mathbb{R}^n$ be a convex body with $\int_x x dx = 0$. Then $\forall a \neq 0$

$$\begin{aligned}\mathrm{Vol}(C \cap \{x : a^T x \leq 0\}) &\leq \left(1 - \left(\frac{n}{n+1}\right)^n\right)\mathrm{Vol}(C) \\ &\leq \left(1 - \frac{1}{e}\right)\mathrm{Vol}(C) \approx 0.63\mathrm{Vol}(C)\end{aligned}$$

Remark. It follows that

$$\mathrm{Vol}(G_t) \leq \left(1 - \frac{1}{e}\right)^t \mathrm{Vol}(X), t \geq 1$$

Convergence of Center of Gravity Method

Theorem. The center of gravity method return $\hat{x}_T \in X$:

$$f(\hat{x}_T) - f^* \leq \left(1 - \frac{1}{e}\right)^{\frac{T}{n}} \cdot \mathrm{Var}_X(f)$$

where $\mathrm{Var}_X(f) = \max_{x \in X} f(x) - \min_{x \in X} f(x)$.

- Linear convergence rate
- Oracle complexity: $N(\epsilon) = \mathcal{O}\left(n \log(\frac{\mathrm{Var}_X(f)}{\epsilon})\right)$
- Main disadvantage: computing the center of gravity is extremely difficult, even for polytopes.

Proof of Convergence

- Note $x^* \in G_t, \forall t \geq 1$ and $\text{Vol}(G_t) \leq (1 - \frac{1}{e})^t \text{Vol}(X)$.
- Consider the neighborhood of x^*:
  $X_\delta = \{x^* + \delta(x - x^*) : x \in X\}$, where $\delta \in ((1 - \frac{1}{e})^{\frac{T}{n}}, 1)$.
- Observe that $X_\delta / G_T \neq 0$.

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

- 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 $g_{t^*}^T (y - x_{t^*}) > 0$, so $f(y) > f(x_{t^*})$.
- Since $y = x^* + \delta(z - x^*)$, by convexity of f,

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

  Hence $f(\hat{x}_T) \leq f(x_{t^*}) \leq f^* + \delta \text{Var}_X(f)$.

Ellipsoid as Localizer

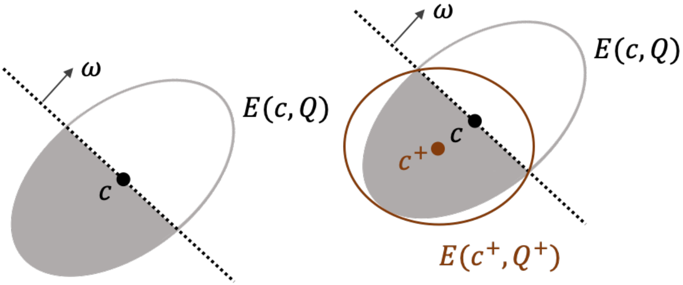
Definition. Let $Q \succ 0$ be symmetric, and c be the center, an *ellipsoid* is uniquely characterized by (c, Q):

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

- $\mathrm{Vol}(E(c, Q)) = \mathrm{Det}(Q^{\frac{1}{2}})\mathrm{Vol}(B_n)$, where B_n is a unit Euclidean ball in $\mathbb{R}^n$ with $\mathrm{Vol}(B_n) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$.

Half Ellipsoid

Let $H_+ = \{x : \omega^T x \leq \omega^T c\}$ be a half space with $\omega \neq 0$ that pass through the center c of the ellipsoid $E(c, Q)$.



Half 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 of the ellipsoid $E(c, Q)$.

- $E \cap H_+ \subseteq E^+ = E(c^+, Q^+)$ with

$$c^+ = c - \frac{1}{n+1} q, \text{ where } q = \frac{Q\omega}{\sqrt{\omega^T Q \omega}},$$

$$Q^+ = \frac{n^2}{n^2+1}\left(Q - \frac{2}{n+1} q q^T\right).$$

- Volume decrease:

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

Ellipsoid Method

(Shor; Nemirovsky, Yudin, 1970s)

- 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 $c_{t-1} \notin X$, call separation oracle and obtain $\omega \neq 0$
  - If $c_{t-1} \in X$, call first order oracle and obtain $\omega \in \partial f(c_t)$
  - Set the new ellipsoid $E(c_t, Q_t)$ with

$$\begin{aligned} c_t &= c_{t-1} - \frac{1}{n+1} \frac{Q_{t-1}\omega}{\sqrt{\omega^T Q_{t-1}\omega}} \\ Q_t &= \frac{n^2}{n^2-1}\left(Q_{t-1} - \frac{2}{n+1} \frac{Q_{t-1}\omega\omega^T Q_{t-1}}{\omega^T Q_{t-1}\omega}\right) \end{aligned}$$

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

Illustration of Ellipsoid Method

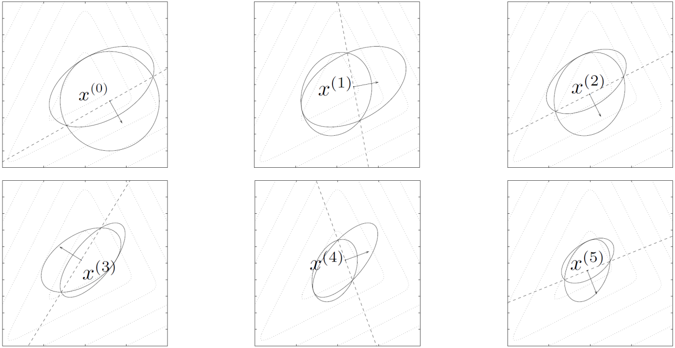


Figure: Illustration

Figure from Boyd, EE364b lecture notes

Convergence of Ellipsoid Method

Theorem. Assume $B(\bar{x}, r^2 I) \subseteq X \subseteq B(0, R^2 I)$. The Ellipsoid method after T steps satisfies:

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

- Linear convergence rate
- Oracle complexity: $N(\epsilon) = \mathcal{O}\left(n^2 \log(\frac{\mathrm{Var}_X(f)}{\epsilon})\right)$
- Modest per iteration computation cost: $O(n^2)$
- Polynomial solvability: as long as it takes polynomial time to call the separation and first-order oracles

Proof of Convergence

- Similar as the proof for the center of gravity method.
- Consider the neighborhood of x^*: $X_\delta = \{x^* + \delta(x - x^*) : x \in X\}$, $\delta \in (\frac{R}{r} \exp\left\{-\frac{T}{2n^2}\right\}, 1)$.
- Note $X_\delta / E(c_T, Q_T) \neq \emptyset$, because

$$\begin{aligned}
\text{Vol}(X_\delta) &= \delta^n \text{Vol}(X) \\
&\geq \delta^n r^n \text{Vol}(B_n) \\
&> R^n \exp\left\{-\frac{T}{2n}\right\} \text{Vol}(B_n) \\
&\geq \text{Vol}(E(c_T, Q_T))
\end{aligned}$$

- Rest is the same as proof of center of gravity method.

Stopping Criterion

- In practice, f^* is often unknown and it is impossible to compute $f(x_t) - f^*$.
- Construct online lower bounds for f^*: $\ell_t \leq f^*$

$$\begin{aligned} f^* &\geq f(x_t) + \omega_t^T(x^* - x_t), \qquad \omega_t \in \partial f(x_t) \\ &\geq f(x_t) + \inf_{x \in E(c_t, Q_t)} \omega_t^T(x - x_t) \\ &= f(x_t) - \sqrt{\omega_t Q_t \omega_t} \end{aligned}$$

- Hence, $\sqrt{\omega_t Q_t \omega_t} \leq \epsilon \Longrightarrow f(x_t) - f^* \leq \epsilon$
- Tighter lower bound:

$$\ell_t = \max_{1 \leq \tau \leq t} \left(f(x_\tau) - \sqrt{\omega_\tau Q_\tau \omega_\tau} \right)$$

Discussion

Advantages and disadvantages of the Ellipsoid method:

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

Experiment on SVM

$$\min_{w,b} \quad \frac{1}{m}\sum_{i=1}^{m} \max(1 - y_i(w^T x_i + b), 0) + \lambda\|w\|_2^2$$

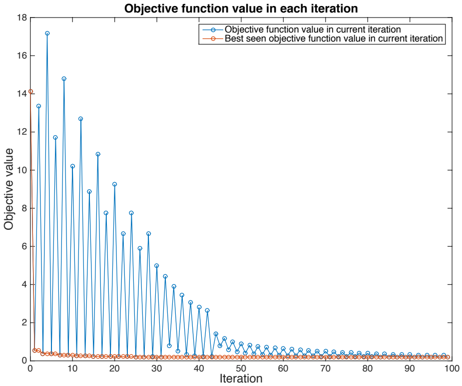


Figure: Ellipsoid Method for SVM on WBDC dataset (n=30)

References

- Ben-Tal & Nemirovski, Chapter 7