IE 521 Convex Optimization

Lecture 18: Interior Point Method

Path Following Scheme

Niao He

16th April 2019

Outline

Revisit Path Following Scheme

$$\min_x \quad c^T x \qquad \text{s.t.} \quad x \in X \tag{P}$$

Barrier Method: solve a series of unconstrained problems

$$x^*(t) := \operatorname*{argmin}_x \ \{\underbrace{t \cdot c^T x + F(x)}_{F_t(x)}\} \quad (t > 0) \tag{P_t}$$

Barrier Function:

- $F : \text{int}(X) \to \mathbb{R}$ and $F(x) \to +\infty$ as $x \to \partial(X)$
- F is twice continuously differentiable and convex
- F is *non-degenerate*, i.e. $\nabla^2 F(x) \succ 0, \forall x \in \text{int}(X)$

Central Path:

$$x^*(t) \in \text{int}(X) \longrightarrow x^*, \text{as } t \longrightarrow \infty$$

Revisit Path Following Scheme

Question: Need to specify

1. the barrier function $F(x)$?
   - Standard self-concordant function and $\mathrm{cl}(\mathrm{dom}(F)) = X$.
2. the method to solve unconstrained problems (P_t)?
   - (Damped) Newton's method

$$x_{k+1} = x_k - \frac{1}{1 + \lambda_{F_t}(x_k)} [\nabla^2 F_t(x_k)]^{-1} \nabla F_t(x_k)$$

   - Local quadratic convergence when $\lambda_{F_t}(x) \leq \frac{1}{4}$.
3. the policy to update the penalty parameter t?
   - when increasing $t \to t'$, we would like to preserve $\lambda_{F_{t'}}(x^*(t)) \leq \frac{1}{4}$ and make t' as large as possible

Policy for Penalty Update

- By definition of Newton's decrement,

$$\begin{aligned}\lambda_{F_{t'}}(x^*(t)) &= \|\nabla F_{t'}(x^*(t))\|_{x^*(t),*} \\ &= \|t' \cdot c + \nabla F(x^*(t))\|_{x^*(t),*}\end{aligned}$$

- From optimality condition of (P_t): $\forall t > 0$,

$$t \cdot c + \nabla F(x^*(t)) = 0$$

$$\begin{aligned}\lambda_{F_{t'}}(x^*(t)) &= \|(t' - t)c + tc + \nabla F(x^*(t))\|_{x^*(t),*} \\ &= \|(t' - t)c\|_{x^*(t),*} \\ &= \|(\frac{t'}{t} - 1)\nabla F(x^*(t))\|_{x^*(t),*} \\ &= (\frac{t'}{t} - 1) \cdot \lambda_F(x^*(t))\end{aligned}$$

$$\lambda_{F_{t'}}(x^*(t)) \leq \frac{1}{4} \Rightarrow t' \leq t\left(1 + \frac{1}{4\lambda_F(x^*(t))}\right)$$

Remark

- To ensure $t \to +\infty$, need $\lambda_F(x)$ to be uniformly bounded from above, namely,

$$\lambda_F^2(x) = \nabla F(x)^T [\nabla^2 F(x)]^{-1} \nabla F(x) \leq \nu$$

for some $\nu > 0$.

- This leads to the definition of self-concordant barriers.

Self-concordant Barrier

Definition. F is a <u>*ν-self-concordant barrier*</u> (ν-s.c.b.) for the set $X = \mathrm{cl}(\mathrm{dom}(F))$ if

- F is standard self-concordant, i.e. $\forall x \in \mathrm{dom}(f), h \in \mathbb{R}^n$:

$$|D^3F(x)[h,h,h]| \leq 2(D^2F(x)[h,h])^{3/2} \qquad (\star)$$

- F also satisfies

$$|DF(x)[h]| \leq \nu^{1/2}\sqrt{D^2F(x)[h,h]} \qquad (\star\star)$$

Remarks

- The following are equivalent:

$$(\star\star) \iff \nabla^2 F(x) \succcurlyeq \frac{1}{\nu}\nabla F(x)[\nabla F(x)]^T$$

- When F is non-degenerate,

$$(\star\star) \iff \lambda_F^2(x) = \nabla F(x)^T[\nabla^2 F(x)]^{-1}\nabla F(x) \le \nu$$

- Lipschitz continuity:

$$(\star\star) \Rightarrow |\nabla F(x)^T h| \le \nu \|h\|_x^2,$$

i.e. F is Lipschitz continuous w.r.t. the local norm.

Questions

Are the following self-concordant functions also s.c.b.?

- Linear function: $f(x) = a^T x + c (a \neq 0)$
- Quadratic function: $f(x) = \frac{1}{2} x^T Q x + q^T x + c,\ Q \succ 0$
- Logarithmic function: $f(x) = -\ln x\ (x > 0)$

Example: Logarithmic Quadratic Function

Example . Recall that the function below is standard self-concordant

$$f(x) = -\ln(q(x)) := \ln\left(-\frac{1}{2}x^TQx + b^Tx + c\right), Q \succeq 0.$$

- $Df(x)[h] = -\frac{1}{q(x)}(b^Th - x^TQh) := \omega_1$
- $D^2f(x)[h,h] = \frac{1}{q^2(x)}(b^Th - x^TQh)^2 + \frac{1}{q(x)}h^TQh := \omega_1^2 + \omega_2$
- $D^3f(x)[h,h,h] = 2\omega_1^3 + 3\omega_1\omega_2$

Q. Is it also a self-concordant barrier?

Operators Preserving Self-concordant Barriers

1. If $F(x)$ is a ν-s.c.b., then
$$\tilde{F}(y) = F(Ay + b) \text{ is a } \nu\text{-s.c.b.}$$
2. If $F_i(x)$ is ν_i-s.c.b., $i = 1, 2$, then
$$\tilde{F}(x) = F_1(x) + F_2(x) \text{ is a } (\nu_1 + \nu_2)\text{-s.c.b.}$$
3. If $F(x)$ is a ν-s.c.b. and $\beta \geq 1$, then
$$\tilde{F}(x) = \beta F(x) \text{ is a } (\beta\nu)\text{-s.c.b.}$$
4. If $F_i(x)$ is a ν_i-s.c.b. and $\beta_i \geq 1$, then
$$\tilde{F}(x) = \sum_i \beta_i F(x) \text{ is a } (\sum_i \beta_i \nu_i)\text{-s.c.b.}$$

Example of Self-concordant Barriers

Example . The function

$$F(x) = -\sum_{i=1}^{m} \ln(b_i - a_i^T x)$$

is a m-self-concordant barrier for the set $\{x : Ax \leq b\}$.

Remark. For any closed convex set $X \subseteq \mathbb{R}^n$ with non-empty interior, there exists a (βn)-self-concordant barrier for X.

Properties of Self-concordant Barriers

Lemma. Let F be a ν-self-concordant barrier for X. Then for any $x \in \text{int}(X), y \in X$, we have

$$\langle \nabla F(x), y - x \rangle \leq \nu.$$

Proof. Consider the function

$$\phi(t) = \langle F'(x + t(y - x)), y - x \rangle.$$

- Note that $\phi'(t) \geq \frac{1}{\nu}\phi(t)$
- It follows that $-\frac{1}{\phi(t)} + \frac{1}{\phi(0)} \geq \frac{t}{\nu}$

Performance Bound on Central Path

Consider the central path

$$x^*(t) = \arg\min_x \left\{ F_t(x) := t \cdot c^T x + F(x) \right\}$$

Theorem. For any $t > 0$, we have

$$c^T x^*(t) - \min_{x \in X} c^T x \leq \frac{\nu}{t}$$

Proof. This is because

$$c^T x^*(t) - c^T x = -t^{-1} \nabla F(x^*(t))^T (x^*(t) - x) \leq \frac{\nu}{t}$$

Performance Bound on Approximate Central Path

Consider an approximate solution $\hat{x}$ that is close to $x^*(t)$:

$$\lambda_{F_t}(\hat{x}) \leq \beta, \text{ where } \beta \text{ is small enough.}$$

Theorem. If $\lambda_{F_t}(\hat{x}) \leq \beta$,

$$c^T\hat{x} - \min_{x\in X} c^T x \leq \frac{1}{t}(\nu + \frac{\sqrt{\nu}\beta}{1-\beta})$$

Proof. This is because

$$\begin{aligned} c^T\hat{x} - c^T x*(t) &\leq \|c\|_{x^*(t),*} \cdot \|x - x^*(t)\|_{x^*(t)} \\ &= t^{-1}\|\nabla F(x^*(t))\|_{x^*(t),*} \cdot \|x - x^*(t)\|_{x^*(t)} \\ &\leq \frac{\sqrt{\nu}}{t} \cdot \frac{\lambda_{F_t}(x)}{1-\lambda_{F_t}(x)} \\ &\leq \frac{\sqrt{\nu}}{t}\frac{\beta}{1-\beta} \end{aligned}$$

Basic Path Following Scheme

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & x \in X \end{aligned} \qquad \text{(P)}$$

---

0. Initialize (x_0, t_0) with $t_0 > 0$ and $\lambda_{F_{t_0}}(x_0) \le \beta \in (0,1)$
1. For $k \ge 0$, do

$$\begin{aligned} t_{k+1} &= t_k\left(1 + \frac{\gamma}{\sqrt{\nu}}\right) \\ x_{k+1} &= x_k - [\nabla^2 F(x_k)]^{-1}[t_{k+1}c + \nabla F(x_k)] \end{aligned}$$

Convergence and Complexity

Theorem. In the above scheme, one has

$$c^T x_k - \min_{x \in X} c^T x \leq O(1)\frac{\nu}{t_0} \exp\left\{-O(1)\frac{k}{\sqrt{\nu}}\right\}$$

where the constant factor $O(1)$ depends solely on β and γ.

Remark.

- The number of Newton steps needed:

$$N(\epsilon) \leq O\left(\sqrt{\nu}\log\frac{\nu}{\epsilon}\right)$$

Two-phase Path Following Scheme

Q. How to find an initial pair (x_0, t_0) such that

$$\lambda_{F_{t_0}}(x_0) \leq \beta \in (0, 1)?$$

Option I: Damped Newton Method for Initialization

---

0. Choose $y_0 \in X$ and $t_0 = 1$
1. For $k \geq 0$, do

$$y_{k+1} = y_k - [\nabla^2 F(y_k)]^{-1}[t_0 c + \nabla F(y_k)]$$

2. Stop if $\lambda_{F_{t_0}}(y_k) \leq \beta$.

Two-phase Path Following Scheme

Option II: Auxiliary Path Following for Initialization

- Let $\hat{x} \in \mathrm{dom}(F)$. Consider the auxiliary path
$$y^*(t) = \arg\min_x[-t\nabla F(\hat{x})^T x + F(x)]$$
- When $t = 1$, $y^*(1) = \hat{x}$.
- When $t \to 0$, $y^*(t) = x_F := \arg\min_x F(x)$.
- We can trace $y^*(t)$ as t decreases from 1 to 0, until we approach to a point (y_0, t_0) such that $\lambda_{F_{t_0}}(x_0) \leq \beta$.

---

0. Choose $y_0 \in X$ and $t_0 = 1$
1. For $k \geq 0$, do
$$t_{k+1} = t_k(1 - \frac{\gamma}{\sqrt{\nu}})$$
$$y_{k+1} = y_k - [\nabla^2 F(y_k)]^{-1}[-t_{k+1}\nabla F(y_0) + \nabla F(y_k)]$$
3. Stop if $\lambda_{F_{t_k}}(y_k) \leq \beta$.

Concluding Remarks

- Number of damped Newton steps for initialization phase:

$$N_{\text{init}} \leq O\left(\sqrt{\nu}\log\nu\right)$$

- Number of damped Newton steps for main phase:

$$N_{\text{main}} \leq O\left(\sqrt{\nu}\log\frac{\nu}{\epsilon}\right)$$

- The total arithmetic cost of finding an ϵ-solution:

$$O\left(\mathcal{M}\sqrt{\nu}\log\left(\frac{\nu}{\epsilon}+1\right)\right)$$

  where $\mathcal{M}$ is the arithmetic cost for computing $\nabla F(x)$, $\nabla^2 F(x)$ and solving a Newton system.

- The algorithm is poly-time if $\mathcal{M}$ is polynomial.

References

- Nesterov (2004), Introductory Lectures on Convex Optimization, Chapter 4.1.4-5
- Nemirovski (2004), Interior Point Polynomial Time Methods in Convex Programming, Chapter 3-4