EE C227C Convex Optimization and Approximation	Lecture 22 - 03/17/2022

Lecture 28: Central Path

Lecturer: Jiantao Jiao *Scribe: Rahul Jain*

In this lecture, we will go over the central path following method and the runtime complexity. We will also go over the preliminaries of Primal-Dual conic problems.

1 Path Following

The Path Following algorithm is used to solve standard constrained minimization problems of the following.

$$\begin{aligned} &\min\nolimits_x \langle c, x\rangle \\ &\text{s.t } x \in \boldsymbol{Q} \end{aligned} \tag{1}$$

Where $\boldsymbol{Q}$ is a bounded closed convex set and we have a v-self-concordant barrier F on $\boldsymbol{Q}$. We will refer to the global minimum of (1) as c^*.

2 Central Path Method

We will solve (1) by tracing the "central path" as discussed in the previous lecture.

$$\begin{aligned} x^*(t) &= \underset{x\in domF}{\arg\min} f(t;x) \\ f(t;x) &= t\langle c, x\rangle + F(x),\ t \geq 0 \end{aligned} \tag{2}$$

We refer to $x^*(t)$ as the "central path". If we take the gradient of $f(t;x)$ and evaluate it at $x^*(t)$, it should be equal to 0, as per the first-order optimality conditions [1]. Thus, we get

$$\nabla_{x^*(t)} f(t;x) = tc + \nabla F(x^*(t)) = 0$$

As we can we can see from before, $x^*(0) = \underset{x\in domF}{\arg\min}\ F(x)$. We will refer to $x^*(0) = x^*_F$ as the "center" of the central path. In order to update/trace the central path (2), we will use the "Approximate Centering Condition" to do so. This is defined as

$$\lambda_{f(t,.}(x) = \|f'(t;x)\|^*_x = \|tc + \nabla F(x)\|^*_x \leq \beta \tag{3}$$

The β parameter is defined as the "centering parameter" as defined by Nesterov, and tends to be between $[0, \frac{1}{2}]$. Now we want to show that tracing the central path will result in a "close" approximation to c^*.

Theorem 1. *5.3.10 (Nesterov) [1]*
For any $t \geq 0$,

$$\langle c, x^*(t)\rangle - c^* \leq \frac{v}{t} \tag{4}$$

More generally, if x satisfies (3), then we have

$$\langle c, x\rangle - c^* \leq \frac{1}{t}\left(v + \frac{(\beta + \sqrt{v})\beta}{1-\beta}\right) \tag{5}$$

When $\beta = 0$, (5) simplifies to (4).

An observation is that we often want F to have a small v-value. We also want to ensure that β does not change too much from one value of t to the next value of t.

Here we present two lemmas that presents how to increment t in an appropriate manner and establish that the growth in t is subsantial (linear).

Lemma 2. *5.2.2 (Nesterov) Let $\tau \in [0, 1/2]$. If x satisfies (3), then if*

$$|\gamma| \leq \tau - \tau^2 \left(1 + \tau + \frac{\tau}{1 + \tau + \tau^2}\right)$$

then

$$\|t_t c + \nabla F(x_t)\|^*_{x_t} \leq \beta \tag{6}$$

Where $t_t = t + \frac{\gamma}{\|c\|^*_x}$ and $x_t = x - \frac{1}{1+\epsilon}[\nabla^2 F(x)]^{-1}(t_t c + \nabla F(x_t))$ and $\epsilon = \frac{\lambda^2}{1+\lambda}$, $\lambda = \|t_t c + \nabla F(x)\|^*_x$

Lemma 3. *5.3.2 (Nesterov)*

$$\|c\|^*_x \leq \frac{1}{t}\left(\beta + \sqrt{v}\right)$$

From these two lemmas we can derive the path following algorithm.

3 Algorithm

1. Set $t_0 = 0$. Choose $\epsilon \geq 0, x_0 \in domF$ such that $\|\nabla F(x_0)\|^*_{x_0} \leq \beta$
2. For $k \geq 0$
   - $t_{k+1} = t_k + \frac{\gamma}{\|c\|^*_{x_k}}$
   - $x_{k+1} = x_k - \frac{1}{1+\epsilon_k}[\nabla^2 F(x_k)]^{-1}(t_{k+1}c + \nabla F(x_k))$

   Where $\lambda_k = \|t_{k+1}c + \nabla F(x_k)\|^*_{x_k}$ and $\epsilon_k = \frac{\lambda_k^2}{1+\lambda_k}$
3. Terminate when $t_k \geq \frac{1}{\epsilon}\left(v + \frac{(\beta+\sqrt{v})\beta}{1-\beta}\right)$

It is not clear how to obtain x_0. We will go back to this later. Keep in mind that the termination condition is not dependent on the number of iterations k, but the value t_k, which doubles roughly (very fast!)

This begs the question, how many iterations do we need?

Theorem 4. *5.3.11 (Nesterov) [1] The path following algorithm terminates after N steps where*

$$N \geq \mathcal{O}\left(\sqrt{v} * \log\left(\frac{v\|c\|^*_{x^*_F}}{\epsilon}\right)\right)$$

Proof Let $r_0 = \|x_0 - x^*_F\|^*_{x_0} \leq \frac{\beta}{1-\beta}$. Using the same terminology from Lemma 2, we get

$$\frac{\gamma}{t_1} = \|c\|^*_{x_0} \leq \frac{1}{1-r_0}\|c\|^*_{x^*_F} \leq \frac{1-\beta}{1-2\beta}\|c\|^*_{x^*_F}$$

Thus

$$t_1 \geq \frac{\gamma(1-2\beta)}{(1-\beta)\|c\|^*_{x^*_F}}$$

More generally,

$$t_k \geq \frac{\gamma(1-2\beta)}{(1-\beta)\|c\|^*_{x^*_F}} \left(1 + \frac{\gamma}{\beta + \sqrt{v}}\right)^{k-1}$$

□

An important thing to note is that the first term within the $\mathcal{O}$ is the dominant term while second term in the bound tends to be very small. This still begs the question: How do we get the initial point x_0? We could use Dampened Newton's method to find x_0 but it is too slow so we will use another "path following" method, but in reverse to find x_0.

4 Auxillary Path Following

Let us define $y^*(t) = \arg\min_{y \in domF}[-t\langle\nabla F(y_0), y\rangle + F(y)]$ We start with $t_0 = 1$, and then we decrease. If we take the gradient of $y^*(t)$ and evaluate at y_0, we will get $\nabla_y y^*(t)|_{y=y_0} = -\nabla F(y_0) + \nabla F(y_0) = 0$. Thus it is trivial to get the first point y_0 in this auxillary path following scheme. Our resulting algorithm will resemble the path following method in reverse to get the initial point x_0 for the original problem.[1]

4.1 Scheme

1. $y_0 \in domF$, $t_0 = 1$
2. For $k \geq 0$
   - $t_{k+1} = t_k + \frac{\gamma}{\|\nabla F(y_0)\|^*_{y_k}}$
   - $y_{k+1} = y_k - \frac{1}{1+\epsilon_k}[\nabla^2 F(y_k)]^{-1}(-t_{k+1}\nabla F(y_0) + \nabla F(y_k))$

   Where $\lambda_k = \|t_{k+1}\nabla F(y_0) - \nabla F(y_k)\|^*_{y_k}$ and $\epsilon_k = \frac{\lambda_k^2}{1+\lambda_k}$
3. Terminate when $\|\nabla F(y_k)\|^*_{y_k} \leq \tau$. Set $\epsilon_k = \frac{\lambda_F(y_k)^2}{1+\lambda_F(y_k)}$ and $\bar{x} = y_k - \frac{1}{1+\epsilon_k}[\nabla^2 F(y_k)]^{-1}(\nabla F(y_k))$

5 Primal-Dual Conic Problems

Definition 5. *A set $K \subset \mathbb{R}^n$ is a convex cone if K is a nonempty convex set with the property that $x \in K, t \geq 0 \Rightarrow tx \in K$*

Definition 6. *A convex cone K is pointed if K does not contain lines (only rays)*

5.1 Dual Cones

Definition 7. $K^* = \{s \in \mathbb{R}^n \mid \langle s, x\rangle \geq 0 \ \forall x \in K\}$

Let us go through an example of finding the dual cone of $K = \{x \in R^n \mid x > 0\}$ (the nonnegative orthant cone). We can see that by simply plugging in $e_1, e_2, ...e_n$, where e_i is the i^{th} basis vector, into the definition of the dual cone of K, we will get $K^* = \{x \in R^n \mid x > 0\} = K$. Because $K = K^*$, we call K self-dual.

Theorem 8. *Let K be a closed convex cone. Then*

1. *K^* is a closed convex cone. Moreover, $(K^*)^* = K$*
2. *K is pointed iff K^* has a nonempty interior (and vice versa)*
3. *$s \in K^*$ is strictly positive on K (besides the 0 vector) $\iff$ $K(s) = \{x \in K \mid \langle s, x \rangle \leq 1\}$*

References

[1] Y. Nesterov, *Lectures on Convex Optimization.* Springer Optimization and Its Applications.