EE C227C Convex Optimization and Approximation	Lecture 17 - 03/15/2022

Lecture 17: Newton's Method, Path Following and S-C Barrier

Lecturer: Jiantao Jiao *Scribe: Hyungki Im, Dilara Aykanat*

In this lecture we continue to study a convergence of variants of the Newton's Method. Also, we study central path following method and self-concordant barrier.

1 Quadratic Convergence of the Newton's Method

We start by reviewing some concepts of variants of Newton's method. Basically, Newton's method have the following update rule.

$$x_{k+1} = x_k - \frac{1}{1+\xi_k}[\nabla^2 f(x_k)]^{-1}\nabla f(x_k) \tag{1}$$

In this lecture, we consider three different variants of Newton's method which can be expressed as:

1. (Standard Newton's Method) $\xi_k = 0$, $\forall k \in \mathbb{N}$
2. (Damped Newton's Method) $\xi_k = M_f \lambda_f(x_k)$
3. (Intermediate Newton's Method) $\xi_k = \frac{M_f^2 \lambda_f^2(x_k)}{1+M_f\lambda_f(x_k)}$

In the previous lecture, we defined $\lambda_f(x_k)$ as $\lambda_f(x_k) = \sqrt{\nabla f(x_k)^\top \left(\nabla^2 f(x_k)\right)^{(-1)} \nabla f(x_k)}$ and this can be shortly expressed as $\|\nabla f(x_k)\|_x^*$. We can see that the Intermediate Newton's methods has more aggressive step size than a step size of the Damped Newton's method because $\frac{M_f^2 \lambda_f^2(x_k)}{1+M_f\lambda_f(x_k)} < M_f\lambda_f(x_k)$. The following theorem explains how a value of $\lambda_f(x_k)$ changes after an update of each Newton's method.

Theorem 1 (Theorem 5.2.2 of [1]). *Suppose we have $x \in domf$ and use λ to denote $\lambda = \lambda_f(x)$. Then the following holds:*

1. *If $\lambda < \frac{1}{M_f}$ and the point x_+ is generated by the Standard Newton's method, then $x_+ \in domf$ and $\lambda_f(x_+) \le \frac{M_f\lambda^2}{(1-M_f\lambda^2)}$ hold.*
2. *If the point x_+ is generated by the Damped Newton's method, then $\lambda_f(x_+) \le M_f\lambda^2(1+\frac{1}{1+M_f\lambda})$ holds.*
3. *If $M_f\lambda + (M_f\lambda)^2 + (M_f\lambda)^3 \le 1$ and the point x_+ is generated by the Intermediate Newton's method, then $x_+ \in domf$ and*

$$\lambda_f(x_+) \le M_f\lambda^2\Big(1 + M_f\lambda + \frac{M_f\lambda}{1+M_f\lambda+(M_f\lambda)^2}\Big) \le M_f\lambda^2(1+2M_f\lambda) \tag{2}$$

Among all the variants of Theorem 1, if we assume $\lambda < \frac{1}{2M_f}$, (2) seems attractive as (2) implies quadratic convergence of λ. Indeed, if $\lambda < \frac{1}{2M_f}$, we have

$$M_f\lambda(1+2M_f\lambda) \le 1 \Rightarrow M_f\lambda^2(1+2M_f\lambda) \le \lambda \tag{3}$$

and this implies the quadratic convergence of λ, if Intermediate Newton's method is used. Moreover, if $\lambda < \frac{1}{2M_f}$, the condition $M_f\lambda + (M_f\lambda)^2 + (M_f\lambda)^3 \le 1$ is satisfied. In [1], Nesterov used a term 'region of quadratic convergence' to denote a set

$$\Big\{x \in \mathrm{dom} f : \lambda_f(x) < \frac{1}{2M_f}\Big\}. \tag{4}$$

Thus, Theorem 1 leads us to the following scheme to solve a minimization problem.

- First Stage: When $\lambda_f(x_k) \geq \frac{1}{2M_f}$. At this stage we use Damped Newton's Method to update our solution candidate(x_k). In the previous lecture, we learned that Damped Newton's Method guarantees

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{M_f^2}\omega(\frac{1}{2}) \tag{5}$$

where we previously defined $\omega : \mathbb{R} \to \mathbb{R}$ as $\omega(t) = t - \log(1+t)$. This implies the number iterations of the first stage is bounded as follows:

$$N \leq \Delta_f(x_0)\omega(\frac{1}{2}) \tag{6}$$

where $\Delta_f(x_0) := M_f^2(f(x_0) - f(x_f^*))$ and x_f^* is the optimal solution of the problem.

- Second Stage: When $\lambda_f(x_k) < \frac{1}{2M_f}$. At this stage, we apply interdiate Newton's Method to update our solution candidate. By theorem 2, this process converges quadratically.

The most of the effort of this scheme are spend at the first stage because quadratic convergence is very fast.

2 Path Following

Here we begin by introducing a modern view of optimization. Lets assume $\text{dom}f = \{x \in \mathbb{R}^n : |f(x)| < \infty\}$ is a open set and denote $\text{Dom}f$ to denote a closure of $\text{dom}f(\text{cl}(\text{dom}f))$.

Definition 2. *We define a standard constrained minimization problem as*

$$\begin{aligned} \min \quad & \langle c, x\rangle \\ s.t. \quad & x \in Q \end{aligned} \tag{7}$$

where Q is a closed convex set and have a self-concordant function f such that $domf = Q$.

Now we consider a path-following technique to solve a standard constrained minimization problem.
We will solve (7) by tracing the *central path* defined as follows:

$$x^*(t) = \operatorname*{arg\,min}_{x \in \text{dom}f} f(t,x)$$

where $f(t;x) = t\langle c, x\rangle + F(x)$. It is clear that when we have $t = 0$ we minimize $F(x)$ and we minimize $\langle c, x\rangle$ when t goes to infinity. Suppose that $x = x^*(t)$ is given for a particular t, if we increase the value of t to obtain $t_+ = t + \Delta$, $x^*(t)$ will no longer be optimal. However, it should still be in the regime of quadratic convergence. In other words, we should guarantee that $\lambda_{f(t+\Delta;x)}(x) \leq \beta < \frac{1}{2}$ where β is called *the centering parameter*. Since a change in t does not effect the Hessian, we also have $\nabla^2 f(t+\Delta;x) = \nabla^2 f(t;x)$. Remember that $x = x^*(t)$ equates the gradient to zero such that $tc + \nabla f(x) = 0$ implying $c = \frac{-\nabla f(x)}{t}$. So we should satisfy the following *approximate centering condition*:

$$\begin{aligned} \lambda_{f(t+\nabla;x)}(x) = \|\nabla f(t+\Delta;x)\|_x^* &= \|t_+ c + \nabla f(x)\|_x^* \\ &= \|\Delta c\|_x^* \\ &= \Delta\|c\|_x^* \\ &= \frac{\Delta}{t}\|\nabla f(x)\|_x^* \leq \beta \end{aligned}$$

We see that $f(x)$ should not only be self-concordant but also for every x, $\|\nabla f(x)\|_x^*$ must be small as well in order to satisfy the latter condition. Thus, we come to a definition of a *self-concordant barrier.*

3 Self-concordant Barrier

Definition 3 (Definition 5.3.2 of [1])**.** *Let $F(.)$ be a standard self-concordant function. We call it a ν self-concordant barrier for the set Dom F, if:*

$$\nabla^2 F(x) \geq \frac{1}{\nu} \nabla F(x) \nabla F(x)^T, \quad \text{all x} \in \text{dom F}$$

If $\nabla^2 f(x) > 0$ then the hessian matrix is invertible, we have $\left\langle \nabla F(x), [\nabla^2 F(x)]^{-1} \nabla F(x) \right\rangle \leq \nu$. A good $F(.)$ is then characterized by a small ν to allow for large Δ in the algorithm. However, it was shown that for generic self-concordant barrier functions, ν cannot be too small. In one-dimensional setting for example, $\nu \geq 1$. Another detail is that this requirement is not too strong in the sense that Nesterov was able to show $F(.)$'s constructibility for any domain.

Example 4 (Example 5.3.1 of [1])**.**

1. *Linear function:* $f(x) = \alpha + \langle a, x \rangle$, dom $f = \mathbb{E}$. Clearly, f is a self-concordant function as $\nabla^3 f(x) = 0$ but for $a \neq 0$ this function is not a self-concordant barrier since $\nabla^2 f(x) = 0$.

2. *Convex quadratic function:* Let A symmetric, $A > 0$. Consider the function:

$$f(x) = \alpha + \langle a, x \rangle + \frac{1}{2} \langle Ax, x \rangle, \quad \text{dom}\, f = \mathbb{R}^n$$

Then $\nabla f(x) = a + Ax$ and $\nabla^2 f(x) = A$. Therefore,

$$\begin{aligned} \left\langle [\nabla^2 f(x)]^{-1} \nabla f(x), \nabla f(x) \right\rangle &= \left\langle A^{-1}(Ax + a), Ax + a \right\rangle \\ &= \langle Ax, x \rangle + 2 \langle a, x \rangle + \left\langle A^{-1} a, a \right\rangle \end{aligned}$$

Clearly, this value is unbounded from above on $\mathbb{R}^n$. Thus, a quadratic function is not a self-concordant barrier.

3. *Logarithmic barrier for a ray:* Consider the following function of one variable:

$$F(x) = -\ln x, \quad \text{dom}\, F = \{x \in \mathbb{R} \mid x > 0\}$$

Then $\nabla F(x) = \frac{-1}{x}$ and $\nabla^2 F(x) = \frac{1}{x^2} > 0$. Therefore

$$\frac{(\nabla F(x))^2}{\nabla^2 F(x)} = \frac{1}{x^2} . x^2 = 1.$$

Thus, $F(.)$ is a ν-self-concordant barrier for the set $\{x \geq 0\}$ with $\nu = 1$.

Theorem 5 (Theorem 5.3.2 of [1])**.** *Let F_i be ν_i-self-concordant barriers, $i = 1, 2$. Then the function $F(x) = F_1(x) + F_2(x)$ is a self-concordant barrier for a convex set $Dom\, F = Dom\, F_1 \bigcap Dom\, F_2$ with the parameter $\nu = \nu_1 + \nu_2$.*

Theorem 6 (Theorem 5.3.4 of [1])**.** *Let the function f be self-concordant with constant $M_f \geq 0$. Suppose that the set*

$$\mathcal{L}(\beta) = \{x \in dom\, f \;:\; f(x) \leq \beta\}$$

has nonempty interior and $f(x) \geq f^*$ for all $x \in dom\, f$. Then the function

$$F(x) = -\nu \ln(\beta - f(x))$$

with any $\nu \geq 1 + M_f^2(\beta - f^*)$ is a ν-self-concordant barrier for the level set $\mathcal{L}(\beta)$.

Properties of self-concordant barrier

Theorem 7 (Theorem 5.3.7 of [1]). *F is a ν-self-concordant barrier. For any x and y from dom F, we have*

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

Theorem 8 (Theorem 5.3.9 of [1]). *Let F be a ν-self-concordant barrier for the set Dom F . The point $x_F^* = \arg\min_{x \in \operatorname{dom} F} F(x)$ is called the analytic center of the convex set Dom F, generated by the barrier F. Assume that the analytic center of a ν-self-concordant barrier F exists. Then for any $x \in$ Dom F we have:*

$$\|x - x_F^*\|_{x_F^*} \leq \nu + 2\sqrt{\nu}$$

Corollary 9 (Corollary 5.3.4 of [1]). *Let Dom F be bounded. Then for any $x \in \operatorname{dom} F$ and $v \in \mathbb{R}^n$ we have*

$$\|v\|_x \leq (\nu + 2\sqrt{\nu})\|v\|_{x_F^*}^*$$

In other words, for any $x \in \operatorname{dom} F$ we have

$$\nabla^2 F(x) \geq \frac{1}{(\nu + 2\sqrt{\nu})^2} \nabla^2 F(x_F^*).$$

References

[1] Y. Nesterov *et al.*, *Lectures on convex optimization.* Springer, 2018, vol. 137.