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

Lecture 20: Interior Point Method - Part IV – April 10

*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

- Newton method for self-concordant functions
- Damped Newton method

**Reference**: Nesterov, Introductory Lectures on Convex Optimization, 2004, Chapter 4.1.5

## 20.1 Recall

In the last lecture, we discussed the unconstrained minimization of self-concordant function

$$\min_x f(x)$$

where $f(x)$ is standard self-concordant and non-degenerate.

Dikin ellipsoid: $W_r^0(x) = \{y : \| y - x \|_x < r\}$

For standard self-concordant function, $W_1^0(x) \subseteq dom(f), \forall x \in dom(f)$. Moreover, $f$ has some nice behavior insider the Dikin ellipsoid. $\forall y : \| y - x \|_x = \gamma < 1$, it holds

1. $(1-r)^2 \nabla^2 f(x) \preccurlyeq \nabla^2 f(y) \preccurlyeq \frac{1}{(1-r)^2} \nabla^2 f(x)$
2. $\frac{\gamma^2}{1+\gamma} \leq \langle \nabla f(y) - \nabla f(x), y - x \rangle \leq \frac{\gamma^2}{1-\gamma}$
3. $\omega(\gamma) \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \omega_*(\gamma)$

where $\omega(t) = t - \ln(1+t)$, $\omega_*(t) = -t - \ln(1-t)$.

Newton's decrement: $\lambda_f(x) = \sqrt{\nabla f(x)^T [\nabla^2 f(x)]^{-1} \nabla f(x)}$

- Note that $\lambda_f(x) = \| \nabla f(x) \|_{x,*} = \| d(x) \|_x$, where $d(x) = [\nabla^2 f(x)]^{-1} \nabla f(x)$
- If $\lambda_f(x) < 1$, the point $x_+ = x - d(x) \in dom(f)$
- If $x^*$ is a minimizer of $f$, then $\lambda_f(x^*) = 0$
- $\lambda_f(x_0) < 1$ for some $x_0 \in dom(f)$, then $f$ has a unique minimizer.

## 20.2 Newton Method for Self-concordant Function

<u>Basic Newton method</u>: initialize $x_0 \in dom(f)$ and update via

$$x_{k+1} = x_k - [\nabla^2 f(x_k)]^{-1} \nabla f(x_k), k = 0, 1, 2, ...$$

Unlike the classical analysis, we should adopt different measures of error that are independent of Euclidean metric. For instance,

- Function gap: $f(x_k) - f(x^*)$
- Newton's decrement: $\lambda_f(x_k) = \| \nabla f(x_k) \|_{x_k,*}$
- Local distance to the minimizer: $\| x_k - x^* \|_{x_k}$
- Distance to the minimizer under fixed metric $\| x_k - x^* \|_{x^*}$

Indeed, all of these measures are equivalent locally.

**Proposition 20.1** *When $\lambda_f(x) < 1$, we have*

1. $f(x) - f(x^*) \leq \omega_*(\lambda_f(x)) \leq \frac{\lambda_f(x)^2}{2(1-\lambda_f(x))^2}$
2. $\| x - x^* \|_x \leq \frac{\lambda_f(x)}{1-\lambda_f(x)}$
3. $\| x - x^* \|_{x^*} \leq \frac{\lambda_f(x)}{1-\lambda_f(x)}$

*Proof:* See Theorem 4.1.13 in (Nesterov, 2004). ■

For this reason, we will focus mainly on the convergence in terms of $\lambda_f(x)$.
For simplicity, in the following, let us denote $\lambda_k := \lambda_f(x_k), k = 0, 1, 2...$

**Theorem 20.2 (Local convergence)** *If $x_k \in dom(f)$ and $\lambda_k < 1$, then $x_{k+1} \in dom(f)$ and*

$$\lambda_{k+1} \leq \left( \frac{\lambda_k}{1 - \lambda_k} \right)^2$$

*Proof:* Note $\| x_{k+1} - x_k \|_{x_k} = \lambda_f(x_k) = \lambda_k < 1$, so $x_{k+1} \in dom(f)$. Also, it holds that

$$\nabla^2 f(x_{k+1}) \succcurlyeq (1 - \lambda_k)^2 \nabla^2 f(x_k)$$

Hence,

$$\lambda_{k+1} = \sqrt{\nabla f(x_{k+1})^T \Big[\nabla^2 f(x_{k+1})\Big]^{-1} \nabla f(x_{k+1})} \leq \frac{1}{1 - \lambda_k} \sqrt{\nabla f(x_{k+1})^T \Big[\nabla^2 f(x_k)\Big]^{-1} \nabla f(x_{k+1})}$$

Note

$$\begin{aligned}\nabla f(x_{k+1}) &= \nabla f(x_{k+1}) - \nabla f(x_k) - [\nabla^2 f(x_k)](x_{k+1}-x_k)\\ &= \Big[\int_0^1 \nabla^2 f(x_k + t(x_{k+1}-x_k)) - \nabla^2 f(x_k) dt\Big](x_{k+1}-x_k)\\ &:= G(x_{k+1}-x_k)\end{aligned}$$

Hence,

$$\begin{aligned}\lambda_{k+1} &\le \frac{1}{1-\lambda_k}\sqrt{(x_{k+1}-x_k)^T G^T [\nabla^2 f(x_k)]^{-1} G(x_{k+1}-x_k)}\\ &\le \frac{1}{1-\lambda_k}\|x_{k+1}-x_k\|_{x_k}\cdot \parallel [\nabla^2 f(x_k)]^{-1/2} G [\nabla^2 f(x_k)]^{-1/2} \parallel_2\\ &\le \frac{\lambda_k}{1-\lambda_k} \parallel \underbrace{[\nabla^2 f(x_k)]^{-1/2} G [\nabla^2 f(x_k)]^{-1/2}}_{H} \parallel_2\end{aligned}$$

We have

$$\begin{aligned} G &\succcurlyeq \nabla^2 f(x_k)\int_0^1 \Big[(1-t\lambda_k)^2 - 1\Big]dt = \Big(\frac{\lambda_k^2}{3}\Big)\nabla^2 f(x_k)\\ G &\preccurlyeq \nabla^2 f(x_k)\int_0^1 \Big[\frac{1}{(1-t\lambda_k)^2} - 1\Big]dt = \frac{\lambda_k}{1-\lambda_k}\nabla^2 f(x_k)\end{aligned}$$

Hence, $\parallel H \parallel_2 \le \max\left\{\lambda_k - \frac{\lambda_k^2}{3}, \frac{\lambda_k}{1-\lambda_k}\right\} = \frac{\lambda_k}{1-\lambda_k}$

This leads to the conclusion $\lambda_{k+1} \le (\frac{\lambda_k}{1-\lambda_k})^2$ ■

**Remark**: Let $\lambda^*$ be such that $\frac{\lambda^*}{(1-\lambda^*)^2} = 1$. Then if $\lambda_k < \lambda^*$, $\lambda_{k+1} < \lambda_k$. The region of quadratic convergence is $\lambda_f(x) \le \lambda^* = \frac{3-\sqrt{5}}{2} \approx 0.38$.

**Question**: T he Newton method for self-concordant function still might diverge if not started with a point with $\lambda_f(x)$ small enough. How to modify the Newton method to ensure global convergence?

The remedy is to use Newton method with line-search or damping factors

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

where $\gamma_k > 0$ is some stepsize.

## 20.3 Damped Newton Method

Damped Newton method: initialize $x_0 \in dom(f)$ and update via

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

Note this is essentially Newton method with particular stepsize $\gamma_k = \frac{1}{1+\lambda_f(x_k)}$.

**Remark.**

1. Since $\| x_{k+1} - x_k \|_{x_k} = \frac{\lambda_f(x_k)}{1+\lambda_f(x_k)} < 1$, $x_{k+1} \in W_1^0(x_k) \subseteq dom(f)$. The damped Newton procedure is always well-defined.
2. The Newton direction $d_k = -[\nabla^2 f(x_k)]^{-1} \nabla f(x_k)$ is a descent direction when $f$ is non-degenerate. If $f$ is bounded below, then $x_t$ converges to the unique minimizer of $f$.

**Theorem 20.3 (Global convergence of damped Newton)** *The damped Newton method satisfies that*

1. *(Descent phase)* $\forall k \geq 0, f(x_{k+1}) \leq f(x_k) - \omega(\lambda_f(x_k))$
2. *(Quadratic convergence phase) If* $\lambda_k(x_k) < \frac{1}{4}$, *then* $\lambda_f(x_{k+1}) \leq 2[\lambda_f(x_k)]^2$

*Proof:*

1. In view of previous theorem,
$$\begin{aligned} f(x_{k+1}) &\leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \omega_*(\| x_{k+1} - x_k \|_{x_k}) \\ &= f(x_k) - \frac{\lambda_f(x_k)}{1+\lambda_f(x_k)} + \omega_*\Big(\frac{\lambda_f(x_k)}{1+\lambda_f(x_k)}\Big) \end{aligned}$$
where $\omega_*(t) = -t - \ln(1-t)$.
This can be further simplified as $f(x_{k+1}) \leq f(x_k) - \omega(\lambda_f(x_k))$
2. Proof follows similarly as the earlier theorem for basic Newton method.

∎

**Remark**: A general strategy for solving the self-concordant minimization:

- Damped Newton stage: when $\lambda_f(x_k) \geq \beta$, where $\beta \in (0, 1/4)$
$$f(x_{k+1}) \leq f(x_k) - \omega(\beta)$$
The number of steps of this stage is bounded by $N_1 \leq \frac{f(x_0) - f(x^*)}{\omega(\beta)}$.
- (Damped/Basic) Newton stage: when $\lambda_f(x_k) < \beta$
$$\lambda_f(x_{k+1}) \leq 2\Big[\lambda_f(x_k)\Big]^2$$
The number of steps to find a solution with $\lambda_f(x) \leq \epsilon$ is bounded by $N_2 \leq O(1) \log_2 \log_2(\frac{1}{\epsilon})$.

The total complexity does not exceed
$$O(1)[f(x_0) - f^* + \log\log(\frac{1}{\epsilon})]$$