Optimization Theory and Algorithm	**Lecture 9 - 05/25/2021**

Lecture 9

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

1 Algorithm and Theory

For the convex and β-smooth function f, we use the same iterative method, that is

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \frac{1}{\beta}\nabla f(\mathbf{x}^t).$$

Theorem 1 *Let f be a convex and β-smooth function, and $\{\mathbf{x}^t\}_{t=0}^{\infty}$ is generated by the gradient descent algorithm with $1/\beta$ as the step size. Then for any $\epsilon > 0$, take $T \geq \frac{\beta}{\epsilon}\|\mathbf{x}^0 - \mathbf{x}^*\|^2$,*

$$f(\mathbf{x}^T) - f^* \leq \epsilon. \tag{1}$$

Proof 1 *Recall that*

$$m_t(\mathbf{x}) = f(\mathbf{x}^t) + \langle f(\mathbf{x}^t), \mathbf{x} - \mathbf{x}^t\rangle + \frac{\beta}{2}\|\mathbf{x} - \mathbf{x}^t\|^2,$$

then we can further prove that, for any $\mathbf{x}, \mathbf{y}$

$$m_t(\mathbf{y}) \geq m_t(\mathbf{x}) + \langle m_t(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{\beta}{2}\|\mathbf{y} - \mathbf{x}\|^2. \tag{2}$$

We know that $\mathbf{x}^{t+1}$ is the minimizer for the quadratic model $m_t(\mathbf{x})$. In Eq.(2), take $\mathbf{y} = \mathbf{x}^*$ and $\mathbf{x} = \mathbf{x}^{t+1}$, then we have,

$$m_t(\mathbf{x}^*) \geq m_t(\mathbf{x}^{t+1}) + \frac{\beta}{2}\|\mathbf{x}^{t+1} - \mathbf{x}^*\|^2.$$

Since m_t is a global quadratic upper bound of f, then

$$\begin{aligned}
f(\mathbf{x}^{t+1}) &\leq m_t(\mathbf{x}^{t+1}) \\
&\leq m_t(\mathbf{x}^*) - \frac{\beta}{2}\|\mathbf{x}^{t+1} - \mathbf{x}^*\|^2 \\
&= \underbrace{f(\mathbf{x}^t) + \langle f(\mathbf{x}^t), \mathbf{x}^* - \mathbf{x}^t\rangle + \frac{\beta}{2}\|\mathbf{x}^* - \mathbf{x}^t\|^2}_{m_t(\mathbf{x}^*)} - \frac{\beta}{2}\|\mathbf{x}^* - \mathbf{x}^{t+1}\|^2 \\
&\leq \overbrace{f(\mathbf{x}^t) + \langle f(\mathbf{x}^t), \mathbf{x}^* - \mathbf{x}^t\rangle}^{\leq f^*} + \frac{\beta}{2}\|\mathbf{x}^* - \mathbf{x}^t\|^2 - \frac{\beta}{2}\|\mathbf{x}^* - \mathbf{x}^{t+1}\|^2 \\
&\leq f^* + \frac{\beta}{2}(\|\mathbf{x}^* - \mathbf{x}^t\|^2 - \|\mathbf{x}^* - \mathbf{x}^{t+1}\|^2).
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{t=1}^{T-1}(f(\mathbf{x}^{t+1}) - f^*) &\leq \frac{\beta}{2}\sum_{t=1}^{T-1}(\|\mathbf{x}^* - \mathbf{x}^t\|^2 - \|\mathbf{x}^* - \mathbf{x}^{t+1}\|^2) \\
&= \frac{\beta}{2}(\|\mathbf{x}^* - \mathbf{x}^0\|^2 - \|\mathbf{x}^* - \mathbf{x}^T\|^2).
\end{aligned}$$

So, we have

$$f(\mathbf{x}^T) - f^* \leq \frac{1}{T}\sum_{t=0}^{T-1}(f(\mathbf{x}^{t+1}) - f^*) \leq \frac{\beta}{2T}\|\mathbf{x}^* - \mathbf{x}^0\|^2. \tag{3}$$

Remark 1 *Let us discuss the convergence speed. Suppose that, take $\epsilon = 10^{-2}$, then according to Theorem 1, it should be $T \geq 10^2$. If we take $\epsilon = 10^{-3}$, then $T \geq 10^3$.*

2 Gradient Descent for Beta Smooth and Alpha Strong Convex Function

Definition 1 *We say that a C^1-smooth function f is α-strongly convex (with $\alpha \geq 0$) if the following inequality holds,*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2 \quad \forall \mathbf{x}, \mathbf{y} \in (f).$$

An immediate consequence of the definition is the following lemma.

Lemma 1 *If $\mathbf{x}^*$ is the local minimal point of f, the it has*

$$f(\mathbf{y}) \geq f(\mathbf{x}^*) + \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}^*\|^2, \quad \forall \mathbf{y}.$$

Theorem 2 *f is α-strongly convex if and only if $f(\mathbf{x}) - \frac{\alpha}{2}\|\mathbf{x}\|^2$ is convex.*

Theorem 3 *Suppose $f \in \mathcal{C}^1$. Then the following are equivalent:*

1. *f is α-strongly convex.*
2. *$\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle \geq \alpha\|\mathbf{x} - \mathbf{y}\|^2$.*
3. *Additionally, if $f \in C^2$, then $\nabla^2 f \succeq \alpha I$ everywhere ($\nabla^2 f$ is positive definite).*

Theorem 4 *Assume that f is a β-smooth and α-strongly convex function, and $f^* = \inf f(\mathbf{x})$ exists, then for any $\epsilon > 0$, choose $T \geq \frac{2\beta}{\alpha}\log\frac{\|\mathbf{x}^0 - \mathbf{x}^*\|}{\epsilon}$, it has $\|\mathbf{x}^T - \mathbf{x}^*\| \leq \epsilon$.*

Proof 2

$$\begin{aligned}
\left\|\mathbf{x}^{t+1} - \mathbf{x}^*\right\|^2 &= \left\|\mathbf{x}^t - \mathbf{x}^* - \frac{1}{\beta}f(\mathbf{x}^t)\right\|^2 \\
&= \|\mathbf{x}^t - \mathbf{x}^*\|^2 + \frac{2}{\beta}\underbrace{\langle f(\mathbf{x}^t), \mathbf{x}^* - \mathbf{x}^t\rangle}_{\substack{\leq f(\mathbf{x}^*) - f(\mathbf{x}^t) - \frac{\alpha}{2}\|\mathbf{x}^t - \mathbf{x}^*\|^2 \\ \text{follows from strong convexity}}} + \frac{1}{\beta^2}\|f(\mathbf{x}^t)\|^2 \\
&\leq \left(1 - \frac{\alpha}{\beta}\right)\|\mathbf{x}^t - \mathbf{x}^*\|^2 + \underbrace{\frac{2}{\beta}\left(f^* - f(\mathbf{x}^t) + \frac{1}{2\beta}\|f(\mathbf{x}^t)\|^2\right)}_{\substack{\leq 0,\ \text{ignore} \\ f^* \leq f(\mathbf{x}^{t+1}) \leq f(\mathbf{x}^t) - \frac{1}{2\beta}\|f(\mathbf{x}^t)\|^2}}
\end{aligned}$$

$$\begin{aligned}
\|\mathbf{x}^T - \mathbf{x}^*\|^2 &\leq \left(1 - \frac{\alpha}{\beta}\right)\|\mathbf{x}^{T-1} - \mathbf{x}^*\|^2 \\
&\leq \left(1 - \frac{\alpha}{\beta}\right)^2\|\mathbf{x}^{T-2} - \mathbf{x}^*\|^2 \leq \ldots \leq \left(1 - \frac{\alpha}{\beta}\right)^T\|\mathbf{x}^0 - \mathbf{x}^*\|^2 \\
&\leq \exp\left(-\frac{\alpha}{\beta}T\right)\|\mathbf{x}^0 - \mathbf{x}^*\|^2,
\end{aligned}$$

where $\left(1-\frac{\alpha}{\beta}\right)^T \leq \exp\left(-\frac{\alpha}{\beta}T\right)$.

Example 1 *Let us consider the portfolio management problem again. That is*

$$\min_{\mathbf{x}} f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \Sigma \mathbf{x} - \mu^\top \Sigma, \tag{4}$$

where $\Sigma \in \mathcal{S}^n_{++}$.

- $\nabla f(\mathbf{x}) = \Sigma\mathbf{x} - \mu$, then it can be derived that f is $\lambda_{\max} := \lambda_{\max}(\Sigma)$-smooth.
- $\nabla^2 f(\mathbf{x}) = \Sigma \succeq \lambda_{\min}(\Sigma)I$. So, it is $\lambda_{\min}$-strongly convex.
- Algorithm: $\mathbf{x}^{t+1} = \mathbf{x}^t - \frac{1}{\lambda_{\max}}(\Sigma\mathbf{x}^t - \mu) = (I - \frac{\Sigma}{\lambda_{\max}})\mathbf{x}^t + \frac{\mu}{\lambda_{\max}}$.
- Convergence: $\|\mathbf{x}^{t+1} - \mathbf{x}^*\|^2 \leq (1 - \frac{\lambda_{\min}}{\lambda_{\max}})\|\mathbf{x}^t - \mathbf{x}^*\|^2$.
- $\frac{\lambda_{\max}}{\lambda_{\min}}$ is the *condition number* of Σ.
- The higher the condition number, the lower the convergence.

References