Notes for Lecture 10

Scribe: Tingkai Jia

1 Convergence of Newton's Method

Theorem 1. *Suppose the twice differentiable function $f : \mathbb{R}^d \to \mathbb{R}$ has L_2-Lipschitz continous Hessian and local minimizer $\mathbf{x}^*$ with $\nabla^2 f(\mathbf{x}^*) \succeq \mu I$, then the Newton's method*

$$\mathbf{x}_{t+1} = \mathbf{x}_t - (\nabla^2 f(\mathbf{x}_t))^{-1} \nabla f(\mathbf{x}_t)$$

with $\|\mathbf{x}_0 - \mathbf{x}^*\|_2 \le \mu/(2L_2)$ holds that

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 \le \frac{L_2}{\mu} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2.$$

Proof. It follows that

$$\begin{aligned}
\mathbf{x}_{t+1} - \mathbf{x}^* &= \mathbf{x}_t - (\nabla^2 f(\mathbf{x}_t))^{-1} \nabla f(\mathbf{x}_t) - \mathbf{x}^* \\
&= \mathbf{x}_t - (\nabla^2 f(\mathbf{x}_t))^{-1} (\nabla f(\mathbf{x}_t) - \nabla f(\mathbf{x}^*)) - \mathbf{x}^* \\
&= \mathbf{x}_t - \mathbf{x}^* - (\nabla^2 f(\mathbf{x}_t))^{-1} \int_0^1 \nabla^2 f(\mathbf{x}^* + \tau(\mathbf{x}_t - \mathbf{x}^*))(\mathbf{x}_t - \mathbf{x}^*) \, \mathrm{d}\tau \\
&= (\nabla^2 f(\mathbf{x}_t))^{-1} \int_0^1 \left(\nabla^2 f(\mathbf{x}_t) - \nabla^2 f(\mathbf{x}^* + \tau(\mathbf{x}_t - \mathbf{x}^*)) \right) (\mathbf{x}_t - \mathbf{x}^*) \, \mathrm{d}\tau \\
\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 &\le \lambda_{\min}^{-1}(\nabla^2 f(\mathbf{x}_t)) \int_0^1 L_2(1-\tau) \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \, \mathrm{d}\tau,
\end{aligned}$$

and with the L_2-Lipschitz continuous Hessian, we have

$$\begin{aligned}
\|\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{x}^*)\|_2 &\le L_2 \|\mathbf{x} - \mathbf{x}^*\|_2 \\
-L_2 \|\mathbf{x} - \mathbf{x}^*\|_2 \le \lambda_i(\nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{x}^*)) &\le L_2 \|\mathbf{x} - \mathbf{x}^*\|_2 \\
-L_2 \|\mathbf{x} - \mathbf{x}^*\|_2 I \le \nabla^2 f(\mathbf{x}) - \nabla^2 f(\mathbf{x}^*) &\le L_2 \|\mathbf{x} - \mathbf{x}^*\|_2 I \\
\nabla^2 f(\mathbf{x}) &\ge \nabla^2 f(\mathbf{x}^*) - L_2 \|\mathbf{x} - \mathbf{x}^*\|_2 I \\
\lambda_{\min}(\nabla^2 f(\mathbf{x})) &\ge \mu - L_2 \|\mathbf{x} - \mathbf{x}^*\|_2.
\end{aligned}$$

Since $\|\mathbf{x}_0 - \mathbf{x}^*\|_2 \le \mu/(2L_2)$, we can inductively show that $\|\mathbf{x}_t - \mathbf{x}^*\|_2 \le \frac{\mu}{2L_2}$. Thus we have

$$\mu - L_2 \|\mathbf{x}_t - \mathbf{x}^*\|_2 \ge \frac{\mu}{2}.$$

We finally obtain

$$\begin{aligned}
\lambda_{\min}^{-1}(\nabla^2 f(\mathbf{x}_t)) \int_0^1 L_2(1-\tau) \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \, \mathrm{d}\tau &= \frac{1}{2} \lambda_{\min}^{-1}(\nabla^2 f(\mathbf{x}_t)) L_2 \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \\
&\le \frac{L_2}{\mu} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2 \\
\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 &\le \frac{L_2}{\mu} \|\mathbf{x}_t - \mathbf{x}^*\|_2^2.
\end{aligned}$$

□

2 The SR1 Method

Theorem 2. *We consider secant condition and the symmetric rank one (SR1) update*

$$\begin{cases} \mathbf{y}_t = \mathbf{G}_{t+1}\mathbf{s}_t, \\ \mathbf{G}_{t+1} = \mathbf{G}_t + \mathbf{z}_t\mathbf{z}_t^\top. \end{cases}$$

where $\mathbf{s}_t = \mathbf{x}_{t+1} - \mathbf{x}_t$ *and* $\mathbf{y}_t = \nabla f(\mathbf{x}_{t+1}) - \nabla f(\mathbf{x}_t)$. *It implies*

$$\mathbf{G}_{t+1} = \mathbf{G}_t + \frac{(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top}{(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top\mathbf{s}_t}.$$

Proof. It follows that

$$\begin{aligned} \mathbf{y}_t &= (\mathbf{G}_t + \mathbf{z}_t\mathbf{z}_t^\top)\mathbf{s}_t \\ &= \mathbf{G}_t\mathbf{s}_t + (\mathbf{z}_t^\top\mathbf{s}_t)\mathbf{z}_t \\ \mathbf{s}_t^\top\mathbf{y}_t &= \mathbf{s}_t^\top\mathbf{G}_t\mathbf{s}_t + (\mathbf{z}_t^\top\mathbf{s}_t)^2 \\ (\mathbf{z}_t^\top\mathbf{s}_t)^2 &= \mathbf{s}_t^\top\mathbf{y}_t - \mathbf{s}_t^\top\mathbf{G}_t\mathbf{s}_t, \end{aligned}$$

we also have

$$\begin{aligned} \mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t &= (\mathbf{z}_t^\top\mathbf{s}_t)\mathbf{z}_t \\ (\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top &= (\mathbf{z}_t^\top\mathbf{s}_t)^2\mathbf{z}_t\mathbf{z}_t^\top \\ \mathbf{z}_t\mathbf{z}_t^\top &= \frac{(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top}{(\mathbf{z}_t^\top\mathbf{s}_t)^2} \\ &= \frac{(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top}{\mathbf{s}_t^\top\mathbf{y}_t - \mathbf{s}_t^\top\mathbf{G}_t\mathbf{s}_t}. \end{aligned}$$

Finally we have

$$\mathbf{G}_{t+1} = \mathbf{G}_t + \frac{(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top}{(\mathbf{y}_t - \mathbf{G}_t\mathbf{s}_t)^\top\mathbf{s}_t}.$$

□