Lecture 13

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

Theorem 1 *Suppose $h(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$ where f is β-smooth and α-convex. Then*

$$h(\mathbf{y}) \geq h(\mathbf{x}^+) + \langle G_\gamma(\mathbf{x}), \mathbf{y} - \mathbf{x}\rangle + \gamma\left(1 - \frac{\beta\gamma}{2}\right)\|G_\gamma(\mathbf{x})\|^2 + \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

Proof 1

$$\begin{aligned}
h(\mathbf{x}^+) =& f(\mathbf{x} - \gamma G_t(\mathbf{x})) + g(\mathbf{x}^+) \\
\leq& f(\mathbf{x}) - \gamma\langle \nabla f(\mathbf{x}), G_\gamma(\mathbf{x})\rangle + \frac{\beta}{2}\gamma^2\|G_\gamma(\mathbf{x})\|^2 + g(\mathbf{x}^+) \qquad \textit{(β smoothness)} \\
\leq& f(\mathbf{y}) + \langle \mathbf{x} - \mathbf{y}, \nabla f(\mathbf{x})\rangle - \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2 \quad \textit{(α convexity)} \\
& - \gamma\langle \nabla f(\mathbf{x}), G_\gamma(\mathbf{x})\rangle + \frac{\beta}{2}\gamma^2\|G_\gamma(\mathbf{x})\|^2 + g(\mathbf{x}^+) \\
\leq& f(\mathbf{y}) + \langle \nabla f(\mathbf{x}), \mathbf{x}^+ - \mathbf{y}\rangle - \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \frac{\beta\gamma^2}{2}\|G_\gamma(\mathbf{x})\|^2 + g(\mathbf{x}^+) \\
\leq& f(\mathbf{y}) + \langle \nabla f(\mathbf{x}), \mathbf{x}^+ - \mathbf{y}\rangle - \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2 + g(\mathbf{y}) \\
& + \langle G_\gamma(\mathbf{x}) - \nabla f(\mathbf{x}), \mathbf{x}^+ - \mathbf{y}\rangle + \frac{\beta\gamma^2}{2}\|G_\gamma(\mathbf{x})\|^2 \\
=& h(\mathbf{y}) + \langle G_\gamma(\mathbf{x}), \mathbf{x}^+ - \mathbf{y}\rangle - \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \frac{\beta\gamma^2}{2}\|G_\gamma(\mathbf{x})\|^2 \\
=& h(\mathbf{y}) + \langle G_\gamma(\mathbf{x}), \mathbf{x} - \mathbf{y}\rangle - \frac{\alpha}{2}\|\mathbf{y} - \mathbf{x}\|^2 - (\gamma - \frac{\beta\gamma^2}{2})\|G_\gamma(\mathbf{x})\|^2
\end{aligned}$$

Remark 1

- *α could be zero!*
- *If $\gamma = 1/\beta$ and $\mathbf{x} = \mathbf{y}$, then*

$$h(\mathbf{x}) \geq h(\mathbf{x}^+) + \frac{1}{2\beta}\|G_{1/\beta}(\mathbf{x})\|^2. \tag{1}$$

  This is the same with β-smooth function.
- *If $\gamma = 1/\beta$ and $\mathbf{y} = \mathbf{x}^*$, then*

$$h(\mathbf{x}^*) \geq h(\mathbf{x}^+) + \langle G_{1/\beta}(\mathbf{x}), \mathbf{x}^* - \mathbf{x}\rangle + \frac{1}{2\beta}\|G_{1/\beta}(\mathbf{x})\|^2 + \frac{\alpha}{2}\|\mathbf{x}^* - \mathbf{x}\|^2. \tag{2}$$

Theorem 2 *Consider problem $\min h = f + g$, if f is β-smooth and g is convex, the sequence generated by the proximal gradient descent algorithm satisfies,*

$$h(\mathbf{x}^T) - h^* \leq \frac{\beta}{2T}\|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

If further we assume f to be α-strongly convex, we have,

$$\|\mathbf{x}^T - \mathbf{x}^*\|^2 \leq \exp\left(-\frac{\alpha T}{\beta}\right)\|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

Where we h^* denote the optimal function value, and $\mathbf{x}^*$ optimal solution.

Proof 2 *If we set* $\gamma = \frac{1}{\beta}$ *and* $\mathbf{y} = \mathbf{x}^*$ *in the inequality in Theorem 1, then*

$$0 \geq h(\mathbf{x}^+) - h(\mathbf{x}^*) + \langle G_{1/\beta}(\mathbf{x}), \mathbf{x}^* - \mathbf{x} \rangle + \frac{1}{2\beta} \|G_{1/\beta}(\mathbf{x})\|^2 + \frac{\alpha}{2} \|\mathbf{x}^* - \mathbf{x}\|^2 \tag{3}$$

and in particular

$$\langle G_{1/\beta}(\mathbf{x}), \mathbf{x} - \mathbf{x}^* \rangle \geq \frac{1}{2\beta} \|G_{1/\beta}(\mathbf{x})\|^2 + \frac{\alpha}{2} \|\mathbf{x}^* - \mathbf{x}\|^2. \tag{4}$$

These allow to show convergence guarantees similar to those we obtained for smooth minimization. For instance, if h *is as above,*

$$\begin{aligned} h(\mathbf{x}^t) - h(\mathbf{x}^*) &\leq -\langle G_{1/\beta}(\mathbf{x}^{t-1}), \mathbf{x}^* - \mathbf{x}^{t-1} \rangle - \frac{1}{2\beta} \|G_{1/\beta}(\mathbf{x}^{t-1})\|^2 \text{ by equation (3)} \\ &= \frac{\beta}{2} \left(\|\mathbf{x}^{t-1} - \mathbf{x}^*\|^2 - \|\mathbf{x}^t - \mathbf{x}^*\|^2 \right) \end{aligned}$$

Adding up, and using telescoping, we get

$$h(\mathbf{x}^T) - h(\mathbf{x}^*) \leq \frac{\beta}{2T} (\|\mathbf{x}^0 - \mathbf{x}^*\|^2 - \|\mathbf{x}^T - \mathbf{x}^*\|^2) \leq \frac{\beta}{2T} \|x^0 - x^*\|^2.$$

If we assume further that f *is* α *strongly convex, then*

$$\begin{aligned} \|\mathbf{x}^{t+1} - \mathbf{x}^*\|^2 &= \|\mathbf{x}^t - \frac{1}{\beta} G_{1/\beta}(\mathbf{x}^t) - \mathbf{x}^*\|^2 \\ &\leq \|\mathbf{x}^t - \mathbf{x}^*\|^2 - \frac{2}{\beta} \langle G_{1/\beta}(\mathbf{x}^t), \mathbf{x}^t - \mathbf{x}^* \rangle + \frac{1}{\beta^2} \|G_{1/\beta}(\mathbf{x}^t)\|^2 \\ &\leq \|\mathbf{x}^t - \mathbf{x}^*\|^2 - \frac{2}{\beta} \left(\frac{1}{2\beta} \|G_{1/\beta}(\mathbf{x}^t)\|^2 + \frac{\alpha}{2} \|\mathbf{x}^k - \mathbf{x}^*\|^2 \right) + \frac{1}{\beta^2} \|G_{1/\beta}(\mathbf{x}^t)\|^2 \text{ by equation (4)} \\ &= \|\mathbf{x}^t - \mathbf{x}^*\|^2 - \frac{\alpha}{\beta} \|\mathbf{x}^t - \mathbf{x}^*\|^2. \end{aligned}$$

Recursively, we get

$$\|\mathbf{x}^T - x^*\|^2 \leq \left(1 - \frac{\alpha}{\beta}\right)^T \|x_0 - x^*\|^2 \leq \exp\left(-\frac{\alpha T}{\beta}\right) \|\mathbf{x}^0 - \mathbf{x}^*\|^2.$$

References