

Proof of Theorem 2

Let $\theta \in \text{dom } g$. Note from **Theorem 1** that we have for all $k \geq 0$

$$\begin{aligned} F(\theta_{k+1}; x) &\leq \mathcal{L}(\theta_k; x) + \nabla \mathcal{L}(\theta_k; x)^T (\theta_{k+1} - \theta_k) \\ &\quad + \frac{1}{2\eta} \|\theta_{k+1} - \theta_k\|_2^2 + \|\theta_{k+1}\|_1, \end{aligned} \quad (1)$$

Theorem A1: Let F be proper, closed, and strongly convex, for any $\theta \in R^n$, it holds that

$$\frac{\sigma}{2} \|\theta - \hat{\theta}\|_2^2 \leq F(\theta) - F(\hat{\theta}), \quad (2)$$

where $\sigma > 0$ and is called a strong convexity modulus. Generally, under the condition of **Theorem 1**, σ can be set $\frac{\sqrt{v_t}}{\eta}$. For all $k \geq 0$

$$\begin{aligned} F(\theta_{k+1}; x) &\leq \mathcal{L}(\theta_k; x) + \nabla \mathcal{L}(\theta_k; x)^T (\theta - \theta_k) \\ &\quad + \frac{1}{2\eta} \|\theta - \theta_k\|_2^2 + \|\theta\|_1 - \|\frac{\sqrt{v_t}}{\eta}\|_2 \|\theta_{k+1} - \theta\|_2^2. \end{aligned} \quad (3)$$

Setting $x = x^k$, we get

$$F(\theta_{k+1}; x) \leq \mathcal{L}(\theta_k; x) + \|\theta\|_1 - \|\frac{\sqrt{v_t}}{\eta}\|_2 \|\theta_{k+1} - \theta_k\|_2^2 \leq F(\theta_k; x), \quad (4)$$

showing that $F(\theta_k; x)$ is nonincreasing. Now, pick any $\hat{\theta} \in \text{Arg min } F$ and let $\theta = \hat{\theta}$, then

$$\begin{aligned} F(\theta_{k+1}; x) &\leq \mathcal{L}(\theta_k; x) + \nabla \mathcal{L}(\theta_k; x)^T (\hat{\theta} - \theta_k) \\ &\quad + \frac{1}{2\eta} \|\hat{\theta} - \theta_k\|_2^2 + \|\hat{\theta}\|_1 - \|\frac{\sqrt{v_t}}{\eta}\|_2 \|\theta_{k+1} - \hat{\theta}\|_2^2 \\ &\leq \mathcal{L}(\hat{\theta}; x) + \|\hat{\theta}\|_1 + \frac{1}{2\eta} \|\hat{\theta} - \theta_k\|_2^2 - \|\frac{\sqrt{v_t}}{\eta}\|_2 \|\theta_{k+1} - \hat{\theta}\|_2^2. \end{aligned} \quad (5)$$

Thus, we obtain:

$$\begin{aligned} (k+1)[F(\theta_{k+1}; x) - F(\hat{\theta}; x)] &\leq \sum_{i=0}^k (F(\theta_{i+1}; x) - F(\hat{\theta})) \\ &\leq \frac{1}{2\eta} \sum_{i=0}^k \|\hat{\theta} - \theta_i\|_2^2 - \|\theta_{i+1} - \hat{\theta}\|_2^2 \leq \|\frac{\sqrt{v_t}}{\eta}\|_2 \|\hat{\theta} - \theta_0\|_2^2 \leq \frac{G}{\eta} \|\hat{\theta} - \theta_0\|_2^2. \end{aligned} \quad (6)$$

Hence,

$$F(\theta_k; x) - F(\hat{\theta}) \leq \frac{G}{2k\eta} \|\theta_0 - \hat{\theta}\|_2^2. \quad (7)$$