

$$\min_y \hat{f}(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2\epsilon} \|y-x\|^2$$

$$\Leftrightarrow \min_y \frac{1}{\epsilon} \|y - \underbrace{(x - \epsilon \nabla f(x))}_{\text{subgradient}}\|^2$$

$$y = x - \epsilon \nabla f(x).$$

$$\nabla^2 f \leq LI$$

$$A \geq 0.$$

$$x^T A x \geq 0 \quad \forall x.$$

Proof.

$$\forall x, y \quad f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2} (x-y)^T \nabla^2 f^2(z) (x-y)$$

for some $z \in [x, y]$ Taylor remainder thm

$$\nabla^2 f \leq LI$$

$$(x-y)^T (\nabla^2 f - LI) (x-y) \leq 0 \quad \forall x, y.$$

$$L \|x-y\|^2 \geq (x-y)^T \nabla^2 f(z) (x-y)$$

$$\bullet \quad \left| f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{L}{2} \|x-y\|^2 \right|$$

$$y = x - t \nabla f(x).$$

$$f(x - t \nabla f(x)) \leq f(x) - \nabla f(x)^T (t \nabla f(x)) + \frac{L t^2}{2} \|\nabla f(x)\|^2$$

$\underbrace{\hspace{10em}}_{x^+}$

$$\bullet \quad = f(x) - (1 - \frac{L t}{2}) t \|\nabla f(x)\|^2$$

$$\bullet \quad t \leq \frac{1}{L} \quad (1 - \frac{L t}{2}) \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

$$\rightarrow \left| \begin{aligned} f(x^+) &\leq f(x) - \frac{t}{2} \|\nabla f(x)\|^2 \\ &\leq f(x^*) + \nabla f(x)^T (x-x^*) - \frac{t}{2} \|\nabla f(x)\|^2 \\ &= f(x^*) + \frac{1}{2t} (\|x-x^*\|^2 - \|x-x^* - t \nabla f(x)\|^2) \end{aligned} \right| \quad \begin{aligned} f(x^*) &\geq f(x) \\ &- \nabla f(x)^T (x-x^*) \\ f(x) &\leq f(x^*) \\ &+ \nabla f(x)^T (x-x^*) \end{aligned}$$

$\|x^+ - x^*\|^2$

$$f(x^+) \leq f(x^*) + \frac{1}{2t} (\|x - x^*\|^2 - \|x^+ - x^*\|^2).$$

$$\begin{aligned} \sum_{i=1}^k (f(x^{(i)}) - f(x^*)) &\leq \frac{1}{2t} \sum_{i=1}^k (\|x^{(i-1)} - x^*\|^2 \\ &\quad - \|x^{(i)} - x^*\|^2) \\ &= \frac{1}{2t} (\|x^{(0)} - x^*\|^2 - \|x^{(k)} - x^*\|^2) \\ &\leq \frac{1}{2t} \|x^{(0)} - x^*\|^2. \end{aligned}$$

$f(x^{(i)})$ is nonincreasing.

$$\begin{aligned} f(x^{(k)}) - f(x^*) &\leq \frac{1}{k} \sum_{i=1}^k (f(x^{(i)}) - f(x^*)) \\ &\leq \frac{\|x^{(0)} - x^*\|^2}{2tk}. \end{aligned}$$

$$f(x) = \frac{1}{2} \|y - Ax\|^2$$

$$\nabla f(x) = -A^T(y - Ax).$$

$$\nabla^2 f(x) = A^T A.$$

$$A^T A \preceq LI, \text{ take } L = \sigma_{\max}^2(A).$$

$$u \in \partial \|\nabla f(x)\|_r,$$