$f(x) = g(x) + h(x)$, composite optimization

$$x^{k+1} = x^k - t^k \nabla f(x^k)$$

f Lipschitz: $|f(x) - f(y)| \leq G\|x-y\| \quad \forall x, y$

∇f Lipschitz: $\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|$

Subgradient Method: f is Lipschitz (bounded gradient)

Subgradient: $f(y) \geq f(x) + \nabla f(x)\cdot(y-x)$

$$t^k = \frac{1}{\sqrt{K}} \Rightarrow f(x^{best}) - f(x^*) \leq \frac{\|x-x^*\|^2 + G^2}{2\sqrt{K}}$$

Gradient Method: ∇f is Lipschitz (bounded Hessian)

$$t^k = \frac{1}{L} \Rightarrow f(x^K) - f(x^*) \leq \frac{L\|x-x^*\|^2}{2K}$$

If $g(x)$ is smooth and $h(x)$ is "simple", we can take advantage of structure and achieve a better rate. Both g, h, are convex.

Recall: $$x^{k+1} = \arg\min_y f(x^k) + \langle \nabla f(x^k), y-x \rangle + \frac{1}{2t^k}\|y-x\|^2$$

Idea, only approximate $g(x)$:

$$x^{k+1} = \arg\min_y \cancel{f(x)}\ g(x^k) + \langle \nabla g(x^k), y-x \rangle + h(y) + \frac{1}{2t^k}\|y-x\|^2$$

If we can solve this optimization "quickly", we are done.

In class, we used $\mathrm{prox}_t(x) = \arg\min_y \frac{1}{2t}\|x-y\|^2 + h(y)$

Note: $x^{k+1} = \mathrm{prox}_{t^k}(x^k - t^k \nabla g(x^k))$

We called the generalized gradient $G_t(x) = \dfrac{x - \mathrm{prox}_t(x - t\nabla g(x))}{t}$

i.e., $x^{k+1} = x^k - t^k G_{t^k}(x^k)$

Gradient Descent Proof:

$$\nabla g \text{ is } L\text{-Lipschitz} \Rightarrow g(y) \overset{+h(y)}{\leq} g(x) + \nabla g(x)\cdot(y-x) + \frac{L}{2}\|y-x\|^2 \overset{+h(y)}{,} \quad \forall x, y$$

Choosing $y = x^+ = x - t\nabla f(x)$

$$g(x^+) \leq g(x) - \left(1 - \frac{Lt}{2}\right)t\|\nabla g(x)\|^2$$

If $0 < t \leq \frac{1}{L}$, $\Rightarrow 1 - \frac{Lt}{2} \geq \frac{1}{2}$

$$\Rightarrow g(x^+) \leq g(x) - \frac{t}{2}\|\nabla g(x)\|^2$$

As $g(x^*) \geq g(x) + \nabla g(x)\cdot(x^* - x)$

$$\Rightarrow g(x^+) \leq g(x^*) + \nabla g(x)\cdot(x - x^*) - \frac{t}{2}\|\nabla g(x)\|^2$$

$$= g(x^*) + \frac{1}{2t}\left[\|x - x^*\|^2 - \|x^+ - x^*\|^2\right] \quad \text{(Expand to check)}$$

$$\Rightarrow \sum_{i=1}^{k}\left(f(x^i) - f(x^*)\right) \leq \frac{1}{2t}\sum_{i=1}^{k}\left[\|x^{i-1} - x^*\|^2 - \|x^i - x^*\|^2\right]$$

$$= \frac{1}{2t}\left[\|x^0 - x^*\|^2 - \|x^k - x^*\|^2\right]$$

$$\leq \frac{1}{2t}\|x^0 - x^*\|^2$$

Non-increasing $\therefore$

$$f(x^k) - f(x^*) \leq \frac{1}{k}\sum_{i=1}^{k}\left[f(x^i) - f(x^*)\right] \leq \frac{\|x^0 - x^*\|^2}{2tk}$$

$$y = x^+ = x - tG_t(x)$$

$$f(x^+) \leq g(x) - t g(x)\cdot G_t(x) + \frac{Lt^2}{2}\|G_t(x)\|^2 + h(x^+)$$

Since $x^+ = \text{prox}_t(x - t\nabla g(x))$

$$= \underset{y}{\text{argmin}}\ \frac{1}{2t}\|x - t\nabla g(x) - y\|^2 + h(y)$$

$$\Rightarrow -\frac{(\cancel{x} - t\nabla g(x) - \cancel{x} + tG_t(x))}{t} + v = 0, \quad v \in \partial h(x^+)$$

$$\Rightarrow -\nabla g(x) + G_t(x) \in \partial h(x^+)$$

$$\Rightarrow h(x) \geq h(x^+) + [G_t(x) - \nabla g(x)]\cdot[x - x^+]$$

$$\Rightarrow f(x^+) \leq f(x) + \frac{Lt^2}{2}\|G_t(x)\|^2 - t g(x)\cdot G_t(x) - [G_t(x) - \nabla g(x)]\cdot tG_t(x)$$

$$= f(x) - \left(1 - \frac{Lt}{2}\right)t\|G_t(x)\|^2$$

ISTA: $\text{prox}_t(x) = \underset{y}{\text{argmin}} \frac{1}{2t}\|y-x\|^2 + h(y)$, $h(y) = \|y\|_1$

$$\frac{y^* - x}{t} + v = 0, \quad v \in \partial h(y^*)$$

Claim: $[\text{prox}_t(x)]_i = \begin{cases} x_i - t & x_i > t \\ x_i + t & x_i < -t \\ 0 & \text{otherwise} \end{cases}$

Case $x_i > t$:

$$\frac{x_i - t - x_i}{t} = -1, \quad [\partial h(x_i - t)]_i = \{1\}$$

Case $x_i < -t$:

$$\frac{x_i + t - x_i}{t} = 1, \quad [\partial h(x_i + t)]_i = \{-1\}$$

Case $|x_i| \leq 1$:

$$-1 \leq \frac{-x_i}{t} \leq 1, \quad [\partial h(0)]_i = [-1, 1]$$

$h(y) = I(y \geq 0) = \begin{cases} 0 & y \geq 0 \\ \infty & \text{otherwise} \end{cases}$

Claim: $\text{prox}_t(x) = \max(x, 0)$

Case $x_i < 0$:

$$\frac{0 - x_i}{t} + v = 0, \text{ show } v \in \partial I(0 \geq 0)$$

show: $f(z) \geq f(0) + \frac{x_i}{t}(z - 0)$

if $z \geq 0 \Rightarrow \frac{x_i z}{t} \leq 0$

$z < 0 \Rightarrow \frac{x_i z}{t} > 0$, but $f(z) = \infty$

$h(y) = I(\|y\| \leq 1)$

Claim: $\text{prox}_t(x) = \begin{cases} x & \text{if } \|x\| \leq 1 \\ \frac{x}{\|x\|} & \text{otherwise} \end{cases}$

Case $\|x\| > 1$: show: $-\frac{\frac{x}{\|x\|} - x}{t} \in \partial h\left(\frac{x}{\|x\|}\right)$

if $\|z\| \leq 1$:

$$-\frac{1}{t}\left(\frac{1}{\|x\|} - 1\right)x \cdot \left(z - \frac{x}{\|x\|}\right) = c\,x \cdot z - c\frac{x \cdot x}{\|x\|}$$
$$\leq c\|x\| - c\|x\| = 0$$

$h(y) = I(a \cdot y \leq 0)$

Claim: $\text{prox}_t(x) = \begin{cases} x & \text{if } a \cdot x \leq 0 \\ x - \frac{a \cdot x}{a \cdot a} a & \text{otherwise} \end{cases}$

Case $a \cdot x > 0$, show $\frac{1}{t}\left[\frac{a \cdot x}{a \cdot a} a\right]$ is subgradient

$$\frac{1}{t}\left[\frac{a \cdot x}{a \cdot a} a\right] \cdot \left[z - x + \frac{a \cdot x}{a \cdot a} a\right]$$

$$= \frac{1}{t}\left[\frac{a \cdot x}{a \cdot a} a\right] \cdot z \leq \begin{cases} 0 & \text{if } a \cdot z \leq 0 \\ \infty & \text{otherwise} \end{cases}$$