

Proximal Gradient Method

- Convex function : $f(x) = g(x) + h(x)$
 - \nearrow convex, cont. diff.
 - \uparrow convex but not diff. everywhere
- Can minimize $f(x)$ using sub-gradient algorithm, but rate of convergence is at best $O\left(\frac{1}{\sqrt{N}}\right)$, i.e.,
$$f_N^* - f^* \sim O\left(\frac{1}{\sqrt{N}}\right)$$

Proximal Gradient Iteration

Since g is cont. diff., if $h(x) = 0$, GD iteration would be

$$x_{k+1} = x_k - \alpha_k \nabla g(x_k)$$

with $h(x) \neq 0$, use "proximal" gradient iterate:

$$x_{k+1} = \arg \min_x \frac{1}{2\alpha_k} \|x - (x_k - \alpha_k \nabla g(x_k))\|^2 + h(x)$$

$\underbrace{\hspace{10em}}$

penalty for deviating from
 $x_k - \alpha_k \nabla g(x_k)$

- Inherent Assumption: h is s.t. above min. is easy (closed-form).

Convergence Analysis:

$$x_{k+1} = \arg \min_x \underbrace{\frac{1}{2\alpha_k} \|x - (x_k - \alpha_k \nabla g(x_k))\|^2}_{\text{convex in } x} + h(x)$$

Write solution x_{k+1} as $x_k - \alpha_k d_k$ for convenience

- d_k is a "generalized" gradient at x_k .

$$x_{k+1} = x_k - \alpha_k d_k$$

Assume g has Lipschitz gradients, and use constant stepsize $\alpha_k = \alpha \forall k$. By Descent Lemma,

$$g(x_{k+1}) \leq g(x_k) - \alpha \nabla g(x_k)^T d_k + \frac{1}{2} L \alpha^2 \|d_k\|^2$$

Now, since $x_{k+1} = x_k - \alpha d_k$ minimizes convex function

$$\underbrace{\frac{1}{2\alpha} \|x - (x_k - \alpha \nabla g(x_k))\|^2}_{\text{gradient}} + h(x)$$

Gradient = $\frac{1}{\alpha} (x - (x_k - \alpha \nabla g(x_k)))$ \uparrow
subgradient v
 $\in \partial h(x)$

$$\Rightarrow 0 = \frac{1}{\alpha} (x_{k+1} - (x_k - \alpha \nabla g(x_k))) + v, \text{ for some } v \in \partial h(x_{k+1})$$

$$\text{i.e., } v = (\nabla g(x_k) - d_k) + v, \text{ for some } v \in \partial h(x_{k+1})$$

i.e. $v = d_k - \nabla g(x_k)$, for some $v \in \partial h(x_{k+1})$

But by definition of $\partial h(x_{k+1})$,

$$h(x_{k+1}) + v^T(x - x_{k+1}) \leq h(x), \quad \forall x \in \mathbb{R}^n$$

i.e.,

$$h(x_{k+1}) + (d_k - \nabla g(x_k))^T(x - x_{k+1}) \leq h(x), \quad \forall x \in \mathbb{R}^n$$

$$\text{i.e. } h(x_{k+1}) \leq (\nabla g(x_k) - d_k)^T(x - x_{k+1}) + h(x), \quad \forall x \in \mathbb{R}^n - (1)$$

Also, from earlier (descent lemma),

$$g(x_{k+1}) \leq g(x_k) - \alpha \nabla g(x_k)^T d_k + \frac{1}{2} L \alpha^2 \|d_k\|^2$$

$$\xrightarrow{\text{Convexity of } g} g(x) + \nabla g(x_k)^T(x_k - x) - \alpha \nabla g(x_k)^T d_k + \frac{1}{2} L \alpha^2 \|d_k\|^2, \quad \forall x \in \mathbb{R}^n - (2)$$

Combine (1), (2) to get

to get

$$f(x_{k+1}) \leq f(x) + \nabla g(x_k)^T(\overbrace{x_k - x + x - x_{k+1} - \alpha d_k}^{=0}) + d_k^T(x_{k+1} - x) + \frac{1}{2} L \alpha^2 \|d_k\|^2$$

i.e.,

$$f(x_{k+1}) \leq f(x) + d_k^T(x_{k+1} - x) + \frac{1}{2} L \alpha^2 \|d_k\|^2$$

$$\xrightarrow{x_{k+1} = x_k - \alpha d_k} = f(x) + d_k^T(x_k - x) - \alpha \|d_k\|^2 + \frac{1}{2} L \alpha^2 \|d_k\|^2$$

$$f(x_{k+1}) \leq f(x) + d_k^T(x_k - x) - \alpha \|d_k\|^2 + \frac{1}{2} L \alpha^2 \|d_k\|^2, \quad \forall x \in \mathbb{R}^n$$

Set $\alpha < \frac{1}{L}$, i.e. $L < \frac{1}{\alpha}$. Then,

$$f(x_{k+1}) \leq f(x) + d_k^T(x_k - x) - \frac{\alpha}{2} \|d_k\|^2, \quad \forall x$$

Setting $x = x_k$ on RHS, we obtain

$$f(x_{k+1}) \leq f(x_k) - \frac{\alpha}{2} \|d_k\|^2 \leq f(x_k)$$

i.e. the algorithm is a descent algorithm (if $\alpha < \frac{1}{L}$)

Also, setting $x = x^*$ on RHS in box, we obtain

$$\begin{aligned} f(x_{k+1}) &\leq f(x^*) + d_k^T(x_k - x^*) - \frac{\alpha}{2} \|d_k\|^2 \\ &= f(x^*) - \frac{\alpha}{2} \left(\|d_k\|^2 - \frac{2}{\alpha} d_k^T(x_k - x^*) \right) \\ &= f(x^*) - \frac{\alpha}{2} \left(\left\| \frac{x_k - x^*}{\alpha} - d_k \right\|^2 - \left\| \frac{x_k - x^*}{\alpha} \right\|^2 \right) \\ &= f(x^*) - \frac{1}{2\alpha} \left(\|x_k - \alpha d_k - x^*\|^2 - \|x_k - x^*\|^2 \right) \\ &= f(x^*) - \frac{1}{2\alpha} \left(\|x_{k+1} - x^*\|^2 - \|x_k - x^*\|^2 \right) \end{aligned}$$

$$f(x_{k+1}) - f(x^*) \leq \frac{1}{2\alpha} \left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2 \right)$$

Sum from $k=0$ to $N-1$ on both sides,

$$\begin{aligned} \sum_{k=0}^{N-1} (f(x_{k+1}) - f(x^*)) &\leq \frac{1}{2\alpha} \left(\|x_0 - x^*\|^2 - \|x_{N+1} - x^*\|^2 \right) \\ &\leq \frac{1}{2\alpha} \|x_0 - x^*\|^2 \end{aligned}$$

$$\text{But } f(x_N) \leq f(x_{N-1}) \leq \dots \leq f(x_0)$$

$$\Rightarrow N(f(x_N) - f(x^*)) \leq \sum_{k=0}^{N-1} f(x_{k+1}) - f(x^*)$$

$$\begin{aligned} \text{i.e. } f(x_N) - f(x^*) &\leq \frac{1}{2\alpha N} \|x_0 - x^*\|^2 \\ &\sim O(\frac{1}{N}) \end{aligned}$$

Note The proximal gradient algorithm is only useful if we can easily solve

$$\min_x \frac{1}{2\alpha} \|x - x_k + \alpha \nabla g(x_k)\|^2 + h(x)$$

$$\text{Example } h(x) = \|x\|_1 = \sum_{i=1}^n |x(i)|$$

$$\text{Let } \bar{z} = x_k - \alpha \nabla g(x_k)$$

$$\underset{x}{\text{minimize}} \quad \frac{1}{2\alpha} \sum_{i=1}^n (x(i) - \bar{z}(i))^2 + \sum_{i=1}^n |x(i)|$$

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^n \frac{1}{2\alpha} (x(i) - z(i))^2 + |x(i)|$$

$$\equiv \text{for each } i, \underset{x(i)}{\text{minimize}} \underbrace{\frac{(x(i) - z(i))^2}{2\alpha} + |x(i)|}_{C(x(i))}$$

Case 1 $x(i) > 0$ optimal

$$x(i) > 0 \Rightarrow C(x(i)) = \frac{(x(i) - z(i))^2}{2\alpha} + x(i)$$

$$\nabla C(x(i)) = 0 \Rightarrow \frac{1}{\alpha} (x(i) - z(i)) + 1 = 0 \\ \Rightarrow x^*(i) = z(i) - \alpha$$

This is solution if $z(i) > \alpha$

Case 2 $x(i) < 0$ is optimal

$$x(i) < 0 \Rightarrow C(x(i)) = \frac{(x(i) - z(i))^2}{2\alpha} - x(i)$$

$$\nabla C(x(i)) = 0 \Rightarrow x^*(i) = z(i) + \alpha$$

This is the solution if $z(i) < -\alpha$.

Case 3 $x(i) = 0$ is optimal

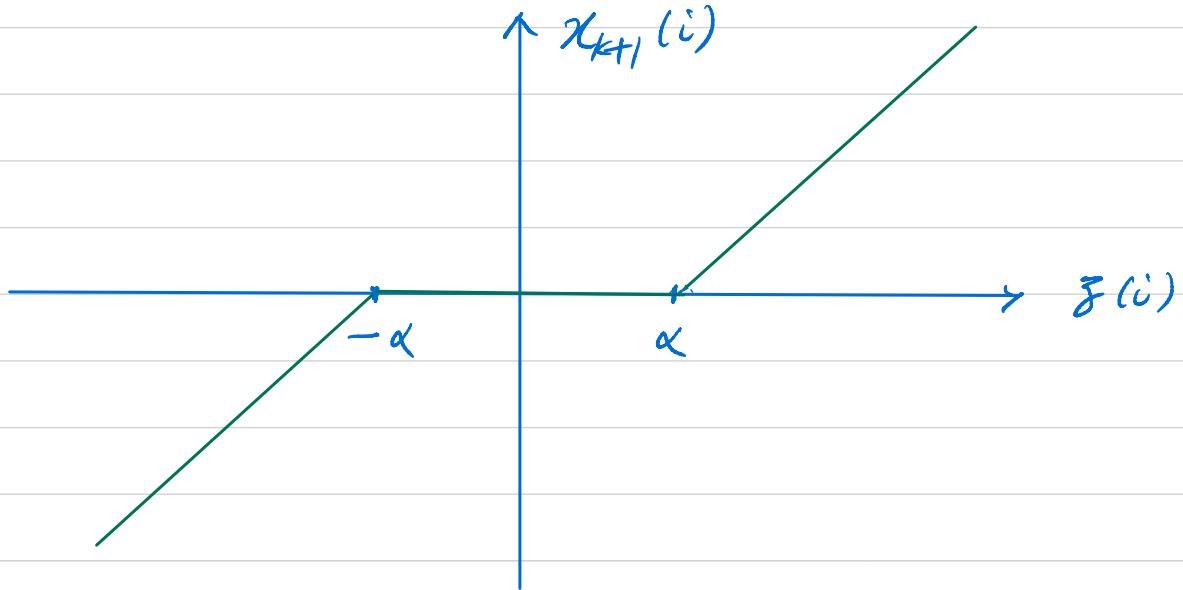
$$\nabla C(0) = -\frac{z(i)}{\alpha} + \varphi, \varphi \in [-1, 1]$$

$$0 \in \partial C(0) \Leftrightarrow -\frac{z(i)}{\alpha} + \varphi = 0 \text{ for some } \varphi \in [-1, 1]$$

$$\Leftrightarrow z(i) \in [-\alpha, \alpha].$$

Thus,

$$x_{k+1}(i) = x^*(i) = \begin{cases} \bar{z}(i) - \alpha & \text{if } \bar{z}(i) > \alpha \\ 0 & \text{if } -\alpha \leq \bar{z}(i) \leq \alpha \\ \bar{z}(i) + \alpha & \text{if } \bar{z}(i) < -\alpha \end{cases}$$



$x_{k+1}(i)$ is obtained by "shrinking" $\bar{z}(i)$ to 0 in α -neighborhood of 0.

$$x_{k+1}(i) = S_\alpha(\bar{z}(i))$$

\uparrow Shrinkage operator

Problems of the form :

$$\underset{x}{\text{minimize}} \quad g(x) + \lambda \|x\|_1$$

\uparrow convex, cont. diff.

arise in many modern applications:
LASSO, compressed sensing, etc.