OP202 HW3

Yanyu ZHOU

February 22, 2023

1 Q1

Consider

$$\min_{x \in X} f(x) \tag{PC}$$

$$\Updownarrow$$

$$\min_{x \in \mathbf{R}^n} f(x) + \iota_X(x) \tag{PS$'$}$$

where $\iota_X(x)$ is an indicator function. When $x \in X$, $\iota_X(x) = 1$, else $\iota_X(x) = \infty$.

Theorem 1. $\partial \iota_X(x) \equiv N_X(x)$

Proof. For (PC), the necessary and sufficient conditions for optimality is

$$\partial f(x^*) + N_X(x^*) \ni \mathbf{0}$$

where $N_X(x)$ is the normal cone operator of X at x. By definition, the *normal cone* of a set $\mathbf{X}$ at a boundary point x_0 is the set of all vectors y such that $y^T(x - x_0) \geq 0$ for all $x \in \mathbf{X}$.

By definition, $\partial \iota_X(x) = \{g \in \mathbf{R}^n : g^T x \geq g^T y \quad \forall y \in \mathbf{X}\}$. Therefore, $\partial \iota_X(x) \equiv N_X(x)$. $\square$

2 Q2

Work out one of the proofs of Theorem 17 (the simplest one is $f_1 \in \mathbf{S}^{1,1}_{m,L}$). Here we prove that Theorem 12 in class 2 is also valid for the proximal gradient method. Referrence: CMU 10-725 Lecture 8.

For $f(x) = f_1(x) + f_2(x)$, we assume that

- f_1 is convex and at least $f_1 \in \mathbf{S}^{1,1}_{0,L}$
- f_2 is convex, $\text{prox}_{\alpha h}(x) = \arg\min_z \left\{ \frac{\|x-z\|^2}{2\alpha} + f_2(z) \right\}$ can be evaluated.

Notation: The update step of proximal gradient method can be written as

$$x_k = x_{k-1} - \alpha G_{\alpha_k}(x_{k-1})$$

where $G_\alpha(x)$ is the generalized gradient and is given by

$$G_\alpha(x) = \frac{x - \operatorname{prox}_\alpha(x - \alpha \nabla f_1(x))}{\alpha}.$$

Theorem 2. *The proximal gradient method with constant step size $\alpha \leq \frac{2}{L}$ satisfies*

$$f(x_k) - f(x^*) \leq \frac{\|x_0 - x*\|^2}{2\alpha k}$$

where x^* is the optimal solution.

Proof. We begin by showing that

$$f(y) \leq f_1(x) + \nabla g(x)^T (y - x) + \frac{L}{2}\|y - x\|^2 + f_2(y) \quad \forall x, y.$$

Since $f_1 \in \mathbf{S}_{0,L}^{1,1}$, by Taylor's expansion

$$\begin{aligned} f_1(y) &= f_1(x) + f_1(x)^T (y - x) + \frac{1}{2}(x - y)^T \nabla^2 f_1(x)(x - y) \\ &\leq f_1(x) + f_1(x)^T (y - x) + \frac{L}{2}\|y - x\|^2 \end{aligned}$$

Therefore

$$f(y) = f_1(y) + f_2(y) \leq f_1(x) + f_1(x)^T (y - x) + \frac{L}{2}\|y - x\|^2 + f_2(y).$$

Substituting $y = x_k = x_{k-1} - \alpha G_\alpha(x_{k-1}), x = x_{k-1}$ we have

$$f(x_k) \leq f_1(x_{k-1}) - \alpha f_1(x_{k-1})^T G_\alpha(x_{k-1}) + \frac{L\alpha^2}{2}\|G_\alpha\|^2 + f_2(x_{k-1} - tG_\alpha(x_{k-1})) \tag{3}$$

Now

$$\begin{aligned} x_{k-1} - \alpha G_\alpha(x_{k-1}) &= \arg\min_z \left\{ \frac{\|x_k - z\|^2}{2\alpha} + f_2(z) \right\} \\ &= \arg\min_z \left\{ \frac{\|x_{k-1} - \alpha \nabla f_1(x_{k-1}) - z\|^2}{2\alpha} + f_2(z) \right\} \\ &= \arg\min_z \left\{ \frac{\|x_{k-1}^2 - z\|^2}{2\alpha} + \nabla f_1^T (z - x_{k-1}) + f_2(z) \right\} \end{aligned}$$

This implies that there exist a $\nu \in \partial f_2(z)$, $f_2(y) \geq f_2(z) + \nu^T(y - z)$, $\forall y$, such that at minima $\nabla f_1(x) + \frac{1}{\alpha}(z - x) + \nu = 0$, but the minimum occurs at $z = x_{k-1} - \alpha G_\alpha(x_{k-1})$, thus

$$\begin{aligned} \nabla f_1(x_{k-1}) - G_\alpha(x_{k-1}) + \nu = 0, \nu \in \partial f_2(x_{k-1} - \alpha G_\alpha(x_{k-1})) \\ G_\alpha(x_{k-1}) - \nabla f_1(x_{k-1}) \in \partial f_2(x_{k-1} - \alpha G_\alpha(x_{k-1})) \end{aligned}$$

Since f_2 is convex,

$$\begin{aligned} f_2(x) &\geq f_2\left(x_{k-1} - \alpha G_\alpha(x)\right) + \left(G_\alpha(x_{k-1}) - \nabla f_1(x_{k-1})\right)^T \alpha G_t(x_{k-1}) \\ f_2\left(x_{k-1} - \alpha G_\alpha(x_{k-1})\right) &\leq f_2(x_{k-1}) - \alpha \left(G_\alpha(x_{k-1}) - \nabla g(x_{k-1})\right)^T G_\alpha(x_{k-1}) \end{aligned} \tag{4}$$

Substituting (4) in (3) we get

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

Since f is convex as well, $f(x_{k-1}) \leq f\left(x^*\right) + G_\alpha(x_{k-1})^T \left(x_{k-1} - x^*\right)$, substituting it above we geometry

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

□

Summing over iteration we have:

$$\begin{aligned} \sum_{i=1}^{k} \left(f\left(x_i - f\left(x^*\right)\right)\right) &\leq \frac{1}{2\alpha}\left(\left\|x_0 - x^*\right\|^2 - \left\|x_k - x^*\right\|^2\right) \\ &\leq \frac{1}{2\alpha}\left\|x_0 - x^*\right\|^2. \end{aligned} \tag{5}$$

Since the difference is non-increasing we have

$$f\left(x_k\right) - f\left(x^*\right) \leq \frac{1}{k}\sum_{i=1}^{k} f(x_i - f\left(x^*\right)) \leq \frac{1}{2\alpha k}\left\|x_0 - x^*\right\|^2. \tag{6}$$

Remark:

- This has the same convergence rate $O(\frac{1}{t})$ as that of gradient descent.

3 Q3

Depending on f, you can derive conditions for which the dual ascent is converging depending on α and f you can have linear convergence, or $O(1/t), O\left(1/t^2\right), O(1/\sqrt{t})\ldots$

- If f is m-strongly convex, then the dual ascent with constant step size $\alpha = m$ converges with the rate of $O(1/t)$.
- If $f \in \mathbf{S}^{1,1}_{m,L}$, then the dual ascent with step size $\alpha = \frac{2}{1/m+1/L}$ converges with the rate of $O(\log(1/t))$.

4 Q4

Homework. Can you think of a dual Newton's method? Can you think of a dual proximal method? When would you apply these methods and with which guarantees?