Subgradient Methods

Lecturer: Jiaming Liang *September 14, 2023*

1 Subgradient

Definition 1. *Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a proper function and let $x \in \mathrm{dom}(f)$. A vector $g \in \mathbb{R}^n$ is called a subgradient of f at x if*

$$f(y) \geq f(x) + \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n.$$

We denote a subgradient of f at x by $f'(x)$.

Definition 2. *The set of all subgradients of f at x is called the subdifferential of f at x and is denoted by $\partial f(x)$:*

$$\partial f(x) \equiv \{g \in \mathbb{R}^n : f(y) \geq f(x) + \langle g, y - x \rangle \quad \forall y \in \mathbb{R}^n\}.$$

If f is convex, then $\partial f(x) \neq \emptyset$. If f is convex and smooth, then $\partial f(x) = \{\nabla f(x)\}$.

2 Localization ideas

We are now interested in the following optimization problem

$$\min_{x \in Q} f(x)$$

where Q is a closed convex set, and the function f is convex on $\mathbb{R}^n$ but may not necessarily be smooth. As compared with the smooth problem, our goal is more challenging. Indeed, even in the simplest situation, when $Q \equiv \mathbb{R}^n$, the subgradient seems to be a poor replacement for the gradient of a smooth function. For example, we cannot be sure that the value of the objective function is decreasing in the direction $-f'(x)$. We cannot expect that $f'(x) \to 0$ as x approaches the solution of our problem (e.g., the absolute function).

Let us fix some optimal solution x_*. It follows from the convexity of f that

$$f(x_*) \geq f(x) + \langle f'(x), x_* - x \rangle \quad \forall x \in Q,$$

and hence that

$$\langle f'(x), x_* - x \rangle \leq f(x_*) - f(x) \leq 0.$$

This implies that for a fixed point $x \in Q$, the optimal solution $x_* \in Q$ lies in the half-space

$$H_x^- = \{y \in \mathbb{R}^n : \langle f'(x), y - x \rangle \leq 0\}.$$

Definition 3. *Let* $\{x_i\}_{i=0}^{\infty}$ *be a sequence in* Q*. Define*

$$S_k = \left\{x \in Q \mid \left\langle f'(x_i), x - x_i \right\rangle \leq 0, i = 0 \ldots k\right\}.$$

We call S_k the localization set generated by the sequence $\{x_i\}_{i=0}^{\infty}$. It is obvious that $x_* \in S_k$ and $S_{k+1} \subset S_k$. With more search points x_k, we can shrink the localization set and hence localize the optimal solution x_*. **This is also the key idea in the cutting-plane/Kelly's method, method of centers of gravity, ellipsoid method, and many others.**

Definition 4. *For some fixed* $\bar{x} \in \mathbb{R}^n$ *and any* $x \in \mathbb{R}^n$ *with* $f'(x) \neq 0$*, define*

$$v_f(\bar{x}; x) = \frac{1}{\|f'(x)\|}\langle f'(x), x - \bar{x}\rangle.$$

If $f'(x) = 0$*, then define* $v_f(\bar{x}; x) = 0$*.*

Geometrically, for a fixed $x \in \mathbb{R}^n$ with $f'(x) \neq 0$ and any $\bar{x} \in H_x^-$, $v_f(\bar{x}, x)$ is the distance from point $\bar{x}$ to the hyperplane

$$H_x = \{y \in R^n : \langle f'(x), y - x\rangle = 0\}.$$

Indeed, consider the point

$$\bar{x}_p = \bar{x} + v_f(\bar{x}, x)\frac{f'(x)}{\|f'(x)\|}. \tag{1}$$

Then

$$\langle f'(x), \bar{x}_p - x\rangle = \langle f'(x), \bar{x} - x\rangle - v_f(\bar{x}, x)\|f'(x)\| = 0$$

and $\|\bar{x}_p - \bar{x}\| = v_f(\bar{x}, x)$. Note that $\bar{x}_p \in H_x$ and it is the projection of $\bar{x}$ onto H_x.

Let

$$v_i = v_f\left(x_*; x_i\right)(\geq 0), \quad v_k^* = \min_{0 \leq i \leq k} v_i.$$

Thus,

$$v_k^* = \max\left\{r \in \mathbb{R}_{++} : \left\langle f'\left(x_i\right), x - x_i\right\rangle \leq 0,\ i = 0 \ldots k,\ \forall x \in B_2\left(x_*, r\right)\right\}.$$

This is the radius of the maximal ball centered at x_*, which is contained in the localization set S_k.

Lemma 1. *If a convex function* f *is* M*-Lipschitz continuous on* $B_2(\bar{x}, R)$ *with constant* $M > 0$*, then*

$$f(x) - f(\bar{x}) \leq M v_f(\bar{x}; x)$$

for all $x \in \mathbb{R}^n$ *with* $0 \leq v_f(\bar{x}; x) \leq R$*. Moreover,*

$$\min_{0 \leq i \leq k} f\left(x_i\right) - f_* \leq M v_k^*.$$

Proof. Consider $\bar{x}_p$ as in (1) which lies in H_x, then

$$f(\bar{x}_p) \geq f(x) + \langle f'(x), \bar{x}_p - x \rangle = f(x).$$

If f is Lipschitz continuous on $B_2(\bar{x}, R)$ and $0 \leq v_f(\bar{x}; x) \leq R$, then $\bar{x}_p \in B_2(\bar{x}, R)$. Hence,

$$f(x) - f(\bar{x}) \leq f(\bar{x}_p) - f(\bar{x}) \leq M\|\bar{x}_p - \bar{x}\| = M v_f(\bar{x}; x).$$

Thus, we have

$$f(x_i) - f(x_*) \leq M v_f(x_*; x_i) = M v_i.$$

Therefore,

$$\min_{0 \leq i \leq k} f(x_i) - f_* = \min_{0 \leq i \leq k} [f(x_i) - f_*] \leq M \min_{0 \leq i \leq k} v_i = M v_k^*.$$

□

3 Subgradient methods

3.1 Normalized variant

Algorithm 1 Subradient method (normalized)

Input: Initial point $x_0 \in Q$
for $k \geq 0$ **do**
  Step 1. Choose $h_k > 0$.
  Step 2. Compute $x_{k+1} = \operatorname{proj}_Q \left(x_k - h_k \frac{f'(x_k)}{\|f'(x_k)\|} \right)$.
end for

Theorem 1. *Assume f is convex and M-Lipschitz continuous on $B_2(x^*, R)$ with $R \geq \|x_0 - x_*\|$ and choose $h_k > 0$ for every $k \geq 0$. Then, the Subgradient Method (normalized) generates a sequence of points $\{x_k\}$ satisfying*

$$\min_{0 \leq i \leq k} f(x_i) - f_* \leq M \frac{R^2 + \sum_{i=0}^{k-1} h_i^2}{2 \sum_{i=0}^{k-1} h_i}, \quad \forall k \geq 0.$$

Proof. **Fact:** (See PS1 (P2)(d)) for any two points $x \in Q$ and $y \in \mathbb{R}^n$, we have

$$\|x - \operatorname{proj}_Q(y)\|^2 + \|\operatorname{proj}_Q(y) - y\|^2 \leq \|x - y\|^2.$$

Let $r_k := \|x_k - x_*\|$. Using the above inequality and convexity, we have

$$\begin{aligned} r_{k+1}^2 &= \left\| \operatorname{proj}_Q \left(x_k - h_k \frac{f'(x_k)}{\|f'(x_k)\|} \right) - x_* \right\|^2 \\ &\leq \left\| x_k - h_k \frac{f'(x_k)}{\|f'(x_k)\|} - x_* \right\|^2 \\ &= r_k^2 - 2h_k \left\langle \frac{f'(x_k)}{\|f'(x_k)\|}, x_k - x_* \right\rangle + h_k^2 \\ &= r_k^2 - 2h_k v_k + h_k^2. \end{aligned}$$

Summing up these inequalities for $i = 0, \ldots, k-1$, we get

$$r_0^2 + \sum_{i=0}^{k-1} h_i^2 \geq 2 \sum_{i=0}^{k-1} h_i v_i + r_k^2 \geq 2 v_{k-1}^* \sum_{i=0}^{k-1} h_i.$$

Thus,

$$v_{k-1}^* \leq \frac{R^2 + \sum_{i=0}^{k-1} h_i^2}{2 \sum_{i=0}^{k-1} h_i}.$$

Since $v_{k-1}^* \leq v_0 \leq \|x_0 - x_*\| \leq R$, the conclusion follows form Lemma 1. □

Two stepsize rules:

1. given a fixed number of iterations $K \geq 1$,

$$h_k = \frac{R}{\sqrt{K}}, \quad k = 0, \ldots, K-1;$$

2. given a solution accuracy $\varepsilon > 0$,

$$h_k = \frac{\varepsilon}{M}, \quad k \geq 0.$$

For option 1, using Theorem 1, we have

$$\min_{0 \leq i \leq K-1} f(x_i) - f_* \leq \frac{MR}{\sqrt{K}}.$$

For option 2, using Theorem 1, we have

$$\min_{0 \leq i \leq K-1} f(x_i) - f_* \leq \frac{M^2 R^2}{2\varepsilon K} + \frac{\varepsilon}{2}.$$

Observation: the two options are equivalent in order to find an ε-solution.

3.2 Unnormalized variant

Algorithm 2 Subradient method (unnormalized)

Input: Initial point $x_0 \in Q$
for $k \geq 0$ **do**
  Step 1. Choose h_k.
  Step 2. Compute $x_{k+1} = \text{proj}_Q\left(x_k - h_k f'(x_k)\right)$.
end for

Theorem 2. (Polyak stepsize) *Suppose we know f_* (and we do in certain cases). Assume f is convex and M-Lipschitz continuous on $B_2\left(x^*, R\right)$ with $R \geq \|x_0 - x_*\|$ and choose*

$$h_k = \frac{f(x_k) - f_*}{\|f'(x_k)\|^2}, \quad k \geq 0.$$

Then, the Subgradient Method (unnormalized) generates a sequence of points $\{x_k\}$ satisfying

$$\min_{0 \leq i \leq k} f\left(x_i\right) - f_* \leq \frac{MR}{\sqrt{k}}, \quad \forall k \geq 0.$$

Proof. Let $r_k := \|x_k - x_*\|$. Recall: for any two points $x \in Q$ and $y \in \mathbb{R}^n$, we have

$$\|x - \text{proj}_Q(y)\|^2 + \|\text{proj}_Q(y) - y\| \leq \|x - y\|^2.$$

Using the above inequality and convexity, we have

$$\begin{aligned}
r_{k+1}^2 &= \left\|\text{proj}_Q\left(x_k - h_k f'(x_k)\right) - x_*\right\|^2 \\
&\leq \left\|x_k - h_k f'(x_k) - x_*\right\|^2 \\
&= r_k^2 - 2h_k\left\langle f'\left(x_k\right), x_k - x_*\right\rangle + h_k^2\left\|f'\left(x_k\right)\right\|^2 \\
&\leq r_k^2 - 2h_k[f(x_k) - f_*] + h_k^2\left\|f'\left(x_k\right)\right\|^2.
\end{aligned}$$

It follows from the Polyak's stepsize rule of h_k that

$$r_{k+1}^2 \leq r_k^2 - \frac{[f(x_k) - f_*]^2}{\|f'(x_k)\|^2},$$

and hence that $r_{k+1} < r_k < r_0 \leq R$. It follows from the Lipschitz continuity assumption and Theorem 3.61 of AB that $\|f'(x_k)\| \leq M$ that

$$r_{k+1}^2 \leq r_k^2 - \frac{[f(x_k) - f_*]^2}{\|f'(x_k)\|^2} \leq r_k^2 - \frac{[f(x_k) - f_*]^2}{M^2},$$

and

$$r_k^2 \leq r_0^2 - \frac{\sum_{i=0}^{k-1}[f(x_i) - f_*]^2}{M^2}.$$

Finally, we have

$$k \left[\min_{0 \le i \le k-1} f(x_i) - f_* \right]^2 \le \sum_{i=0}^{k-1} [f(x_i) - f_*]^2 \le M^2 r_0^2 \le M^2 R^2,$$

and

$$\min_{0 \le i \le k-1} f(x_i) - f_* \le \frac{MR}{\sqrt{k}}.$$

$\square$