Subgradient Methods

R.D.C. Monteiro[1]

[1]School of Industrial and Systems Engineering
Georgia Institute of Technology

January 17, 2024

Assume $f \in \mathrm{Conv}(\mathbb{R}^n)$ and $\bar{x} \in \mathrm{dom}\, f$

Definition

$s \in \mathbb{R}^n$ is called a **subgradient** of f at $\bar{x}$ if

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

The set of all subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$ and the point-to-set map $\partial f(\cdot)$ is called the **subdifferential** of f

Let $f \in \mathrm{Conv}(\mathbb{R}^n)$, $\bar{x} \in \mathrm{dom}\, f$ and $\varepsilon \geq 0$ be given

Definition

$s \in \mathbb{R}^n$ is called an ε-**subgradient** of f at $\bar{x}$ if

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

The set of all ε-subgradients of f at $\bar{x}$ is denoted by $\partial_\varepsilon f(\bar{x})$ and the point-to-set map $\partial_\varepsilon f(\cdot)$ is called the ε-**subdifferential** of f

Proposition

The following statements hold:

a) $\bar{x}$ *is an* ε*-minimizer of* $f_* := \inf\{f(x) : x \in \mathbb{R}^n\}$, *i.e.,*

$$f(\bar{x}) - f_* \leq \varepsilon$$

if and only if $0 \in \partial_\varepsilon f(\bar{x})$

b) $\partial_\varepsilon f(\bar{x})$ *is a closed convex set (possibly empty)*

Subgradient method

Consider the set optimization problem

$$f_* = \min\{(f + \delta_X)(x) : x \in \mathbb{R}^n\} = \min\{f(x) : x \in X\}$$

Assumption:

- $X \subset \mathbb{R}^n$ is nonempty closed convex and f is convex on X
- there exists $M \geq 0$ with the following property: for every $x \in X$, there exists $s(x) \in \partial f(x)$ such that $\|s(x)\| \leq M$
- optimal solution set X_* is nonempty, and hence $f_* \in \mathbb{R}$

Subgradient method (SM)

0) $x_0 \in X$ is given
1) For $k = 0, 1, 2, \ldots,$ do
    - choose stepsize $\lambda_k > 0$ and set $s_k = s(x_k)$
    - $x_{k+1} = Proj_X(x_k - \lambda_k s_k)$

Obs: SM is not descent, i.e., it does not necessarily satisfy $f(x_{k+1}) < f(x_k)$ for every k.

Example: For $a \in (0,1)$, consider

$$f(x_1, x_2) := |x_1| + a|x_2|$$

whose global min is $(0,0)$. Take $x = (0,1)$. Then

$$f(0,1) = a, \quad \partial f(0,1) = [-1,1] \times \{a\}$$

and hence $s_\eta = (\eta, a) \in \partial f(0,1)$ for all $\eta \in [-1,1]$. Now, for any $\lambda \in (0, a^{-1})$, we have

$$\begin{aligned} f(x - \lambda s_\eta) &= f((0,1) - \lambda(\eta, a)) = f(-\lambda\eta, 1 - \lambda a) \\ &\geq \lambda|\eta| + a(1 - \lambda a) = a + \lambda(|\eta| - a^2). \end{aligned}$$

Hence, if $|\eta| > a^2$, have

$$f(x - \lambda s_\eta) > a = f(x) \quad \forall \lambda \in (0, a^{-1})$$

which shows that s_η is not a descent direction

Key result

Proposition

For every $k \geq 0$ *and* $x_* \in X_*$, *have*

$$2\lambda_k \Delta f_k \leq \|x_k - x_*\|^2 - \|x_{k+1} - x_*\|^2 + \lambda_k^2 \|s_k\|^2$$

where

$$\Delta f_k := f(x_k) - f_* \geq 0$$

Hence, if $\lambda_k > 0$ is such that

$$\lambda_k < \frac{2\Delta f_k}{\|s_k\|^2}$$

then $\|x_k - x_*\| > \|x_{k+1} - x_*\|$

Key result (continued)

Let d_0 denote the distance of x_0 to X_*, i.e.,

$$d_0 := \|x_0 - Proj_{X_*}(x_0)\|$$

Proposition

For every $K \geq 0$ and $x_* \in X_*$

$$2 \sum_{k=0}^{K} \lambda_k \Delta f_k \leq \|x_0 - x_*\|^2 - \|x_{K+1} - x_*\|^2 + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2$$

As a consequence,

$$2 \sum_{k=0}^{K} \lambda_k \Delta f_k \leq d_0^2 + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2$$

Last part follows from first one with $x_* = Proj_{X_*}(x_0)$

SM with Polyak stepsize rule

$$\lambda_k = \frac{\Delta f_k}{\|s_k\|^2} \quad \forall k \geq 0$$

Then

$$\lambda_k^2 \|s_k\|^2 = \lambda_k \Delta f_k$$

and hence

$$2 \sum_{k=0}^{K} \lambda_k \Delta f_k \leq \|x_0 - x_*\|^2 - \|x_{K+1} - x_*\|^2 + \sum_{k=0}^{K} \lambda_k \Delta f_k$$

or equivalently,

$$\sum_{k=0}^{K} \lambda_k \Delta f_k \leq \|x_0 - x_*\|^2 - \|x_{K+1} - x_*\|^2 \tag{1}$$

The above ineq with $x_* = Proj_{X_*}(x_0)$ implies that

$$d_0^2 \geq \sum_{k=0}^{K} \lambda_k \Delta f_k$$

Moreover,

$$\lambda_k \Delta f_k = \frac{\Delta f_k^2}{\|s_k\|^2} \geq \frac{\Delta f_k^2}{M^2}$$

Hence,

$$d_0^2 \geq \frac{1}{M^2} \sum_{k=0}^{K} \Delta f_k^2 \geq \frac{1}{M^2}(K+1)\left(\min_{k \leq K} \Delta f_k\right)^2$$

Thus

$$\left(\min_{k \leq K} \Delta f_k\right)^2 \leq \frac{M^2 d_0^2}{K+1} \leq \frac{M^2 d_0^2}{K}$$

Theorem

- $\{x_k\}$ *is bounded*
- *for any tolerance* $\varepsilon > 0$ *and for every*

$$K \geq \frac{d_0^2 M^2}{\varepsilon^2}$$

we have

$$\theta_K := \min_{k \leq K} \Delta f_k \leq \varepsilon$$

Obs: This SM variant has ε-iteration complexity $\mathcal{O}(M^2 d_0^2 / \varepsilon^2)$

Drawback: Polyak rule requires f_*

Exerc: Show that $\{x_k\}$ converges to some point in X_*

SM with constant stepsize

$$\lambda_k = \lambda > 0 \quad \forall k \geq 0$$

Then

$$\lambda_k^2 \|s_k\|^2 = \lambda^2 \|s_k\|^2 \leq \lambda^2 M^2$$

and hence

$$2\lambda\theta_K(K+1) \leq 2\lambda \sum_{k=0}^{K} \Delta f_k \leq d_0^2 + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2$$
$$\leq d_0^2 + (K+1)\lambda^2 M^2$$

So,

$$2\theta_K \leq \lambda M^2 + \frac{d_0^2}{\lambda(K+1)}$$

Take $\lambda = \varepsilon / M^2$. Then

$$2\theta_K \leq \varepsilon + \frac{d_0^2 M^2}{\varepsilon K}$$

So, if

$$K \geq \frac{d_0^2 M^2}{\varepsilon^2}$$

then $\theta_K \leq \varepsilon$

Obs: It is not true that θ_K converges to 0 nor that $\{x_K\}$ is bounded as $K \to \infty$

SM with adaptive stepsize

Take

$$\lambda_k = \frac{\varepsilon}{\|s_k\|^2} \quad \forall k \geq 0$$

Observe that $\lambda_k \geq \varepsilon / M^2 = \lambda$ where λ is the constant stepsize. Have

$$\begin{aligned} 2\varepsilon\theta_K \sum_{k=0}^{K} \frac{1}{\|s_k\|^2} &\leq 2\varepsilon \sum_{k=0}^{K} \frac{\Delta f_k}{\|s_k\|^2} = 2 \sum_{k=0}^{K} \lambda_k \Delta f_k \\ &\leq d_0^2 + \sum_{k=0}^{K} \lambda_k^2 \|s_k\|^2 = d_0^2 + \varepsilon^2 \sum_{k=0}^{K} \frac{1}{\|s_k\|^2} \end{aligned}$$

Hence

$$2\theta_K \leq \frac{d_0^2}{\varepsilon \sum_{k=0}^{K} \frac{1}{\|s_k\|^2}} + \varepsilon \leq \frac{d_0^2 M^2}{\varepsilon(K+1)} + \varepsilon$$

Thus, $K \geq (d_0^2 M^2)/\varepsilon^2$ implies that $\theta_K \leq \varepsilon$

SM with diminishing stepsize

The key result and the second assumption imply

$$2\sum_{k=0}^{K}\lambda_k\Delta f_k \le d_0^2 + \sum_{k=0}^{K}\lambda_k^2\|s_k\|^2 \le d_0^2 + M^2\sum_{k=0}^{K}\lambda_k^2$$

Hence

$$2\theta_K \le \frac{d_0^2 + M^2\sum_{k=0}^{K}\lambda_k^2}{\sum_{k=0}^{K}\lambda_k}$$

A sufficient condition for $\theta_K \to 0$ as $K \to \infty$ is that

$$\sum_{k=0}^{K}\lambda_k \to \infty, \quad \frac{\sum_{k=0}^{K}\lambda_k^2}{\sum_{k=0}^{K}\lambda_k} \to 0$$

SM with diminishing stepsize (continued)

Taking $\lambda_k = a/\sqrt{k}$ where $a > 0$ implies that

$$\sum_{k=0}^{K} \lambda_k \approx 2a\sqrt{K}, \quad \sum_{k=0}^{K} \lambda_k^2 \approx a^2 \log K$$

and hence

$$2\theta_K \leq \frac{1}{2\sqrt{K}} \left(\frac{d_0^2}{a} + aM^2 \log K \right)$$

Taking $a = d_0/M$, have

$$2\theta_K \leq \frac{d_0 M}{2\sqrt{K}} (1 + \log K)$$

Proof of Key result

Lemma

(Non-expansiveness of the projection operator) If X is a nonempty closed convex set, then for every $x, x' \in X$, have

$$\|Proj_X(x) - Proj_X(x')\| \leq \|x - x'\|$$

To show the key result, it suffices to show that

Proposition

For every $k \geq 0$ and $x \in X$, have

$$2\lambda_k[f(x_k) - f(x)] \leq \|x_k - x\|^2 - \|x_{k+1} - x\|^2 + \lambda_k^2 \|s_k\|^2$$

The key result follows from the above one with $x = x_*$.

Proof of Key result (continued)

Proof: Let $x \in X$ be given. Have

$$\begin{aligned} \|x_{k+1} - x\| &= \|Proj_X(x_k - \lambda_k s_k) - x\| \\ &= \|Proj_X(x_k - \lambda_k s_k) - Proj_X(x)\| \\ &\leq \|x_k - \lambda_k s_k - x\| \end{aligned}$$

Hence

$$\begin{aligned} \|x_{k+1} - x\|^2 &= \|x_k - x\|^2 + \lambda_k^2 \|s_k\|^2 + 2\lambda_k \langle s_k, x - x_k \rangle \\ &\leq \|x_k - x\|^2 + \lambda_k^2 \|s_k\|^2 + 2\lambda_k [f(x) - f(x_k)] \end{aligned}$$