Noname manuscript No.
(will be inserted by the editor)

Lipschitz-free Projected Subgradient Method with Time-varying Step-size

Yong Xia, Yanhao Zhang, Zhihan Zhu

Received: date / Accepted: date

Abstract We introduce a novel time-varying step-size for the classical projected subgradient method, offering optimal ergodic convergence. Importantly, this approach does not depend on the Lipschitz assumption of the objective function, thereby broadening the convergence result of projected subgradient method to non-Lipschitz case.

Keywords Projected subgradient method · Step-size · Nonsmooth convex optimization

Mathematics Subject Classification (2000) 90C25, 90C30

1 Introduction

To tackle the nonsmooth convex optimization problem

$$x^* \in \mathrm{argmin}_{x\in\mathcal{X}} f(x),$$

where $\mathcal{X} \subset \mathbb{R}^n$ is a compact convex set enclosed within the Euclidean ball $B(x^*, R)$, and f is a (possibly nonsmooth) convex function, the traditional projected subgradient method (PSG) is employed as follows:

$$\begin{cases} y_{s+1} = x_s - \eta_s g_s, \text{ where } g_s \in \partial f(x_s), \\ x_{s+1} = \mathrm{argmin}_{x\in\mathcal{X}} \|x - y_{s+1}\|, \end{cases}$$

where $\|\cdot\|$ denotes the Euclidean norm throughout this paper.

In the literature, the following common Lipschitz assumption on f is made:

Assumption 1 *There exists an $L > 0$ such that for any $g \in \partial f(x) \neq \emptyset$ and $x \in \mathcal{X}$, it holds that $\|g\| \leq L$.*

Yong Xia, Yanhao Zhang, Zhihan Zhu
School of Mathematical Sciences, Beihang University, Beijing 100191, People's Republic of China. E-mail: {yxia,yanhaozhang,zhihanzhu}@buaa.edu.cn

It is well-established that employing a constant step-size of

$$\eta_s \equiv \frac{R}{L\sqrt{t}},\ s = 1, \cdots, t,$$

allows PSG to attain an optimal ergodic convergence rate, which is given by

$$f\left(\frac{\sum_{s=1}^{t} x_s}{t}\right) - f(x^*) \leq \frac{RL}{\sqrt{t}},$$

see, for example, [1,2].

Recently, a time-varying step-size formula, as presented in [1,2], given by

$$\eta_s = \frac{R}{L\sqrt{s}},\ s = 1, \cdots, t, \tag{1}$$

has been proven to ensure the optimal convergence rate of PSG as well, as stated in [3, Corollary 3.2]. The following succinct result concerning PSG with the step-size given by (1) is credited to [4]:

$$f\left(\frac{\sum_{s=1}^{t} x_s}{t}\right) - f(x^*) \leq \frac{3RL}{2\sqrt{t}}. \tag{2}$$

In [1], a more practical subgradient-normalized time-varying step-size is further examined, given by

$$\eta_s = \frac{R}{\|g_s\|\sqrt{s}},\ s = 1, \cdots, t, \tag{3}$$

which notably does not necessitate the knowledge of the Lipschitz constant beforehand. However, to guarantee the convergence of PSG, Assumption 1 is still required. Additionally, the convergence rate achieved is merely sub-optimal.

The key contribution of this note is to introduce a subtle variation to the step-size (3), which enables us to establish the optimal ergodic convergence rate of PSG, notably without requiring Assumption 1.

2 The main result

We derive the following result without the need for Assumption 1.

Theorem 1 *PSG with the following step-size*

$$\eta_s = \min\left\{\eta_{s-1},\ \frac{R}{\|g_s\|\sqrt{s}}\right\}\ (\eta_0 = +\infty),\ s = 1, \cdots, t, \tag{4}$$

satisfies

$$f\left(\frac{\sum_{s=1}^{t} x_s}{t}\right) - f(x^*) \leq \frac{3R}{2\sqrt{t}} \cdot \max_{s=1,\cdots,t} \|g_s\|. \tag{5}$$

Proof Consider PSG with the step-size (4). We have

$$
\begin{aligned}
f(x_s) - f(x^*) &\le g_s^T(x_s - x^*) \quad \text{(by the definition of subgradient)}\\
&= \frac{1}{\eta_s}(x_s - y_{s+1})^T(x_s - x^*)\\
&= \frac{1}{2\eta_s}(\|x_s - y_{s+1}\|^2 + \|x_s - x^*\|^2 - \|y_{s+1} - x^*\|^2) \qquad (6)\\
&= \frac{1}{2\eta_s}(\|x_s - x^*\|^2 - \|y_{s+1} - x^*\|^2) + \frac{\eta_s}{2}\|g_s\|^2\\
&\le \frac{1}{2\eta_s}(\|x_s - x^*\|^2 - \|x_{s+1} - x^*\|^2) + \frac{R}{2\sqrt{s}}\|g_s\|, \qquad (7)
\end{aligned}
$$

where (6) is derived from the fundamental identity $2a^Tb = \|a\|^2 + \|b\|^2 - \|a - b\|^2$, and (7) holds due to η_s as defined in (4) and the fact that

$$\|y_{s+1} - x^*\|^2 \ge \|x_{s+1} - x^*\|^2,$$

which is a direct consequence of the projection theorem.

Consequently, we have

$$
\begin{aligned}
& f\left(\frac{\sum_{s=1}^t x_s}{t}\right) - f(x^*)\\
\le\ & \frac{1}{t}\sum_{s=1}^t (f(x_s) - f(x^*)) \quad \text{(Jensen's inequality)}\\
\le\ & \frac{1}{t}\sum_{s=1}^t \frac{1}{2\eta_s}(\|x_s - x^*\|^2 - \|x_{s+1} - x^*\|^2) + \frac{1}{t}\sum_{s=1}^t \frac{R}{2\sqrt{s}}\|g_s\| \quad \text{(by (7))}\\
=\ & \frac{1}{2t\eta_1}\|x_1 - x^*\|^2 + \frac{1}{2t}\sum_{s=2}^t\left(\frac{1}{\eta_s} - \frac{1}{\eta_{s-1}}\right)\|x_s - x^*\|^2 - \frac{1}{2t\eta_t}\|x_{t+1} - x^*\|^2 + \frac{1}{t}\sum_{s=1}^t \frac{R}{2\sqrt{s}}\|g_s\|\\
\le\ & \frac{R^2}{2t\eta_1} + \frac{R^2}{2t}\sum_{s=2}^t\left(\frac{1}{\eta_s} - \frac{1}{\eta_{s-1}}\right) + \frac{1}{t}\sum_{s=1}^t \frac{R}{2\sqrt{s}}\|g_s\| \quad \left(\text{since } \frac{1}{\eta_s} - \frac{1}{\eta_{s-1}} \ge 0\right)\\
\le\ & \frac{R^2}{2t\eta_t} + \frac{R}{2t}\left(\max_{s=1,\cdots,t}\|g_s\|\right)\sum_{s=1}^t \frac{1}{\sqrt{s}}\\
\le\ & \frac{R^2}{2t\eta_t} + \frac{R}{\sqrt{t}}\max_{s=1,\cdots,t}\|g_s\| \quad \left(\text{since } \sum_{s=1}^t \frac{1}{\sqrt{s}} < 2\sqrt{t}\right)\\
=\ & \frac{R}{\sqrt{t}}\left(\frac{1}{2}\max_{s=1,\cdots,t}\|g_s\|\sqrt{\frac{s}{t}} + \max_{s=1,\cdots,t}\|g_s\|\right) \quad \text{(by the definition of } \eta_t \text{ (4))}\\
\le\ & \frac{3R}{2\sqrt{t}}\max_{s=1,\cdots,t}\|g_s\|.
\end{aligned}
$$

The proof is complete. □

Remark 1 *In the scenario where $\partial f(x_s)$ is not a singleton, Theorem 1 suggests that selecting g_s from $\partial f(x_s)$ with the minimal norm may possibly enhance the convergence.*

Remark 2 *Given Assumption 1, Theorem 1 allows us to swiftly attain the optimal ergodic convergence result of* (2).

Remark 3 *Even when $\|g_s\|$ is unbounded (i.e., Assumption 1 is violated), convergence of PSG can still be assured by Theorem 1, as long as the growth rate of $\|g_s\|$ during the iteration strictly stays within $\mathcal{O}(\sqrt{s})$.*

Remark 4 *We can apply the same proof techniques to extend the non-Lipschitz convergence result to weak ergodic convergence of PSG in [4], mirror descent and other schemes with time-varying step sizes for solving nonsmooth convex optimization, see [1,2].*

Funding

This research was supported by National Natural Science Foundation of China under grant 12171021, and the Fundamental Research Funds for the Central Universities.

Data Availability

The manuscript has no associated data.

References

1. Y. Nesterov, Lectures on convex optimization, volume 137, Springer, 2018.
2. S. Bubeck, Convex optimization: Algorithms and complexity. Foundations and Trends Trends® in Machine Learning, 8(3-4): 231-357, 2015.
3. G. Lan, First-order and Stochastic Optimization Methods for Machine Learning, Springer-Nature, 2020
4. Z. Zhu, Y. Zhang, Y. Xia, Convergence rate of projected subgradient method with time-varying step-sizes, Optimization Letters, 2024. https://doi.org/10.1007/s11590-024-02142-9