Oracle Complexity of Single-Loop Switching Subgradient Methods for Non-Smooth Weakly Convex Functional Constrained Optimization

Yankun Huang[1], Qihang Lin[1]
[1]Department of Business Analytics, University of Iowa
Supported by NSF Grant No 2147253

Non-Smooth Weakly Convex Constrained Optimization

Problem formulation:

$$f^* \equiv \min_{\mathbf{x}} f(\mathbf{x}) \quad \text{s.t.} \quad g(\mathbf{x}) \leq 0 \tag{P}$$

Assumption 1

- *f and g are real-valued and M-Lipschitz continuous (but not necessarily smooth).*
- *f and g are ρ-weakly convex (i.e., $f(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x}\|^2$ and $g(\mathbf{x}) + \frac{\rho}{2}\|\mathbf{x}\|^2$ are convex).*
- $\underline{f} := \inf f(\mathbf{x}) > -\infty$.

Near ϵ-Stationarity

Following the literature on weakly convex optimization (Davis and Drusvyatskiy, 2019, Davis and Grimmer, 2019, Ma et al., 2020, Jia and Grimmer, 2022) consider the following near ϵ-stationarity.

Definition

$\mathbf{x}$ *is an* ϵ-**stationary point** *if there exist* $\lambda \geq 0$, $\zeta_f \in \partial f(\mathbf{x})$ *and* $\zeta_g \in \partial g(\mathbf{x})$ *s.t.*

$$\left\|\zeta_f + \lambda \zeta_g\right\| \leq \epsilon, \quad |\lambda g(\mathbf{x})| \leq \epsilon^2, \quad g(\mathbf{x}) \leq \epsilon^2, \quad \lambda \geq 0.$$

Definition

$\mathbf{x}$ *is a* **nearly** ϵ-**stationary point** *if there exists* $\widehat{\mathbf{x}}$ *s.t.* $\widehat{\mathbf{x}}$ *is an* ϵ*-stationary point and* $\|\widehat{\mathbf{x}} - \mathbf{x}\| \leq \epsilon$.

Existing Techniques

- Solving (P) means to find a nearly ϵ-stationary point of (P).
- Existing double-loop methods (Ma et al., 2020, Boob et al., 2023, Jia and Grimmer, 2022) find a nearly ϵ-stationary point of (P) with oracle complexity $O(1/\epsilon^4)$ under different CQs.
- The **oracle complexity** is the total number of times for which the algorithm queries the subgradient or function value of f or g.

Main contribution

- Study the classical **switching subgradient (SSG) method** (Polyak, 1967) and show that,
- as a single-loop first-order algorithm, SSG can also find a nearly ϵ-stationary point of (P) with oracle complexity $O(1/\epsilon^4)$.
- Invent a **switching step-zize rule** to accompany the switching subgradient.

Switching Subgradient Method

Algorithm 1: Switching Subgradient (SSG) method

Input: $\mathbf{x}^{(0)}$, T, step-sizes $\eta_t > 0$ and tolerances $\epsilon_t \geq 0$.
for $t = 0, 1, \cdots, T-1$ **do**
  if $g(\mathbf{x}^{(t)}) \leq \epsilon_t$ **then**
    $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta_t \zeta_f^{(t)}$ for some $\zeta_f^{(t)} \in \partial f(\mathbf{x}^{(t)})$ and, $I = I \cup \{t\}$.
  else
    $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} - \eta_t \zeta_g^{(t)}$ for some $\zeta_g^{(t)} \in \partial g(\mathbf{x}^{(t)})$ and, $J = J \cup \{t\}$.
  end
end
Output: $\mathbf{x}^{(\tau)}$ where τ is sampled from $I \cup J$ using $\text{Prob}(\tau = t) = \eta_t / \sum_{s \in I \cup J} \eta_s$.

Technical Lemmas

Assumption 2 (Uniform Slater's condition in Ma et al. (2020))

There exist $\bar{\epsilon} > 0$, $\theta > 0$ *and* $\bar{\rho} > \rho$ *such that Slater's condition*

$$\exists\, \mathbf{y} \text{ s.t. } g(\mathbf{y}) + \frac{\bar{\rho}}{2}\|\mathbf{y} - \mathbf{x}\|^2 \leq -\theta$$

holds for any $\mathbf{x}$ *satisfying* $g(\mathbf{x}) \leq \bar{\epsilon}^2$. *(This is the CQ for SSG in our results.)*

Denote: $g_+(\mathbf{x}) = \max\{g(\mathbf{x}), 0\}$, $\mathcal{L} = \{\mathbf{x} \mid g(\mathbf{x}) = 0\}$, $\mathcal{S} = \{\mathbf{x} \mid g(\mathbf{x}) \leq 0\}$.

Lemma

1. *Subgradient of* g *is bounded away from zero on* $\mathcal{L}$*:*

$$\min_{\zeta_g \in \partial g(\mathbf{x})} \|\zeta_g\| \geq \nu := \sqrt{2\theta(\hat{\rho} - \rho)},\ \forall\, \mathbf{x} \in \mathcal{L} \text{ for some } \hat{\rho} \in (\rho, \bar{\rho}].$$

2. *Local error bound holds:*

$$(\nu/2) \cdot \operatorname{dist}(\mathbf{x}, \mathcal{S}) \leq g_+(\mathbf{x}) \quad \text{if} \quad \operatorname{dist}(\mathbf{x}, \mathcal{S}) \leq \nu/\rho.$$

Oracle Complexity

When $g(\mathbf{x}^{(t)}) > \epsilon_t$, SSG is essentially solving a sharp weakly convex unconstrained problem

$$\mathcal{S} = \arg\min_{\mathbf{x}} g_+(\mathbf{x}),$$

and thus Davis et al. (2018) suggests applying the Polyak's step-size in this case for the Q-linear convegence on $\text{dist}(\mathbf{x}^{(t)}, \mathcal{S})$.

Theorem 1

Suppose $\hat{\rho} \in (\rho, \bar{\rho}]$ and $\epsilon \leq \bar{\epsilon}$. Let $\mathbf{x}^{(0)} \in \mathcal{S}$, $\epsilon_t = \frac{\nu}{4} \min\left\{\epsilon^2/M, \nu/(4\rho)\right\}$ and

$$\eta_t = \begin{cases} \frac{\nu}{4M^2} \min\left\{\epsilon^2/M, \nu/(4\rho)\right\} & \text{if } g(\mathbf{x}^{(t)}) \leq \epsilon_t \\ g(\mathbf{x}^{(t)})/\|\zeta_g^{(t)}\|^2 & \text{if } g(\mathbf{x}^{(t)}) > \epsilon_t. \end{cases}$$

Then $g(\mathbf{x}^{(t)}) \leq \epsilon^2$, $\forall t \geq 0$, and SSG finds a nearly ϵ-stationary point of (P) if

$$T \geq \frac{8M^2\left(f(\mathbf{x}^{(0)}) - \underline{f} + 3M^2/(2\hat{\rho})\right)}{\hat{\rho}(1 + 2M/\nu)\nu\epsilon^2 \min\left\{\epsilon^2/M, \nu/(4\rho)\right\}} = O(1/\epsilon^4).$$

The choice of step-sizes $\{\eta_t\}_{t\geq 0}$ in Theorem 1 shows the switching step-zize rule.

[1] Digvijay Boob, Qi Deng, and Guanghui Lan. Stochastic first-order methods for convex and nonconvex functional constrained optimization. *Mathematical Programming*, 197(1):215–279, 2023. doi:10.1007/s10107-021-01742-y.

[2] Damek Davis and Dmitriy Drusvyatskiy. Stochastic model-based minimization of weakly convex functions. *SIAM Journal on Optimization*, 29(1):207–239, 2019. doi:10.1137/18M1178244.

[3] Damek Davis and Benjamin Grimmer. Proximally guided stochastic subgradient method for nonsmooth, nonconvex problems. *SIAM Journal on Optimization*, 29(3):1908–1930, 2019. doi:10.1137/17M1151031.

[4] Damek Davis, Dmitriy Drusvyatskiy, Kellie J. MacPhee, and Courtney Paquette. Subgradient methods for sharp weakly convex functions. *Journal of Optimization Theory and Applications*, 179(3):962–982, 2018. doi:10.1007/s10957-018-1372-8.

[5] Zhichao Jia and Benjamin Grimmer. First-order methods for nonsmooth nonconvex functional constrained optimization with or without slater points. *arXiv preprint arXiv:2212.00927*, 2022. doi:10.48550/arXiv.2212.00927.

[6] Runchao Ma, Qihang Lin, and Tianbao Yang. Quadratically regularized subgradient methods for weakly convex optimization with weakly convex constraints. In *International Conference on Machine Learning*, pages 6554–6564. PMLR, 2020. URL http://proceedings.mlr.press/v119/ma20d.html.

[7] Boris T. Polyak. A general method for solving extremal problems. In *Doklady Akademii Nauk SSSR*, volume 174, pages 33–36, 1967. URL https://www.mathnet.ru/eng/dan33049.