Penalty-based Methods for Simple Bilevel Optimization under Hölderian Error Bounds

Pengyu Chen   Xu Shi   Rujun Jiang   Jiulin Wang

Fudan University
School of Data Science

November 11, 2024

Contents

Simple Bilevel Optimization

- Simple bilevel optimization (SBO):

$$\min_{x \in \mathbb{R}^n} F(x) \quad \text{s.t.} x \in \arg\min_{u \in X} G(u). \tag{P}$$

  where $F, G : \mathbb{R}^n \to \mathbb{R} \bigcup \{\infty\}$ are proper, convex, and lower semi-continuous functions.
- Challenge:
  - The feasible set $X_{\text{opt}} := \{x \mid \arg\min_{u \in X} G(u)\}$.
  - Main challenge: The implicit availability of X_{opt} makes it impossible to simply apply standard first-order methods.

Contents

Penalized Framework

SBO:

$$\min_{x \in \mathbb{R}^n} F(x) \quad \text{s.t.} x \in \arg\min_{u \in X} G(u). \tag{P}$$

Penalized SBO:

$$\min_{x \in \mathbb{R}^n} \Phi_\gamma(x) = F(x) + \gamma p(x), \tag{P$_\gamma$}$$

where $p(x) = G(x) - G^*$ is the residual function.

- $p(x) \geq 0$, and $p(x) = 0$ if and only if $x \in X_{\text{opt}}$.
- **Definition.** Given $\epsilon_f > 0$ and $\epsilon_g > 0$. We say $\tilde{x}^*$ is an (ϵ_f, ϵ_g)-optimal solution of (P) if it holds that $F(\tilde{x}^*) - F^* \leq \epsilon_f, \quad G(\tilde{x}^*) - G^* \leq \epsilon_g.$
- Define: $\tilde{x}^*_\gamma$ is an ϵ-optimal solution of (P_γ) if it satisfies the following inequality:

$$\Phi_\gamma(\tilde{x}^*_\gamma) - \Phi^*_\gamma \leq \epsilon.$$

Assumptions

Assumption 1.1 (Hölderian error bound):
The function $p : X \mapsto \mathbb{R}$ satisfies the Hölderian error bound with exponent $\alpha \geq 1$ and $\rho > 0$ on the lower-level optimal solution set X_{opt}, i.e.,

$$\rho p(x) \geq \mathrm{dist}(x, X_{\mathrm{opt}})^{\alpha}.$$

$G(\mathbf{x})$	Remarks	Name	α
$\max_{i\in[m]}\{\langle \mathbf{a}_i, \mathbf{x}\rangle - b_i\}$	$\mathbf{a}_i \in \mathbb{R}^n, i \in [m], b \in \mathbb{R}^m$	piece-wise maximum	1
$\|\mathbf{x}-\mathbf{x}_0\|_Q = \sqrt{(\mathbf{x}-\mathbf{x}_0)^{\mathrm{T}} Q (\mathbf{x}-\mathbf{x}_0)}$	$Q \in \mathbb{S}^n, Q \succ 0, \mathbf{x}_0 \in \mathbb{R}^n$	Q-norm	1
$\|\mathbf{x}-\mathbf{x}_0\|_p$	$\mathbf{x}_0 \in \mathbb{R}^n, p \geq 1$	ℓ_p-norm	1
$\|x\|_1 + \frac{\tau}{2}\|x\|^2$	$\tau > 0$	Elastic net	1 or 2[4]
$\|A\mathbf{x}-b\|^2$	$A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$	Least squares	2
$\frac{1}{m}\sum_{i=1}^m \log(1+\exp(-\mathbf{a}_i^{\mathrm{T}}\mathbf{x}b_i))$	$\mathbf{a}_i \in \mathbb{R}^n, i \in [m], b \in \mathbb{R}^m, A \in \mathbb{R}^{m\times n}$	Logistic loss	2
$\eta(\mathbf{x}) + \frac{\sigma}{2}\|\mathbf{x}\|^2$	η convex, $\sigma > 0$	Strongly-convex	2

Assumptions

SBO:

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n} & F(x) \triangleq f_1(x) + f_2(x) \\ \text{s.t.} & x \in \arg\min\limits_{u \in X} G(u) \triangleq g_1(u) + g_2(u), \end{array} \tag{P}$$

- **Assumption 1.2**
  The set $S := \bigcup_{x \in X_{\text{opt}}} \partial F(\mathbf{x})$ is bounded with a diameter $l_F := \max_{xi \in S} \|xi\|$.
  When the upper-level objective F is non-convex, we replace the assumption with the condition that the upper-level objective is Lipschitz continuous.

Penalized Framework

Relationship between (ϵ_f, ϵ_g)-optimal solution of (P) and ϵ-optimal solution of (P_γ):

- **Lemma 1.** (Motivated by [1][1]) Suppose that Assumptions 1.1 and 1.2 hold with $\alpha > 1$. Then, for any $\epsilon > 0$, a global solution of (P) is an ϵ-optimal solution of (P_γ) when $\gamma \geq \gamma^* = \rho l_F^\alpha (\alpha - 1)^{\alpha-1} \alpha^{-\alpha} \epsilon^{1-\alpha}$.
- **Lemma 2.** Suppose that Assumptions 1.1 and 1.2 hold with $\alpha = 1$ and $\gamma \geq \gamma^* = \rho l_F$. Then there is an **exact penalization**:
  - A global optimal solution of (P) is also a global optimal solution of (P_γ);
  - Conversely, a global optimal solution of (P_γ) is also a global optimal solution of (P).

* Note: we use γ^* to denote the lower bound of γ in both cases.

[1] H. Shen and T. Chen. On penalty-based bilevel gradient descent method. In Proceedings of the 40th International Conference on Machine Learning, volume 202 of Proceedings of Machine Learning Research, pages 30992–31015. PMLR, 2023.

Main Result

Relationship between (ϵ_f, ϵ_g)-optimal solution of (P) and ϵ-optimal solution of (P_γ):

- Suppose that Assumptions 1.1 and 1.2 hold. For any given $\epsilon > 0$ and $\beta > 0$, let

$$\gamma = \gamma^* + \begin{cases} 2l_F^{\beta}\epsilon^{1-\beta} & \text{if } \alpha > 1, \\ l_F^{\beta}\epsilon^{1-\beta} & \text{if } \alpha = 1. \end{cases}$$

  If $\tilde{x}_\gamma^*$ is an ϵ-optimal solution of problem (P_γ), then $\tilde{x}_\gamma^*$ is an $(\epsilon, l_F^{-\beta}\epsilon^{\beta})$-optimal solution of problem (P).
- Suppose that the Assumptions 1.1 and 1.2 hold. Then, $\tilde{x}_\gamma^*$ satisfies the following suboptimality lower bound,

$$F(\tilde{x}_\gamma^*) - F^* \geq -l_F(\rho l_F^{-\beta}\epsilon^{\beta})^{\frac{1}{\alpha}}.$$

THANK YOU

THANK YOU FOR YOUR LISTENING!