Stochastic Regret Guarantees for Online Zeroth- and First-Order Bilevel Optimization

Parvin Nazari[1] Bojian Hou[2] Davoud Ataee Tarzanagh[3]
Li Shen[2] George Michailidis[4]

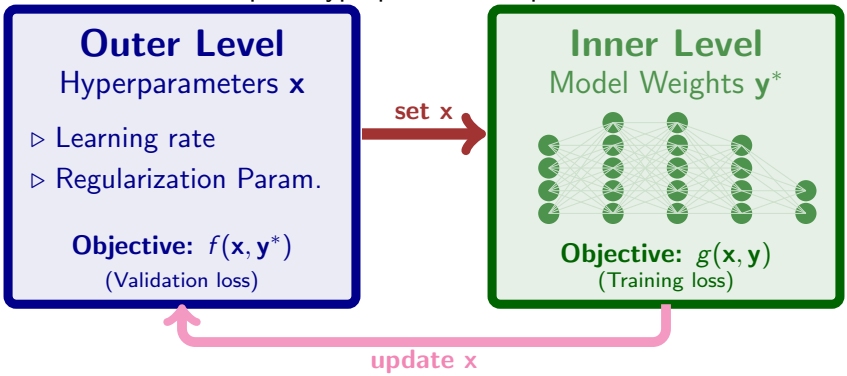
[1]Amirkabir University of Technology [2]University of Pennsylvania
[3]Samsung SDS Research America [4]UCLA

November 10, 2025

Bilevel Optimization (BO)

$$\min_{\mathbf{x}\in\mathbb{R}^{d_1}} f(\mathbf{x}, \mathbf{y}^*(\mathbf{x})) \quad \text{s.t.} \quad \mathbf{y}^*(\mathbf{x}) = \arg\min_{\mathbf{y}\in\mathbb{R}^{d_2}} g(\mathbf{x}, \mathbf{y})$$

Example: Hyperparameter Optimization



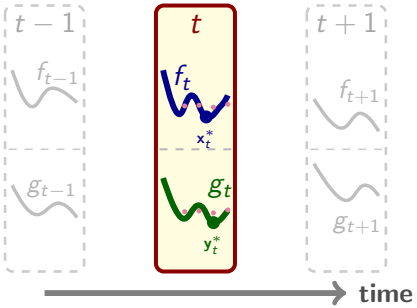
Online BO (OBO) as Leader-Follower Game

At each timestep $t \in [T]$: **Leader** chooses $\mathbf{x}_t \in \mathcal{X} \subset \mathbb{R}^{d_1}$ to minimize:

$$f_t(\mathbf{x}, \mathbf{y}_t^*(\mathbf{x})) := \mathbb{E}_{\xi_t \sim \mathcal{D}_{f,t}}[f_t(\mathbf{x}, \mathbf{y}_t^*(\mathbf{x}); \xi_t)]$$

Follower responds with $\mathbf{y}_t \in \mathbb{R}^{d_2}$ by solving:

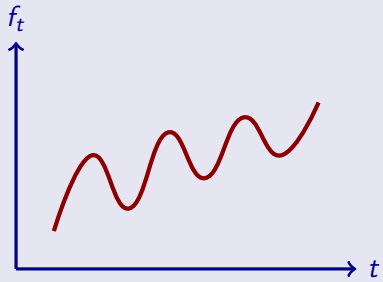
$$\mathbf{y}_t^*(\mathbf{x}) \in \arg\min_{\mathbf{y}} \left\{ g_t(\mathbf{x}, \mathbf{y}) := \mathbb{E}_{\zeta_t \sim \mathcal{D}_{g,t}}[g_t(\mathbf{x}, \mathbf{y}; \zeta_t)] \right\}$$



- Objectives f_t, g_t are time-varying, stochastic, and unknown in advance

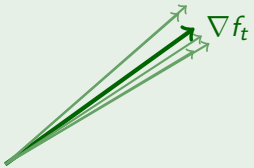
Key Challenges in OBO (1/2)

Time-Varying Objectives



Functions can change rapidly

Stochastic Observations

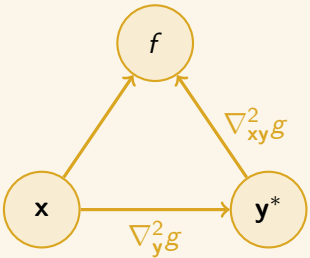


Minibatch noise (with variance σ)

Only noisy gradient estimates available from finite samples

Key Challenges in OBO (2/2)

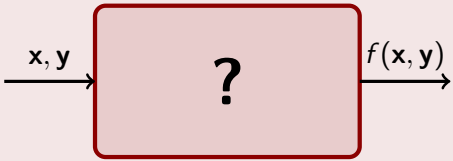
Hypergradient Complexity



Implicit derivatives

Computing hypergradients requires Hessians, Jacobians, ands system solves.

Limited Oracle Access



Function values only

Zeroth-order setting: no direct access to gradients, Hessians, and Jacobians.

Limitations of Prior Work

OAGD (Tarzanagh et al., 2024), SOBOW (Lin et al., 2024), SOBBO (Bohne et al., 2024)

L1. Window Smoothing Required:
search direction : $\frac{1}{w}\sum_{s=t-w+1}^{t}\nabla f_s(x_t, y_t)$

L2. Multiple System Iterations:
$O(\kappa_g \log \kappa_g)$
Computationally expensive inner loops

L3. Strong Oracle Requirements:
Need gradients, Hessians, and Jacobians

L4. Suboptimal Variance Dependence:
$O(T^{-1/2}\sigma^2)$
Can be improved to $O(T^{-2/3}\sigma^2)$

$\Longrightarrow$ **Can we eliminate window smoothing and improve regret bounds?**

Algorithm: SOGD (Simultaneous Online Gradient Descent)

Novel Search Direction: current gradient + momentum + drift correction

At each timestep t: Draw samples $\mathcal{B}_t, \bar{\mathcal{B}}_t$ and compute search directions:

$$\mathbf{d}_t^{\mathbf{yy}} = \nabla_{\mathbf{y}} g_t(\mathbf{z}_t; \bar{\mathcal{B}}_t), \qquad \mathbf{d}_t^{\mathbf{y}} = \mathbf{d}_t^{\mathbf{yy}} + (1-\gamma_t)\mathbf{d}_{t-1}^{\mathbf{y}} + (1-\gamma_t)[\mathbf{d}_t^{\mathbf{yy}} - \mathbf{d}_{t-1}^{\mathbf{yy}}]$$
$$\mathbf{d}_t^{\mathbf{vv}} = \nabla_{\mathbf{y}} f_t + \nabla_{\mathbf{y}}^2 g_t \mathbf{v}_t, \qquad \mathbf{d}_t^{\mathbf{v}} = \mathbf{d}_t^{\mathbf{vv}} + (1-\lambda_t)\mathbf{d}_{t-1}^{\mathbf{v}} + (1-\lambda_t)[\mathbf{d}_t^{\mathbf{vv}} - \mathbf{d}_{t-1}^{\mathbf{vv}}]$$
$$\mathbf{d}_t^{\mathbf{xx}} = \nabla_{\mathbf{x}} f_t + \nabla_{\mathbf{xy}}^2 g_t \mathbf{v}_t, \qquad \mathbf{d}_t^{\mathbf{x}} = \mathbf{d}_t^{\mathbf{xx}} + (1-\eta_t)\mathbf{d}_{t-1}^{\mathbf{x}} + (1-\eta_t)[\mathbf{d}_t^{\mathbf{xx}} - \mathbf{d}_{t-1}^{\mathbf{xx}}]$$

where $\mathbf{d}_{t-1}^{\mathbf{yy}} = \nabla_{\mathbf{y}} g_t(\mathbf{z}_{t-1}; \bar{\mathcal{B}}_t)$, $\mathbf{d}_{t-1}^{\mathbf{vv}} = \nabla_{\mathbf{y}} f_{t-1} + \nabla_{\mathbf{y}}^2 g_{t-1}\mathbf{v}_{t-1}$, $\mathbf{d}_{t-1}^{\mathbf{xx}} = \nabla_{\mathbf{x}} f_{t-1} + \nabla_{\mathbf{xy}}^2 g_{t-1}\mathbf{v}_{t-1}$

Update:

$$\mathbf{y}_{t+1} = \mathbf{y}_t - \beta_t \mathbf{d}_t^{\mathbf{y}}, \quad \mathbf{v}_{t+1} = \Pi_{\mathcal{Z}_p}[\mathbf{v}_t - \delta_t \mathbf{d}_t^{\mathbf{v}}], \quad \mathbf{x}_{t+1} = \Pi_{\mathcal{X}}[\mathbf{x}_t - \alpha_t \mathbf{d}_t^{\mathbf{x}}]$$

Addresses: **L1** (gradient drift correction, $w = 1$), **L2** (single iteration)

Performance Measure: Bilevel Local Regret

Leader's Goal

Minimize **bilevel local regret** over T rounds:

$$\text{BL-Reg}_T := \sum_{t=1}^{T} \mathbb{E}\left[\|\mathcal{P}_{\mathcal{X},\alpha_t}(\mathbf{x}_t; \nabla f_t(\mathbf{x}_t, \mathbf{y}_t^*(\mathbf{x}_t)))\|^2\right]$$

Projected Gradient Mapping:

$$\mathcal{P}_{\mathcal{X},\alpha_t}(\mathbf{x}_t; \nabla f_t) = \frac{1}{\alpha_t}\Big(\mathbf{x}_t - \Pi_{\mathcal{X}}[\mathbf{x}_t - \alpha_t \nabla f_t(\mathbf{x}_t, \mathbf{y}_t^*(\mathbf{x}_t))]\Big)$$

Interpretation: Measures distance from current decision $\mathbf{x}_t$ to stationary points $\mathbf{x}_t^*$ of the leader's objective, where:

$$\mathcal{P}_{\mathcal{X},\alpha_t}(\mathbf{x}_t^*; \nabla f_t(\mathbf{x}_t^*, \mathbf{y}_t^*(\mathbf{x}_t^*))) = 0.$$

Main Results: SOGD Regret Bound

Theorem (SOGD Regret Bound)

With polynomially decaying stepsizes $\alpha_t = O(t^{-1/3})$:

$$\boxed{\text{BL-Reg}_T = O\left(T^{1/3}(\sigma^2 + \Delta_T) + T^{2/3}\Psi_T\right)}$$

Function Variation: $V_T = \sum_{t=2}^{T} \sup_{\mathbf{x}} |f_{t-1}(\mathbf{x}, \mathbf{y}_{t-1}^*) - f_t(\mathbf{x}, \mathbf{y}_t^*)|$
Path Length: $H_{2,T} = \sum_{t=2}^{T} \sup_{\mathbf{x}} \|\mathbf{y}_{t-1}^*(\mathbf{x}) - \mathbf{y}_t^*(\mathbf{x})\|^2$
Gradient Drift: $D_T = \sum_{t=2}^{T} \sup_{\mathbf{x},\mathbf{y}} \|\nabla f_{t-1}(\mathbf{x}, \mathbf{y}) - \nabla f_t(\mathbf{x}, \mathbf{y})\|^2$
where $\Delta_T = E_1 + V_T$ and $\Psi_T = H_{2,T} + G_T + D_T$.

Addresses: **L4** (improved variance rate $T^{-2/3}\sigma^2$ vs. prior $T^{-1/2}\sigma^2$ (Bohne et al., 2024))

Algorithm: ZO-SOGD (Zeroth-Order SOGD)

Key Challenge: No access to gradients, Hessians, or Jacobians – only function values!

Two-Point Gradient Estimation:

$$\hat{\nabla}_{\mathbf{y}} g_t(\mathbf{x}, \mathbf{y}) = \frac{d_2}{2\bar{b}\rho} \sum_{i=1}^{\bar{b}} [g_t(\mathbf{y} + \rho \mathbf{r}_i) - g_t(\mathbf{y} - \rho \mathbf{r}_i)] \mathbf{r}_i$$

Hessian-Vector Estimation:

$$\hat{\nabla}^2_{\mathbf{y}} g_t(\mathbf{x}, \mathbf{y}) \mathbf{v} = \frac{1}{2\bar{b}\rho} \sum_{i=1}^{\bar{b}} [\hat{\nabla}_{\mathbf{y}} g_t(\mathbf{y} + \rho \mathbf{v}) - \hat{\nabla}_{\mathbf{y}} g_t(\mathbf{y} - \rho \mathbf{v})]$$

ZO-SOGD: Use these estimates in SOGD with same momentum structure

Addresses: **L1, L2** (same as SOGD), **L3** (function values only!)

Main Results: ZO-SOGD Regret Bound

Theorem (ZO-SOGD Regret Bound)

With appropriate smoothing parameters and batch sizes:

$$\boxed{\text{BL-Reg}_T = O\left((d_1+d_2)^{3/4} T^{1/3}(\hat{\sigma}^2 + \hat{\Delta}_T) + (d_1+d_2)^{3/2} T^{2/3}\hat{\Psi}_T\right)}$$

First regret guarantee for zeroth-order bilevel optimization!

Comparison with SOGD:

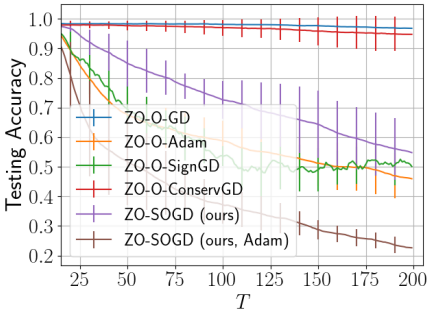
- Dimension dependence: $(d_1+d_2)^{3/4}$ and $(d_1+d_2)^{3/2}$ factors
- Additional regularities: $\hat{\Delta}_T, \hat{\Psi}_T$ include perturbation effects

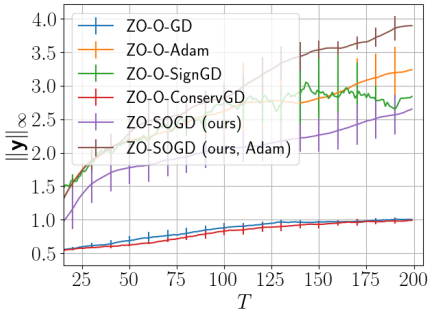
Comparison with Prior Work

Method	Window w	System Iters	ZO	Variance Rate
OAGD (Tarzanagh et al., 2024)	$o(T)$	∞ (exact)	✗	T/w
SOBOW (Lin et al., 2024)	$o(T)$	$O(\kappa \log \kappa)$	✗	T/w
SOBBO (Bohne et al., 2024)	$o(T)$	$O(\kappa \log \kappa)$	✗	$T^{-1/2}\sigma^2/w$
SOGD (Ours)	**1**	**1**	✗	$\mathbf{T^{-2/3}\sigma^2}$
ZO-SOGD (Ours)	**1**	**1**	✓	$\mathbf{(d)^{3/4}T^{-2/3}\sigma^2}$

Experiments: Online Black-Box Adversarial Attacks

Problem: **Inner** learns adversarial perturbations, **outer** tunes regularization to maximize attack strength.

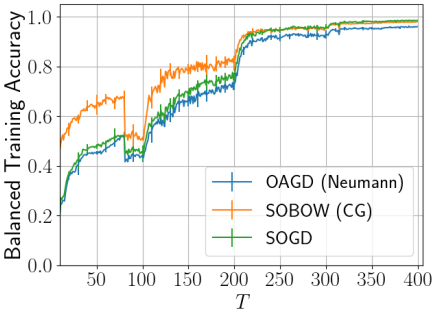


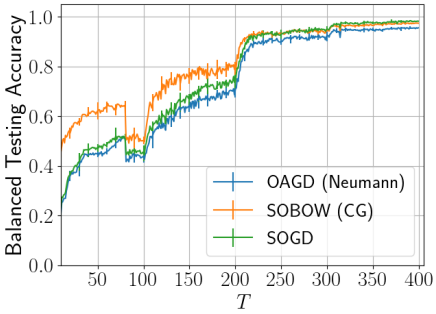


Result: **ZO-SOGD** outperforms all baselines achieving the **strongest attack** while maintaining **imperceptible perturbations** ($\|\mathbf{y}\|_\infty < 4$)

Experiments: Online Loss Tuning for Imbalanced Data

Problem: **Inner** trains with parametric cross-entropy, **outer** tunes logit adjustments for balanced accuracy.





Result: **SOGD** outperforms baselines with **competitive accuracy** and **fastest runtime**, quickly adapting to distribution shifts

Conclusions

We introduced **SOGD** and **ZO-SOGD**, the first online bilevel optimization algorithms with:

- ✓ No window smoothing ($w = 1$)
- ✓ Single iteration per timestep
- ✓ Improved regret bounds: $O(T^{-2/3}\sigma^2)$ vs. prior $T^{-1/2}\sigma^2$ (Bohne et al., 2024)
- ✓ Zeroth-order capability

Code: github.com/Tarzanagh/SOGD

Acknowledgments

SAMSUNG

SAMSUNG SDS

Thank you!