BOME! Bilevel Optimization Made Easy
A Simple First-order Approach

***Bo Liu**[1], *Mao Ye[1], Stephen Wright[2], Peter Stone[1,3], Qiang Liu[1]

[1]The University of Texas at Austin, [2]University of Wisconsin, [3]Sony AI

Problem

We consider the bilevel optimization (**BO**) problem:

$$\underbrace{\min_{v,\theta} f(v,\theta)}_{\text{outer problem}} \quad \text{s.t.} \quad \theta \in \underbrace{\arg\min_{\theta'} g(v,\theta')}_{\text{inner problem}}$$

Example: Hyper-parameter tuning:

$$\min_{v,\theta} L_{\text{val}}(v,\theta) \quad \text{s.t.} \quad \theta \in \arg\min_{\theta'} L_{\text{train}}(v,\theta)$$

Challenges

Scalability: prior **BO** methods often require computing 2nd order gradient each iteration.

Theory: lack non-asymptotic convergence result when f, g are non-convex w.r.t. v, θ.

BOME! (General Idea)

Idea: Convert **BO** into a constrained optimization problem, where g is required to be less than a certain threshold (ideally its optimal value for the given v) .

Optimize the outer problem s.t.
*the **optimality gap** for inner problem is 0*

BOME! (Algorithm)

Step 1: Compute the **approximate value function** (the optimality gap of g)

$$\hat{q}(v,\theta) = g(v,\theta) - g(v, \theta_k^{(T)}).$$

obtained by **T-step** gradient descent ,
then **stop gradient**

Step 2: Descent the **outer** s.t. the **inner** also improves

$$(v_{k+1}, \theta_{k+1}) \leftarrow (v_k, \theta_k) - \xi \delta_k$$

where $\delta_k = \arg\min_{\delta} \underbrace{||\nabla f - \delta||^2}_{\text{descend } f} \quad \text{s.t.} \quad \underbrace{\langle \nabla \hat{q}, \delta \rangle \geq \phi \geq 0}_{\hat{q} \text{ does not ascend}}$

Theory

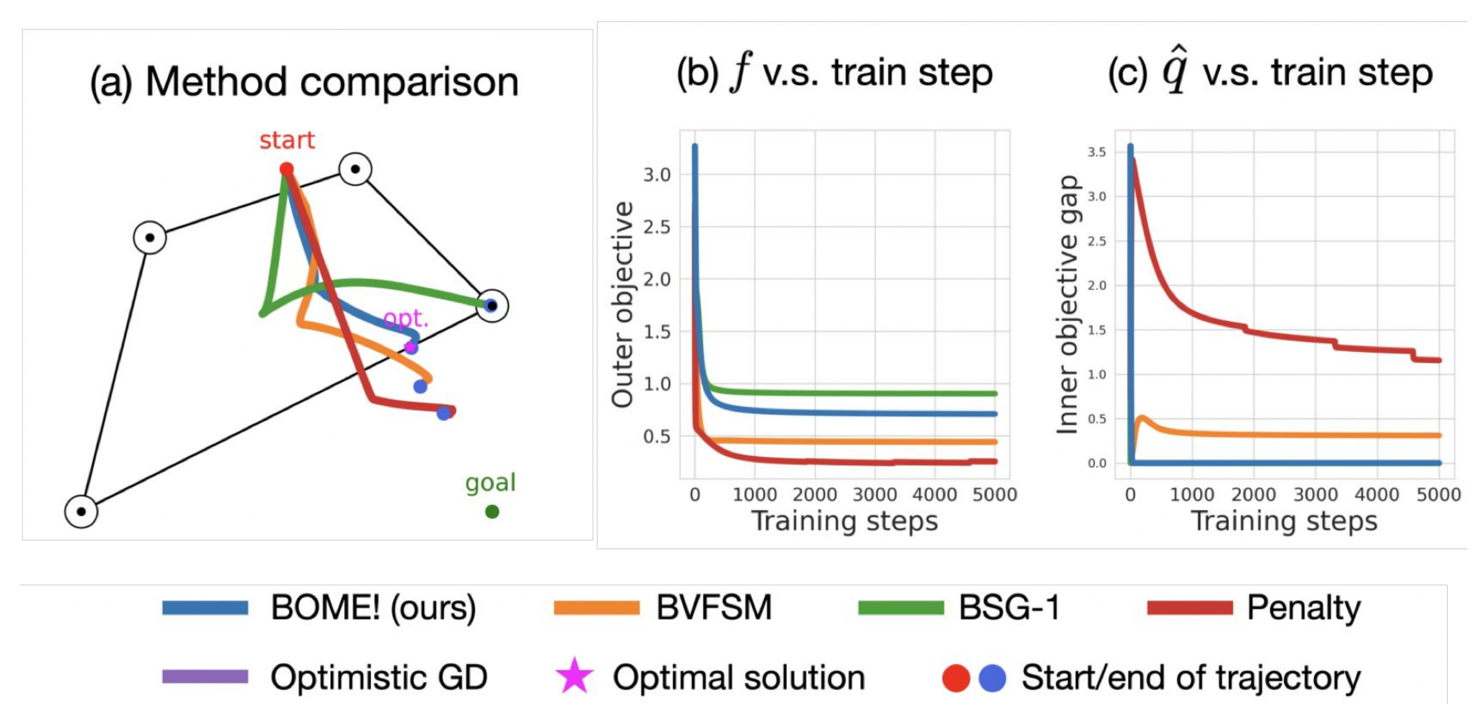
Message: For smooth and (possibly) non-convex inner/outer objectives,

- **non-convex** g, rate is $O(K^{-1/4} + \exp(-bT))$
- **convex** g, rate improves to $O(K^{-1/3} + \exp(-aT))$

Experiment (on a toy example)

Coreset

Find the closest point in the trapezoid to the target goal.



More

Paper Link