Towards Robust Learning to Optimize with Theoretical Guarantees

Qingyu Song (CUHK), Wei Lin (CUHK), Juncheng Wang (HKBU), Hong Xu (CUHK)

香港中文大學 The Chinese University of Hong Kong

CVPR SEATTLE, WA JUNE 17-21, 2024

1. Introduction and Motivation

- What is Learning to Optimize (L2O)?

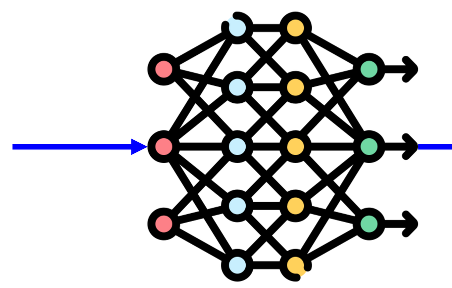
Optimization Problem: $\min_x f(x)$, $x \in R^n, f: R^n \to R$ → Human-Designed Algorithm: Gradient Descent

↓

Learning to Optimize: ML/DL Model + Training

- Benefits
  - Better optimality (potential).
  - Better convergence/efficiency [1].
- Workflow of L2O (Inference, Toy example):

Given initial point x_i^0

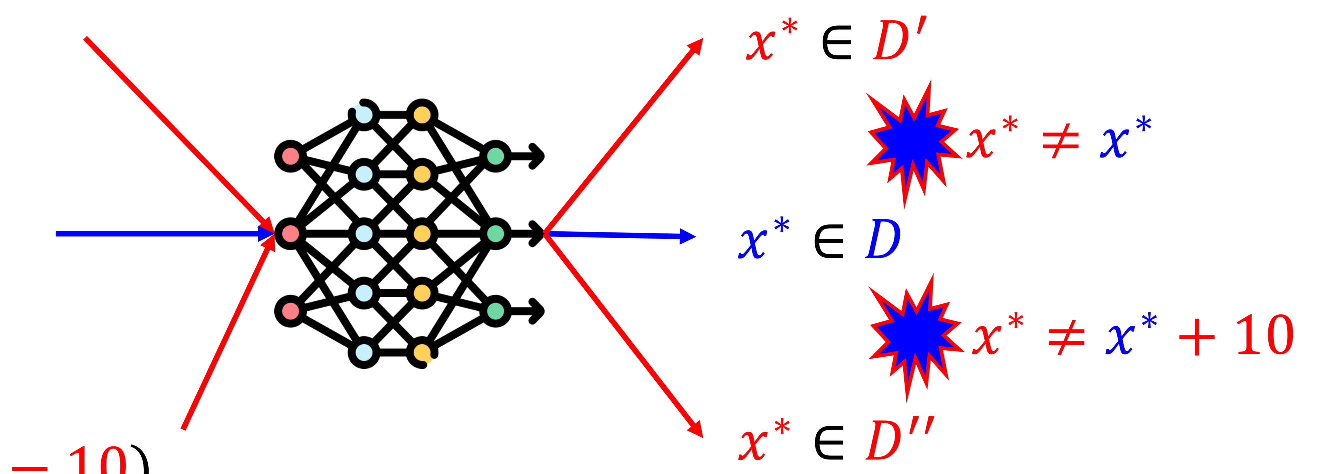
$x_i^0 \in N(0,1), f(x_i^0)$ →  → x_i^*

- L2O Model 's Failure in Out-of-Distribution (OOD) Scenarios

① x OOD
$x^0 \in N(10,1),\ f(x)$

In-Distribution (InD)
$x^0 \in N(0,1),\ f(x)$

② f OOD
$x^0 \in N(0,1),\ f = f(x-10)$



- Q1, How does OOD influence the behavior of x ?
  - Motivation 1: Convergence Analysis in OOD Scenarios.
- Q2, Can we achieve better x^* in OOD scenarios ?
  - Motivation 2: Convergence Improvement in OOD Scenarios.

2. Methods (Step-by-Step, Paper at)

- Backbone L2O Model: Math-Inspired L2O [1]
  - L2O model construction: Necessary condition of convergence.
  - SOTA empirical convergence.

1. Definition:
   - Optimization Problem

   $\min_x f(x) + r(x), x \in R^n$. $f, r: R^n \to R$, f: smooth, r: non-smooth, proper.
   - L2O OOD Definition (Fixed Dimensional Space):

   Define InD Problems ⟶ L2O OOD: Solving OOD Problems
     - Two OOD cases: Initial point x^0 and objective f.
2. OOD Formulation: Formulate OOD with OOD-InD **Difference**
   - Sequences(trajectory in the paper): Variable and Input Feature (defined in terms of objective and variable).

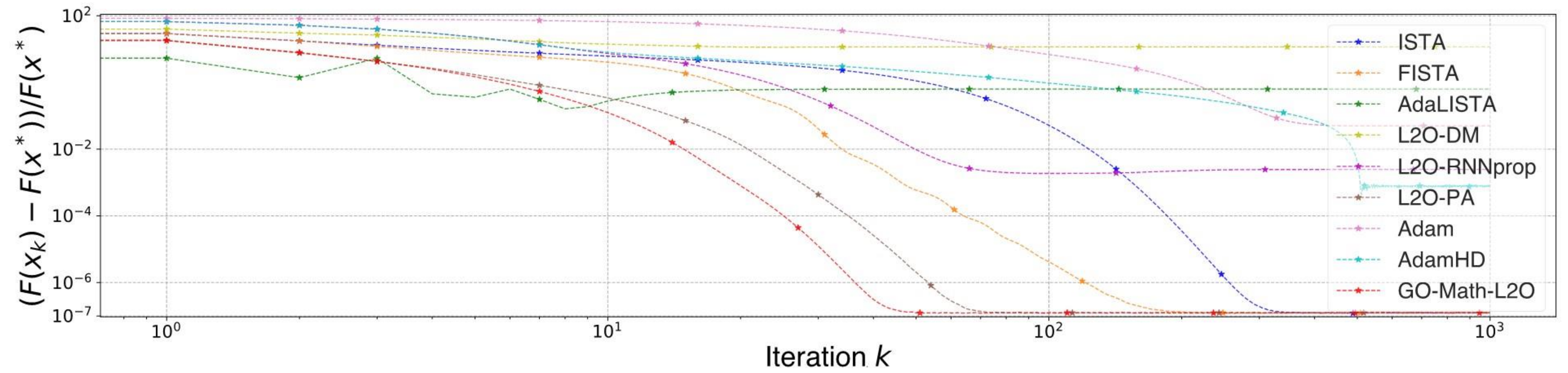
   OOD Sequences − InD Sequences = Virtual Sequences (Extent of OOD)
3. L2O Model's **OOD Behavior** (Output) Formulation:
   - Formulate OOD behavior by InD behavior.
   - Eq. (3) in the paper, extension of Mean Value Theorem for Vector-Valued Function.

   $d(\text{OOD Input Feature}) = d(\text{InD Input Feature}) + \mathbf{J}_d(\text{Difference}) = d(z) + \mathbf{J}_d s'$

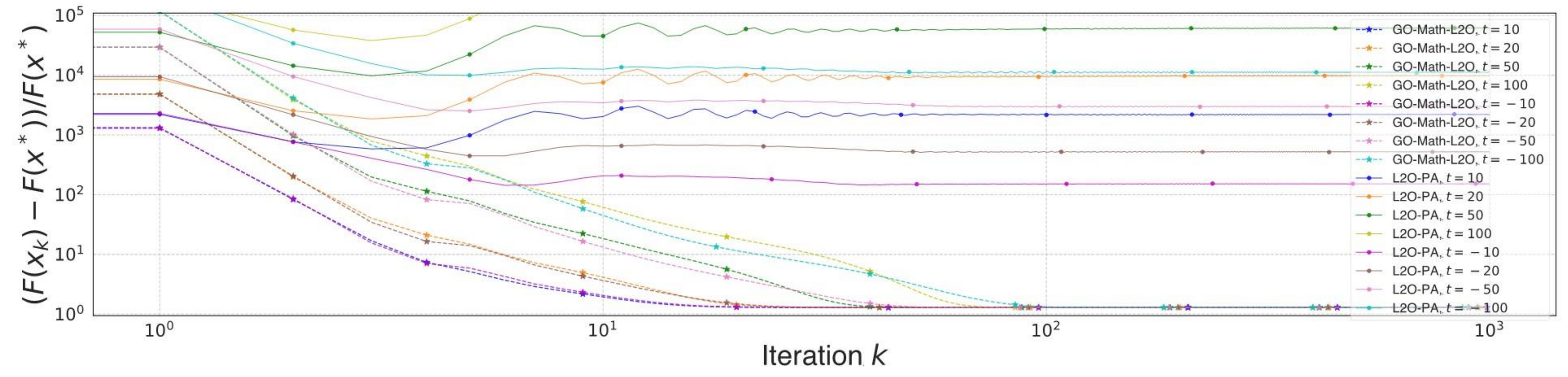
   Bounded "Jacobian" Matrix; Virtual Feature
4. Convergence Analysis: **Convergence Rate Upper Bound**
   - Training-free: Stabilize L2O model's InD behavior by optimal upper bound.
   - OOD Convergence Analysis:
     - Split OOD sequence into InD part and virtual part.
     - InD: Deterministic convergence.
     - OOD: Formulated by virtual variable.
5. Convergence Improvement:
   - Upper bound relaxation on convergence rate: Magnitude of input feature determines convergence rate.
   - Improvement Strategy: Reduce input features.
   - Proposed method: A gradient-only math-inspired L2O model.
     - Eliminate variable-related features.
     - Posterior non-smooth sub-gradient construction.

3. Results（Project at ）

1. Theoretical Results
- InD case: Lemma 1, Corollary 1 in the paper.
  - In general, convergence rate of L2O is as best as a conventional algorithm.
- OOD case: Theorems 1 and 2, Lemmas 2 and 3 in the paper.
  - Convergence rate is deteriorated by $\mathcal{O}(\text{virtual variable})$ $\mathcal{O}(\text{gradient})$.
  - Convergence rate is upper bounded by $\mathcal{O}(\|\text{virtual variable}\|)$.
2. Empirical Results (Our Model: GO-Math-L2O)
- InD Outperformance



- OOD Outperformance over SOTA Math-Inspired L2O [1]



References

[1]. Jialin Liu, Xiaohan Chen, Zhangyang Wang, Wotao Yin, and HanQin Cai. Towards Constituting Mathematical Structures for Learning to Optimize. In ICML, 2023.