1. Introduction and Motivation

- **What is Learning to Optimize (L2O)?**
  - **Virtual f:** Formulated by virtual variable.

- **In Distribution (InD) vs. Out of Distribution (OOD)**
  - **InD:** Formulated by virtual variable.
  - **OOD:** OOD Sequences − InD Sequences = Virtual Sequences

- **Convergence Improvement:**
  - Upper bound relaxation on convergence rate: Magnitude of input feature determines convergence rate.

- **Workflow of L2O (Inference, Toy example):**
  - Given initial point $x_0 \in N(0, 1), f(x_0)$
  - $x^* \in D' 
  - $x^* = x^* + 10$

- **L2O Model’s Failure in Out-of-Distribution (OOD) Scenarios**
  - $x^d \in \mathbb{D}^1 (0, 1), f(x^d)$
  - $x^e \in \mathbb{D}^2 (0, 1), f = f(x^d - 10)$

- **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.

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

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

- **Definition:**
  - Optimization Problem $\min f(x) + r(x), x \in \mathbb{R}^n; r: \mathbb{R}^n \to \mathbb{R}$; 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 $f$.

- **OOD Formulation:**
  - Formulate OOD behavior by $\Omega$.
  - Sequences (trajectory in the paper): Variable and Input Feature (defined in terms of objective and variable).

- **OOD Sequences − InD Sequences = Virtual Sequences**

3. Results (Project at )

- **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 $\Omega$ (virtual variable) $O(gradient)$**.
  - Convergence rate is upper bounded by $O([virtual\ variable])$.

- **Empirical Results (Our Model: GO-Math-L2O)**
  - InD Outperformance
  - OOD Outperformance over SOTA Math-Inspired L2O [1]

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.

References