Generalization Bounded Implicit Learning of Nearly Discontinuous Functions

Penn Engineering | GRASP Laboratory
General Robotics, Automation, Sensing & Perception Lab

Bibit Bianchini, Mathew Halm, Nikolai Matni, Michael Posa

DAIRLab
Dynamic Autonomy & Intelligent Robotics

INTRODUCTION

- **Goal:** Learn a dynamics model of an unknown system.
- **Challenges:**
  - Near discontinuities are rampant in robotics applications in which systems make and break contact with other surfaces.
  - Typical generalization guarantees scale with Lipschitz constants and thus fare poorly for stiff functions.
  - Implicit representations can model high stiffness elegantly, though learning them can still inherit the poor generalizability of the original function if optimized with a prediction-based loss.
- **Our Approach:** Design an implicit loss that, in conjunction with implicit representations, accomplishes:
  - generalizing with bounds that are independent from the original Lipschitz constant, and
  - relating directly to errors in graph distance.

PROBLEM FORMULATION

We aim to learn a function $y = f(x)$ with structure such that it can be formulated as an implicit optimization problem:

$$y = g(x, \lambda)$$
$$\text{such that} \quad \lambda = \underset{\lambda \in \Lambda}{\arg\min}\, h(x, y, \lambda)$$

This class can represent smooth and discontinuous functions alike. Specifically, this structure naturally accommodates rigid body contact dynamics where λ represents frictional forces maximizing energy dissipation.

We compare three approaches defined by the loss functions:

- **Explicit (exp):** explicitly learn f with prediction loss.

$$l^{\theta}_{\text{exp}}(x_i, y_i) = \left\| y_i - f^{\theta}(x_i) \right\|^2$$

- **Naive Implicit (nimp):** implicitly learn g with prediction loss. This is how differentiable simulation is formulated.

$$l^{\theta}_{\text{nimp}}(x_i, y_i) = \left\| y_i - g^{\theta}(x_i, \lambda_i) \right\|^2$$
$$\text{such that} \quad \lambda_i = \underset{\lambda \in \Lambda}{\arg\min}\, h^{\theta}(x_i, g^{\theta}(x_i, \lambda), \lambda)$$

- ***Ours:* Violation Implicit (vimp):** implicitly learn g with a loss that trades off between prediction loss and violation of constraints.

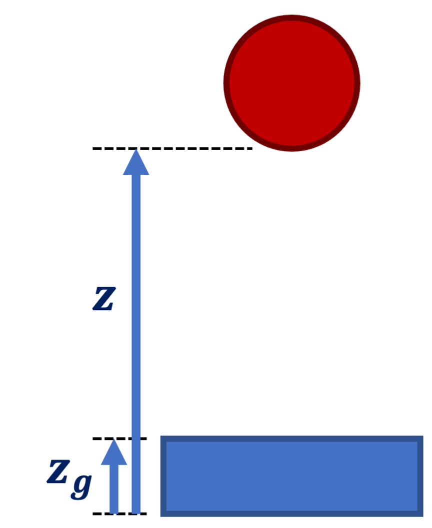
$$l^{\theta}_{\text{vimp}}(x_i, y_i) = \min_{\lambda \in \Lambda} \left\| y_i - g^{\theta}(x_i, \lambda) \right\|^2 + \frac{1}{\epsilon} h^{\theta}(x_i, y_i, \lambda)$$

CONNECT WITH BIBIT

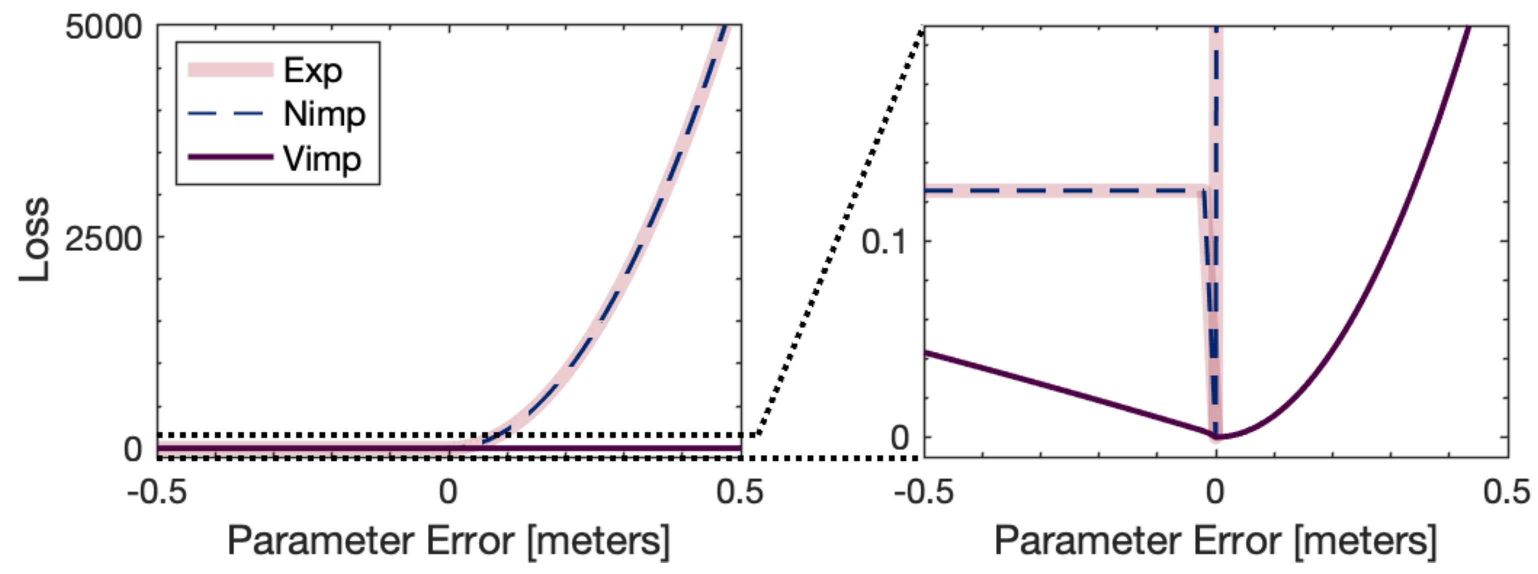
TOY PROBLEM EXAMPLE

We ground our analysis with a physically-relevant 1-D toy problem of a point mass inelastically colliding with a rigid ground.

All approaches (exp, nimp, vimp) share the same parameterization: learning the scalar value of the ground height, z_g.



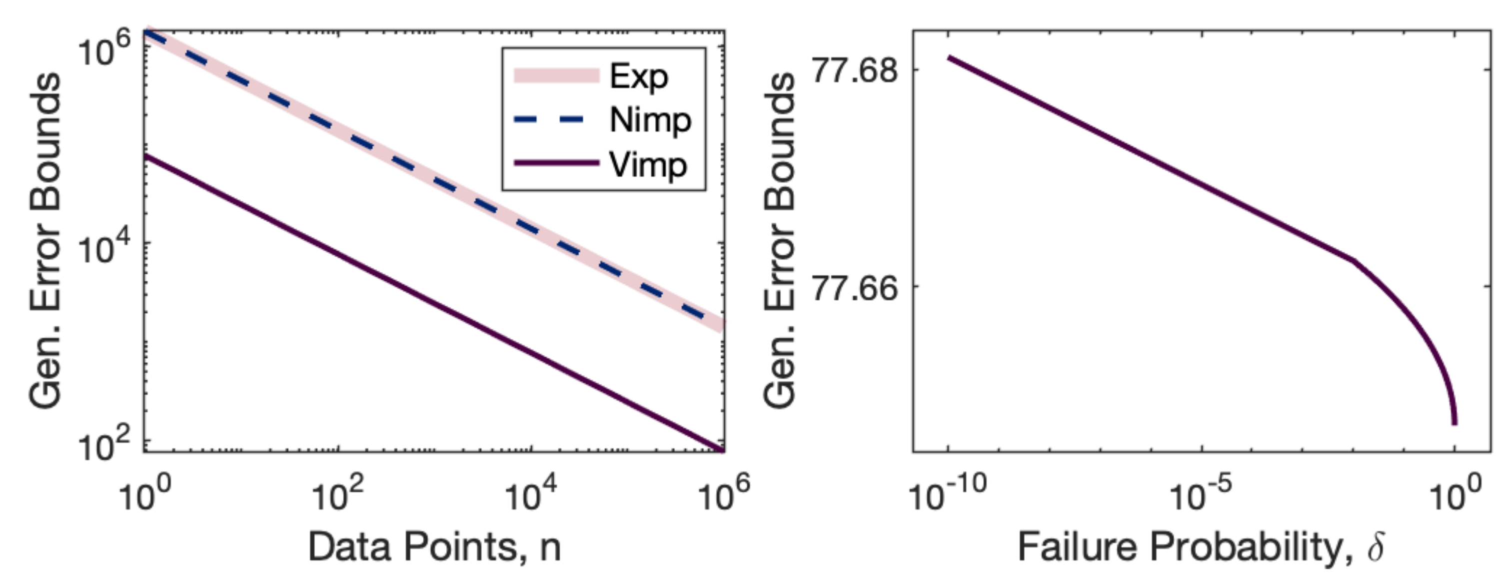
The losses feature **identical landscapes for exp and nimp** approaches, with a **better conditioned landscape for vimp**.



GENERALIZATION ERROR BOUNDS

We combine prior works using Rademacher complexity and Dudley's entropy integral to bound generalization error for our approaches.

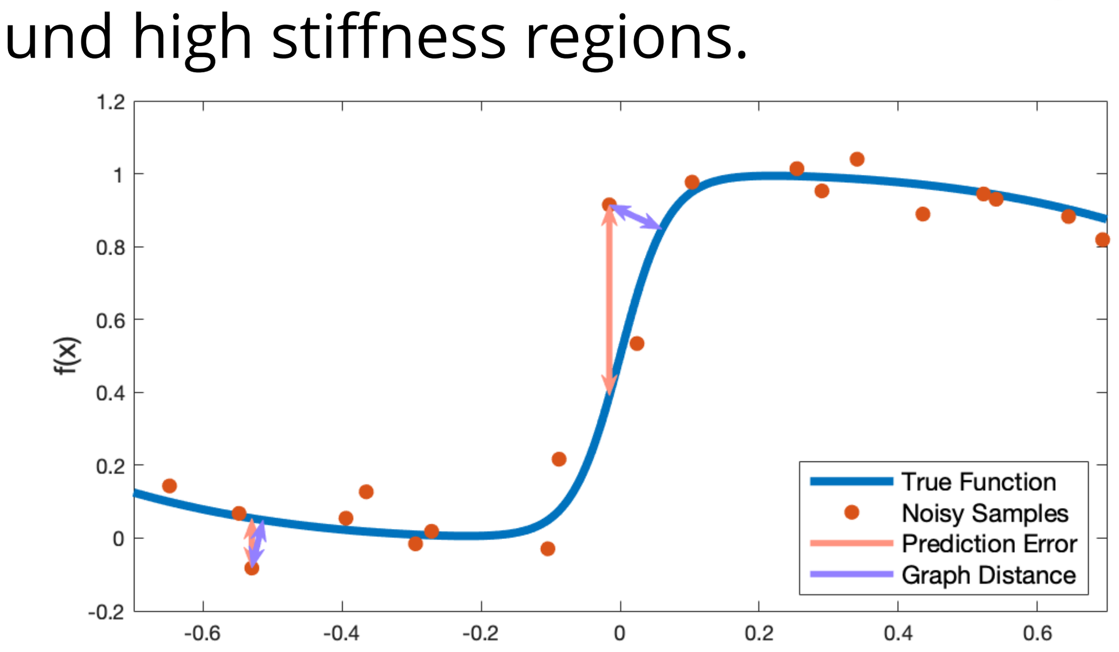
On our toy problem with a choice of ϵ informed by the graph distance analysis, **Lipschitz constants scale like $1/\Delta t$ for exp and nimp, independently of Δt for vimp**. We see **over an order of magnitude reduction in generalization error bound**.



RELATIONSHIP TO GRAPH DISTANCE

Graph distance, or total least squares / errors in variables distance, handles noise in inputs and outputs alike.

For smooth functions, graph distance approaches prediction loss. However, these metrics can significantly differ around high stiffness regions.



Through quadratic growth (QG) arguments, we can prove that **vimp loss can tightly bound graph distance**.

{bibit, mhalm, nmatni, posa}@seas.upenn.edu

This work was supported by an NDSEG Fellowship and an NSF Graduate Research Fellowship under Grant No. DGE-1845298. Toyota Research Institute also provided funds for this work. Nikolai Matni is funded by NSF awards CPS-2038873, CAREER award ECCS-2045834, and a Google Research Scholar award.