Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Progressive Optimal Path Sampling for Closed-Loop Optimal Control Design with Deep Neural Networks

Xuanxi Zhang [*1], Jihao Long [†2], Wei Hu[‡2], Weinan E [§3,4,5], and Jiequn Han [¶6]

[1]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, United States.
[2]Institute for Advanced Algorithms Research, Shanghai 201308, P.R. China.
[3]AI for Science Institute, Beijing 100080, P.R. China.
[4]School of Mathematical Science, Peking University, Beijing 100871, P.R. China.
[5]Center for Machine Learning Research, Peking University, Beijing 100871, P.R. China.
[6]Center for Computational Mathematics, Flatiron Institute, New York, NY 10010, United States.

Abstract. Closed-loop optimal control design for high-dimensional nonlinear systems has been a long-standing challenge. Traditional methods, such as solving the associated Hamilton-Jacobi-Bellman equation, suffer from the curse of dimensionality. Recent literature proposed a new promising approach based on supervised learning, by leveraging powerful open-loop optimal control solvers to generate training data and neural networks as efficient high-dimensional function approximators to fit the closed-loop optimal control. This approach successfully handles certain high-dimensional optimal control problems but still performs poorly on more challenging problems. One of the crucial reasons for the failure is the so-called distribution mismatch phenomenon brought by the controlled dynamics. In this paper, we investigate this phenomenon and propose the progressive optimal path sampling method to mitigate this problem. We theoretically prove that this enhanced sampling strategy outperforms both the vanilla approach and the widely used dataset aggregation method on the classical linear-quadratic regulator by a factor proportional to the total time duration. We further numerically demonstrate that the proposed sampling strategy significantly improves the performance on tested control problems, including the optimal landing problem of a quadrotor and the optimal reaching problem of a 7-DoF manipulator.

Keywords:
Closed-loop optimal control,
Distribution mismatch,
Adaptive sampling,
Supervised learning.

Article Info.:
Volume: 4
Number: 4
Pages: 223 - 263
Date: December/2025
doi.org/10.4208/jml.250213

Article History:
Received: 13/02/2025
Accepted: 08/08/2025

Communicated by:
Jianfeng Lu

1 Introduction

Optimal control aims to find a control for a dynamical system over a period of time such that a specified cost function is minimized. This cost often reflects a combination of task-specific goals such as energy usage, deviation from a tracking target, and control effort. Finding such an optimal control should be distinguished from classical stabilization control [21], which focuses only on keeping the system state bounded or driving it to an equi-

[*]xuanxizhang@nyu.edu
[†]longjh1998@gmail.com
[‡]weihu.math@gmail.com
[§]weinan@math.pku.edu.cn
[¶]Corresponding author. jiequnhan@gmail.com

librium, without regard to minimizing a cost. Generally speaking, there are two types of optimal controls: open-loop optimal control and closed-loop (feedback) optimal control. Open-loop optimal control, also known as trajectory optimization, deals with the problem with a given initial state, and its solution is a function of time for the specific initial data, independent of the other states of the system. In contrast, closed-loop optimal control aims to find the optimal control policy as a function of the state that gives us optimal control for general initial states.

By the nature of the problem, solving the open-loop control problem is relatively easy and various open-loop control solvers can handle nonlinear problems even when the state lives in high dimensions [6, 50]. Closed-loop control is much more powerful than open-loop control since it can cope with different initial states, and it is more robust to the disturbance of dynamics. The classical approach to obtaining a closed-loop optimal control function is by solving the associated Hamilton-Jacobi-Bellman (HJB) equation. However, traditional numerical algorithms for HJB equations such as the finite difference method or finite element method face the curse of dimensionality [5] and hence can not deal with high-dimensional problems.

There is a long history of employing neural networks (NNs) to solve the optimal control problems, see e.g. [1, 9, 19, 23, 28, 35–37, 43–46, 48, 54], and it is getting more attention recently since neural networks have demonstrate superior representation and generalization capabilities.

Generally speaking, there are two categories of methods in this promising direction. One is direct policy search approach [2, 9, 23, 35, 43, 59], which parameterizes the policy function by NNs, samples the total cost with various initial points, and directly minimizes the average total cost. When learning complex policies with hundreds of parameters or solving problems with a long time span and high nonlinearity, the corresponding optimization problems can be extremely hard and may get stuck in poor local minima [35, 58]. The other category of methods is based on supervised learning [37, 43–47]. Combining various techniques for open-loop control, one can solve complex high-dimensional open-loop optimal control problems, see [6, 30, 50] for surveys. Consequently, we can collect optimal trajectories for different initial points as training data, parameterize the control function (or value function) using NNs, and train the NN models to fit the closed-loop optimal controls (or optimal values). This work focuses on the second approach and aims to improve its performance through adaptive sampling.

As demonstrated in [46, 56, 58], NN controllers trained by the vanilla supervised-learning-based approach can perform poorly even when both the training error and test error on collected datasets are fairly small. Some existing works attribute this phenomenon to the fact that the learned controller may deteriorate badly at some difficult initial states even though the error is small in the average sense. Several adaptive sampling methods regarding the initial points are hence proposed (see Section 4 for a detailed discussion). However, these methods all focus on choosing optimal paths according to different initial points and ignore the effect of dynamics. This is an issue since the paths controlled by the NN will deviate from the optimal paths further and further over time due to the accumulation of errors. As shown in Section 6, applying adaptive sampling only on initial points is insufficient to solve challenging problems.