Multistep Quasimetric Learning for Scalable Goal-conditioned Reinforcement Learning

Bill Chunyuan Zheng[1] **Vivek Myers**[1] **Benjamin Eysenbach**[2] **Sergey Levine**[1]

[1]UC Berkeley [2] Princeton University

November 9, 2025

Abstract

Learning how to reach goals in an environment is a longstanding challenge in AI, yet reasoning over long horizons remains a challenge for modern methods. The key question is how to estimate the temporal distance between pairs of observations. While temporal difference methods leverage local updates to provide optimality guarantees, they often perform worse than Monte Carlo methods that perform global updates (e.g., with multi-step returns), which lack such guarantees. We show how these approaches can be integrated into a practical GCRL method that fits a quasimetric distance using a multistep Monte-Carlo return. We show our method outperforms existing GCRL methods on long-horizon simulated tasks with up to 4000 steps, even with visual observations. We also demonstrate that our method can enable stitching in the real-world robotic manipulation domain (Bridge setup). Our approach is the first end-to-end GCRL method that enables multistep stitching in this real-world manipulation domain from an unlabeled offline dataset of visual observations.[1]

1 Introduction

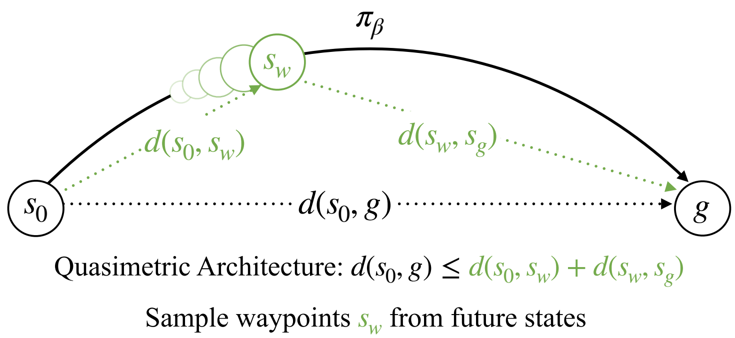

Figure 1: In this paper, we present Multistep Quasimetric Estimation (MQE). Unlike prior work in quasimetric distance learning that use single-step TD updates (Wang et al., 2023) or Monte-Carlo updates (Myers et al., 2024), MQE is the *first* work to incorporate multistep returns with real-world success.

It is natural for humans to use inherent ideas of distances to represent task progress: a GPS will tell you how far you are from the destination and a cookbook will tell you how long a recipe will take. Humans

[1]Website and code: `https://mqe-paper.github.io`