Ergodic Imitation for Adaptive Exploration around Demonstrations

Ziyi Xu[1*], Cem Bilaloglu[2,1*], Yiming Li[2,1], and Sylvain Calinon[2,1]

***Abstract*— In robotics, a common challenge in imitation learning is the mismatch between training and deployment conditions, caused, for example, by environmental changes or imperfect observation and control. When a robot follows a nominal trajectory under such mismatch, it may become stuck and fail to complete the task. This calls for adaptive online exploration strategies that remain grounded in demonstrations. To this end, we propose an adaptive ergodic imitation approach that constructs a target distribution from the geometry of the retrieved demonstrations and uses it to generate trajectories that adaptively interpolate between tracking and exploration. Our method extends ergodic control beyond its traditional role in area-coverage and search by incorporating demonstrations into a retrieval-based receding-horizon framework for adaptive imitation.**

I. INTRODUCTION

Imitation learning (IL) is a practical paradigm for robot programming, where the intended behavior is learned by demonstrations without an explicit reward. In practice, however, the objective is rarely to replicate a demonstration exactly; rather, the agent must adapt the observed behaviors to novel environments that might diverge from the training distribution. Recent critiques of deep generative models in robotics suggest that these architectures often overfit the demonstration data, essentially memorizing specific action sequences rather than learning generalizable policies [1], [2]. Consequently, current IL approaches remain notoriously brittle even under minimal distribution shifts. This necessitates a shift toward imitation paradigms that prioritize situational adaptation over pure trajectory replay.

Although recent adaptive exploration methods allow systems to sample different modes from the dataset, they are largely restricted to discrete transitions [3], [4]. Consequently, they lack the continuous exploration capabilities needed for the subtle, state-space adjustments typical of tasks like robotic assembly. We employ a different perspective that treats demonstrations as references to be tracked under nominal conditions, by using them as informed priors for exploration when environmental shifts render pure tracking insufficient. Instead of considering tracking and exploring as two different objectives, we use the ergodic control methodology to formulate tracking as a special case of exploration. This results in a continuous behavioral spectrum, from rigid imitation to adaptive exploration, governed by the same underlying controller.

*Equal contribution.

The authors are with the [1] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland and with the [2] Idiap Research Institute, Martigny, Switzerland ziyi.xu@epfl.ch, cem.bilaloglu@idiap.ch, yiming.li@idiap.ch, sylvain.calinon@idiap.ch

Ergodic control synthesizes trajectories whose time-averaged state visitation statistics converge to a target spatial distribution [5]. Traditionally, this framework is employed for exploration [6], area coverage [7], or search operations [8], using mobile robots and UAVs. More recently, several works have extended ergodic control to the IL setting, applying it to tasks such as robotic insertion [9], [10], cart-pole inversion, and surface cleaning [11]. In this imitation context, demonstrations are leveraged to construct a static target distribution that the ergodic controller used for task reproduction. A primary advantage of ergodic imitation is its reliance on statistical state-visitation rather than strict temporal sequencing; this allows the robot to synthesize successful behaviors without being tethered to the demonstrator's specific timeline [11]. While this flexibility is beneficial for tasks like cleaning or insertion, temporal dependency remains critical for most manipulation tasks. To address this, we propose an adaptive formulation that strikes a balance between these two paradigms. Our approach reproduces the temporal characteristics of the demonstration when the environment permits, yet autonomously transitions to ergodic exploration if the robot becomes "stuck" and can not track the reference motion.

In this work, we introduce an adaptive ergodic imitation approach that brings the target distribution-driven exploration mechanism of ergodic control into imitation learning for runtime adaptation. Our work positions ergodic control as a promising direction for generalization in imitation learning, not only as a tool for search or coverage, but as a principled mechanism for adaptive execution around demonstrations. Our contributions are as follows:

- a unified method for a continuous spectrum of tracking and exploration behaviors using adaptive ergodic imitation;
- task progress estimation via demonstration to adaptively modify target distributions;
- geometry-guided anisotropic diffusion for synthesizing target distributions inducing tracking and exploration;
- using Maximum Mean Discrepancy (MMD) ergodic metric [12] in a retrieval-based imitation learning framework.

II. METHODOLOGY

We consider a dataset $\mathcal{D} = \left\{\Gamma^{(i)}\right\}_{i=1}^{N}$ of N expert demonstrations, where each demonstration is a state trajectory $\Gamma^{(i)} = \left\{\boldsymbol{x}_t^{(i)}\right\}_{t=0}^{T_i}$. Since the proposed framework utilizes an ergodic controller to derive control laws based on target distributions, the expert actions $\boldsymbol{u}_t^{(i)}$ are omitted, and we focus exclusively on the state trajectories. We define the aggregate

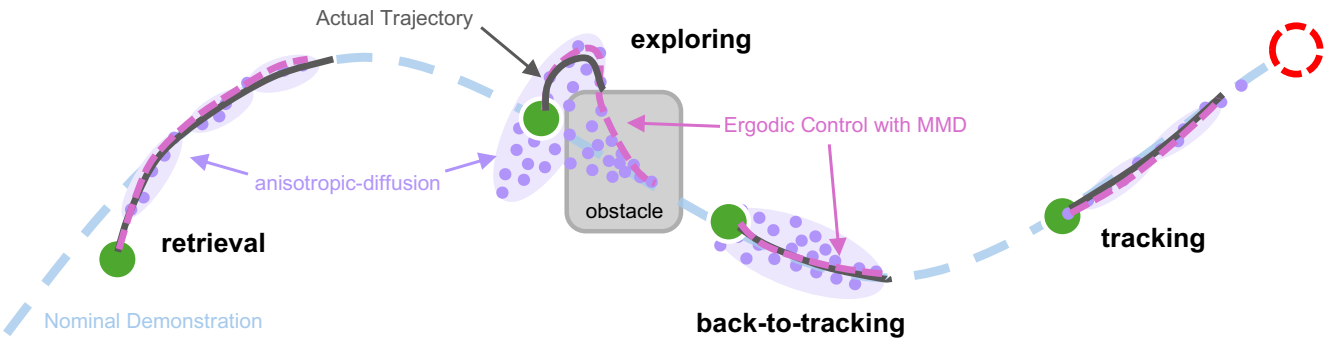

Fig. 1: Overview of ADAPTIVE ERGODIC IMITATION. A nominal trajectory induces tracking behavior when execution remains aligned with the demonstration. Under mismatch, the target particle distribution expands and the ergodic planner promotes exploration around the reference. Once the obstacle is bypassed, the score-based kernel contracts the distribution and pulls the agent back toward the demonstrated trajectory.

state set $\mathcal{P} = \bigcup_{i,t} \left\{ \boldsymbol{x}_t^{(i)} \right\}$ as the flattened collection of all expert states.

A. Background

A dynamical system's time-averaged state trajectory defines an empirical distribution, or coverage, as $p(t, \boldsymbol{x}) = \frac{1}{t} \int_0^t \delta(\boldsymbol{x} - \boldsymbol{x}(\tau)) d\tau$, where $\delta(\cdot)$ is the Dirac delta function. The system is ergodic with respect to a target distribution $q(\boldsymbol{x}) \in \mathcal{S}(\mathcal{X})$ if its coverage converges weakly to the target:

$$\lim_{T \to \infty} \mathbb{E}_{\boldsymbol{x} \sim p_T}[\phi(\boldsymbol{x})] = \mathbb{E}_{\boldsymbol{x} \sim q}[\phi(\boldsymbol{x})], \quad \forall \phi \in \mathcal{C}(\mathcal{X}). \quad (1)$$

The objective of ergodic control is to synthesize control commands that ensure the system's long-term coverage matches the desired distribution $q(\boldsymbol{x})$, typically by minimizing a discrepancy measure between p and q. MMD provides a discrepancy measure for ergodicity [12], particularly effective when the target distribution is known only through a set of discrete samples $\{\boldsymbol{q}_i\}_{i=1}^N \subset \mathcal{X}$. For a trajectory represented by discrete points $\boldsymbol{x}_t$, the squared MMD between the trajectory distribution $p_{\boldsymbol{x}}$ and the target distribution q is approximated as:

$$\begin{aligned} \overline{\mathrm{MMD}}_k^2(p, q) = \frac{1}{T^2} \sum_{t=0}^{T-1} \sum_{t'=0}^{T-1} k(\boldsymbol{x}_t, \boldsymbol{x}_{t'}) \\ - \frac{2}{TN} \sum_{t=0}^{T-1} \sum_{i=1}^{N} k(\boldsymbol{x}_t, \boldsymbol{q}_i) + z(q), \end{aligned} \quad (2)$$

where $z(q) = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N k(\boldsymbol{q}_i, \boldsymbol{q}_j)$ is a constant term that depends solely on the target distribution.

B. Phase Retrieval

At each re-planning interval, the system queries the dataset $\mathcal{D}$ using the current robot state $\boldsymbol{x}_q = \boldsymbol{x}(t)$ to identify the most relevant expert context. We define the projected state $\boldsymbol{x}'$ as the element in $\mathcal{P}$ that minimizes the distance to the query state:

$$\boldsymbol{x}' = \arg\min_{\boldsymbol{x}_k^{(i)} \in \mathcal{P}} \|\boldsymbol{x}_q - \boldsymbol{x}_k^{(i)}\|_2. \quad (3)$$

Each $\boldsymbol{x}'$ is associated with a specific temporal index t' from its source trajectory $\Gamma^{(i)}$, representing the "expert phase".

To evaluate progress relative to the demonstration, we introduce a *phase error* $e(t) = \tau(t) - t'$, where $\tau(t)$ is a virtual reference clock. Unlike the continuous execution time t, the reference clock $\tau(t)$ adapts to the robot's performance. The update logic for the reference clock and a stagnation counter c is governed by a threshold ϵ :

- Progressing $\rightarrow$ Tracking: If $e(t) \leq \epsilon$, the robot is successfully tracking the demonstration. We set $c = 0$ and allow the reference clock to advance ($\dot{\tau} = 1$).
- Stagnating $\rightarrow$ Exploration: If $e(t) > \epsilon$, the robot is lagging behind the expert context. We fix the reference clock ($\dot{\tau} = 0$) and increment the stagnation counter c to signal the exploration.

This stagnation signal c provides a heuristic measure of task progress and determines whether to track or explore around demonstrations.

C. Distribution Generation

We use anisotropic diffusion to generate a particle-based distribution $\{\boldsymbol{q}_j\}_{j=1}^N \subset \mathcal{X}^{(i)}$ along the nominal trajectory $\Gamma^{(i)}$ that unifies *tracking* and *exploration* within a single stochastic differential equation (SDE):

$$\begin{aligned} \mathrm{d}\boldsymbol{q}_j = \Big[\underbrace{\kappa(\theta)\big(\boldsymbol{x}^{(i)*}(\boldsymbol{q}_j) - \boldsymbol{q}_j\big)}_{\text{nominal attraction}} + \underbrace{\alpha(\theta)\, \nabla_{\boldsymbol{q}_j} \log p_t(\boldsymbol{q}_j)}_{\text{heat-kernel score}} \Big] \mathrm{d}t \\ + \underbrace{\sqrt{2\, \boldsymbol{\Sigma}(\theta, \boldsymbol{q}_j)}}_{\text{anisotropic diffusion}} \mathrm{d}\mathbf{W}_t, \end{aligned} \quad (4)$$

where $\boldsymbol{x}^{(i)*}(\boldsymbol{q}_j)$ denotes the projection of particle $\mathbf{q}_j$ onto $\Gamma^{(i)}$, $\mathbf{W}_t$ is a standard Wiener process, and $\theta \in [0, 1]$ is a temperature-like progress variable that modulates the balance between tracking and exploration. For small θ, attraction and score-based tracking dominate; as θ increases, these terms weaken and diffusion becomes prominent.

a) Curve attraction: The first drift term pulls each particle toward its nearest point on $\Gamma^{(i)}$. Temperature-dependent coefficient $\kappa(\theta)$ is large at *tracking* and weakens at *exploring*.

b) Heat-kernel score: The second drift term is the score function of a kernel density estimate defined over $\Gamma^{(i)}$, $p_t(\boldsymbol{q}_j) = \frac{1}{T_i}\sum_{t=0}^{T_i} k_t(\boldsymbol{q}_j, \boldsymbol{x}_t^{(i)})$, where k_t is a heat kernel and $\boldsymbol{x}_t^{(i)}$ are samples along $\Gamma^{(i)}$. The corresponding score hence takes the form of a kernel-weighted average of the individual kernel scores:

$$\nabla_{\boldsymbol{q}_j} \log p_t(\boldsymbol{q}_j) = \frac{\sum_{T_i} k_t(\boldsymbol{q}_j, \boldsymbol{x}_t^{(i)})\, \nabla_{\boldsymbol{q}_j} \log k_t(\boldsymbol{q}_j, \boldsymbol{x}_t^{(i)})}{\sum_{T_i} k_t(\boldsymbol{q}_j, \boldsymbol{x}_t^{(i)})}. \tag{5}$$

This term biases particles toward regions of high reference density, helping them stay close to the reference distribution during the tracking phase.

c) Anisotropic diffusion: To promote goal-aware exploration around the reference trajectory, we design the diffusion to be anisotropic using the geometry of the demonstrations: noise is larger in directions normal to the trajectory than along its tangent, and larger near the beginning of the trajectory than near its end. Let $\hat{\boldsymbol{t}}(\boldsymbol{x}^{(i)*}(\boldsymbol{q}_j))$ denote the local unit tangent at the reference trajectory projection of particle $\boldsymbol{q}_j$. The diffusion term is split as:

$$\begin{aligned}\sqrt{2\,\boldsymbol{\Sigma}(\theta, \boldsymbol{q}_j)}\,\mathrm{d}W_t = &\sqrt{2D_{\parallel}(\theta)}\,\left(\hat{\boldsymbol{t}}\hat{\boldsymbol{t}}^{\top}\right)\mathrm{d}\mathbf{W}_t \\ &+ \sqrt{2D_{\perp}(\theta)}\,\left(\mathbf{I} - \hat{\boldsymbol{t}}\hat{\boldsymbol{t}}^{\top}\right)\mathrm{d}\mathbf{W}_t,\end{aligned} \tag{6}$$

where $D_{\parallel}$ and $D_{\perp}$ are the tangential and orthogonal diffusion coefficients, respectively. Choosing $D_{\perp} \gg D_{\parallel}$ promotes exploration primarily in directions normal to the trajectory while preserving coherence along it.

Particle diffusion is further bounded by a Laplacian envelope defined along $\Gamma^{(i)}$, parameterized by arc length s, with A controlling its amplitude and b its spatial decay:

$$E(s) = \frac{A}{2b}\exp\left(-\frac{|s|}{b}\right). \tag{7}$$

As the phase error $e(t)$ and stagnation signal c increases, the envelope broadens, allowing exploration over a larger neighborhood of the reference trajectory and eventually approaching the uniform case corresponding to pure exploration.

D. Coverage-Aware Ergodic Control with MMD

Since we use a sample-based representation of the target distribution, we adopt the MMD ergodic metric introduced by Hughes *et al.*. We implement a receding horizon controller and include the last ten trajectories from previous planning steps in the MMD objective (Eq. (2)). This allows previously covered regions to be reflected in the objective, encouraging each new plan to complement rather than retrace past coverage.

MMD and our method extend naturally to $SE(3)$ and to any other curved space where we can define a geometry-aware kernel such as the heat kernel [13], since we can define the MMD objective. In $SE(3)$ we can exploit the product structure $SE(3) \cong SO(3) \times \mathbb{R}^3$, such that the resulting kernel captures rotational and translational components in a unified representation.

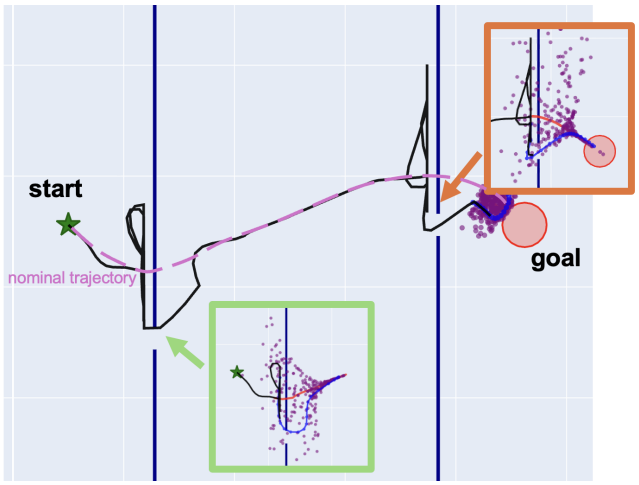

Fig. 2: Adaptive exploration in the maze environment.

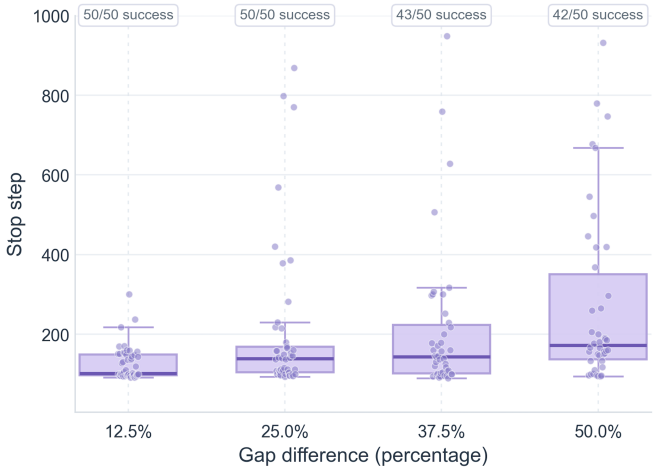

Fig. 3: Quantitative maze results under gate-location perturbations sampled around the nominal layout.

III. RESULTS

We evaluated our method in a 2D navigation environment with narrow vertical gaps and a cluttered goal region, shown in Fig. 2. Successful demonstrations are first collected in a nominal layout, after which the gap locations are shifted at test time to induce deployment mismatch. Under this perturbation, the agent must adaptively explore around the demonstrated behavior rather than simply replay the nominal trajectory. When blocked by a wall, the agent detects a task-progress mismatch via the accumulated phase error. This discrepancy broadens the target distribution with anisotropic diffusion, resulting in exploration through the ergodic controller.

We additionally tested 50 gate location offsets with respect to the nominal gate position around different gap differences using a Gaussian, shown in Fig. 3. *Success* is defined as reaching the goal within 1000 steps, and we used the same nominal trajectory. In such obstacle blocking case, retrieval or generative-based methods would have 0 success rate due to the out-of-distribution gap position, whereas for our method we would always eventually find a solution.

REFERENCES

[1] C. He, X. Liu, G. S. Camps, G. Sartoretti, and M. Schwager, "Demystifying diffusion policies: Action memorization and simple lookup table alternatives," *arXiv preprint arXiv:2505.05787*, 2025.

[2] Y. Li, N. Darwiche, A. Razmjoo, S. Liu, Y. Du, A. Ijspeert, and S. Calinon, "Geometry-aware policy imitation," in *Proc. Intl Conf. on Learning Representations (ICLR)*, 2026.

[3] A. Razmjoo, S. Calinon, M. Gienger, and F. Zhang, "Ccdp: Composition of conditional diffusion policies with guided sampling," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 20 036–20 043.

[4] Y. Jin, J. Lv, W. Yu, H. Fang, Y.-L. Li, and C. Lu, "Sime: Enhancing policy self-improvement with modal-level exploration," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2025, pp. 9792–9799.

[5] G. Mathew and I. Mezic, "Spectral Multiscale Coverage: A uniform coverage algorithm for mobile sensor networks," in *Proceedings of the 48h IEEE Conference on Decision and Control (CDC)*, Dec. 2009, pp. 7872–7877.

[6] L. M. Miller, Y. Silverman, M. A. MacIver, and T. D. Murphey, "Ergodic Exploration of Distributed Information," *IEEE Transactions on Robotics*, vol. 32, pp. 36–52, Feb. 2016.

[7] C. Bilaloglu, T. Löw, and S. Calinon, "Tactile Ergodic Coverage on Curved Surfaces," *IEEE Transactions on Robotics*, vol. 41, pp. 1421–1435, 2025.

[8] S. Ivić, B. Crnković, H. Arbabi, S. Loire, P. Clary, and I. Mezić, "Search strategy in a complex and dynamic environment: The MH370 case," *Scientific Reports*, vol. 10, p. 19640, Nov. 2020.

[9] S. Shetty, J. Silvério, and S. Calinon, "Ergodic Exploration Using Tensor Train: Applications in Insertion Tasks," *IEEE Transactions on Robotics*, vol. 38, pp. 906–921, Apr. 2022.

[10] M. Sun, A. Gaggar, P. Trautman, and T. Murphey, "Fast Ergodic Search with Kernel Functions," Mar. 2024.

[11] A. Kalinowska, A. Prabhakar, K. Fitzsimons, and T. Murphey, "Ergodic imitation: Learning from what to do and what not to do," Mar. 2021.

[12] C. Hughes, H. Warren, D. Lee, F. Ramos, and I. Abraham, "Ergodic trajectory optimization on generalized domains using maximum mean discrepancy," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2025, pp. 01–07.

[13] V. Borovitskiy, A. Terenin, P. Mostowsky, and M. P. Deisenroth, "Mate´rn Gaussian processes on Riemannian manifolds."