

# Symplectic Inductive Bias for Data-Driven Target Reachability in Hamiltonian Systems

Zhuo Ouyang, Jixian Liu, and Enrique Mallada

**Abstract**—Inductive bias refers to restrictions on the hypothesis class that enable a learning method to generalize effectively from limited data. A canonical example in control is linearity, which underpins low sample-complexity guarantees for stabilization and optimal control. For general nonlinear dynamics, by contrast, guarantees often rely on smoothness assumptions (e.g., Lipschitz continuity) which, when combined with covering arguments, can lead to data requirements that grow exponentially with the ambient dimension. In this paper we argue that data-efficient nonlinear control demands exploiting inductive bias embedded in nature itself—namely, structure imposed by physical laws. Focusing on Hamiltonian systems, we leverage symplectic geometry and intrinsic recurrence on energy level sets to solve target reachability problems. Our approach combines the recurrence property with a recently proposed class of policies, called chain policies, which composes locally certified trajectory segments extracted from demonstrations to achieve target reachability. We provide sufficient conditions for reachability under this construction and show that the resulting data requirements depend on explicit geometric and recurrence properties of the Hamiltonian rather than the state dimension.

## I. INTRODUCTION

A central challenge in data-driven control is how to achieve reliable generalization from limited data. In learning theory, this capability is governed by inductive biases, namely, structural restrictions on the hypothesis class—the set of models or control laws considered—that enable generalization [1]–[3]. By limiting the class complexity, inductive bias determines how solutions inferred from finite data extend beyond observed data, and how data requirements scale with problem complexity. In control, such inductive bias typically appears through assumptions on the system model, such as linearity, polynomial structure, or smoothness (e.g., Lipschitz continuity), which define the class of admissible dynamics or policies over which guarantees are derived.

For linear systems, this structure leads to tractable and often sample-efficient data-driven control methods. A broad class of problems—including system identification, stabilization, and optimal control—can be addressed directly from data using a range of techniques, such as convex optimization, LMI-based formulations, and trajectory-based predictive control [4]–[12]. In several cases, these approaches admit rigorous finite-sample guarantees that explicitly characterize

how data requirements scale with system dimension and control objectives [4]–[7], [9]. As a result, the linear setting provides a relatively rich understanding of the interplay between data, computation, and control objectives.

In contrast, for nonlinear systems, there is no a priori canonical inductive bias, and different modeling assumptions—such as polynomial, rational, or Lipschitz models—lead to significantly different methodologies and outcomes [13]–[16]. While these approaches enable controller synthesis with explicit guarantees, sample complexity bounds remain comparatively scarce. A few exceptions include sample complexity results for stability [17], stabilizability [18], and reachability analysis [19]. Notably, in all these cases, the resulting data requirements scale *exponentially with the state dimension*, reflecting the intrinsic difficulty of data-driven control in nonlinear systems.

The above-mentioned results suggest a fundamental limitation in data-driven control for nonlinear systems. In this paper, we argue that this apparent limitation stems from the choice of function class—such as polynomial or Lipschitz models—which fail to capture the rich structure embedded in physical systems. To overcome this, we advocate for an inductive bias grounded in physical laws. In particular, we focus on Hamiltonian systems, where the dynamics are governed by an energy function and evolve on invariant energy level sets. Leveraging their symplectic structure and intrinsic recurrence properties, we study the problem of target reachability and show how these features can be used to design data-driven control policies that fundamentally alter the dependence of sample complexity on system dimension.

Our approach builds on a novel class control policies, called chain policies [18], which construct control strategies by composing locally validated control segments derived from data. By recurrently composing the execution control segments, chain policies guarantee certain recurrent conditions that is sufficient for stability and safety in nonlinear systems [20]–[23]. In the Hamiltonian setting, however, recurrence is not imposed as a design principle but arises intrinsically from the dynamics on invariant energy layers, creating natural opportunities to reuse locally valid behaviors. This leads to a reachability framework where global performance can be achieved from a finite collection of trajectory segments. We establish sufficient conditions under which target reachability is guaranteed and show that the associated data requirements depend on intrinsic geometric and recurrence properties of the Hamiltonian—such as energy variation and ergodic structure—rather than the ambient state dimension.

Zhuo Ouyang is with the College of Engineering, Peking University, BJ 100091, P.R.C. 2200011199@stu.pku.edu.cn. J. Liu and E. Mallada are with the Department of Electrical and Computer Engineering, Johns Hopkins University, MD 21218, U.S.A. jliu376@jh.edu, mallada@jhu.edu.

This work was supported by the NSF Global Centers program under Grant No. 2330450 and by the DOE Office of Science (ASCR) under Award No. 826565.

The remainder of the paper is organized as follows. Section II introduces the Hamiltonian system model and formulates the target reachability problem. Section III presents the main theoretical results, including reachability guarantees and sample complexity bounds under chain policies. Section IV provides numerical validation on representative systems, and Section V concludes with a discussion of future directions.

*Notation:* We denote by  $\|\cdot\|$  the Euclidean norm. For  $x \in \mathbb{R}^n$  and  $r > 0$ , let  $\mathcal{B}_r(x) := \{w \in \mathbb{R}^n \mid \|x - w\| \leq r\}$  be the closed ball of radius  $r$  centered at  $x$ . For a compact set  $S \subseteq \mathbb{R}^n$ , let  $D_S := \sup_{x,y \in S} \|x - y\|$  denote its diameter,  $\bar{S}$  its closure,  $\partial S$  its boundary, and  $\text{int}(S)$  its interior. For  $\forall x \in \mathbb{R}^n$ , the distance from  $x$  to set  $S$  is  $d(x, S) := \inf_{y \in S} \|x - y\|$ .  $\lceil \cdot \rceil$  and  $\lfloor \cdot \rfloor$  denote the ceiling and floor operators, respectively.

## II. PRELIMINARIES AND PROBLEM FORMULATIONS

### A. Hamiltonian System

In this paper, we firstly review mathematical formulation of **Hamiltonian System**.

**Definition 1** (Hamiltonian System). *A Hamiltonian system without energy dissipation is a dynamical system of the form*

$$\dot{x} = f(x, u) = J(x)\nabla H(x) + G(x)u, \quad (1)$$

where  $x \in \mathcal{X} \subset \mathbb{R}^n$  is the state defined on a compact set  $\mathcal{X}$ ,  $H : \mathbb{R}^n \rightarrow \mathbb{R}$  is the Hamiltonian function,  $J(x) \in \mathbb{R}^{n \times n}$  is skew-symmetric, and  $u \in U \subset \mathbb{R}^m$  is the input of the system defined on a compact  $U$ .

Physically,  $H(x)$  describes the energy of system (1), i.e., sum of kinetic and potential energy in mechanical system, while the skew-symmetric matrix  $J(x)$  encodes the intrinsic power-conserving structure. In particular,  $J(x)$  determines how energy flows between state variables without creating or dissipating energy. We make the following assumptions about system (1).

**Assumption 1** (Bounded Hamiltonian Gradient). *The Hamiltonian function  $H$  is continuously differentiable on  $\mathcal{X}$  and there exists an upper bound  $L_H > 0$  such that  $\forall x \in \mathcal{X}$ ,  $\|\nabla H(x)\| \leq L_H$ .*

**Assumption 2** (Lipschitz Continuity of the Dynamics). *The dynamics  $f(x, u)$  in (1) are Lipschitz continuous w.r.t.  $x$  in both the state and the input. Namely, for any  $u \in U$ , there exist constants  $L > 0$  such that*

$$\|f(x_1, u) - f(x_2, u)\| \leq L\|x_1 - x_2\|, \forall x_1, x_2 \in \mathcal{X}.$$

**Remark 1** (Bounded Vector Field). *Since  $\mathcal{X}$  and  $U$  are compact, and  $f$  is continuous by Assumption 2, there exists a constant  $C_f > 0$  such that*

$$\|f(x, u)\| \leq C_f, \forall x \in \mathcal{X}, \forall u \in U. \quad (2)$$

Since the Hamiltonian  $H(\cdot)$  is conserved along the zero-input dynamics, it is natural to partition the state space into invariant energy layers.

**Definition 2** (Energy Layer). *For each energy value  $E \in H(\mathcal{X})$ , the corresponding energy layer is defined as*

$$\Sigma_E := \{x \in \mathcal{X} : H(x) = E\}.$$

$\Sigma_E$  is called an invariant energy layer because, under the zero-input, every trajectory starting in  $\Sigma_E$  remains in  $\Sigma_E$  for all future times.

### B. Target Reachability Problem

We now formalize the control objective considered in this paper. Given the Hamiltonian structure introduced above, our goal is to design control inputs that steer the system from a set of admissible initial conditions to a desired target set.

Let  $\mathcal{U}^{(0,t]}$  denote the set of admissible control signals on  $(0, t]$ , where each  $u : (0, t] \rightarrow U$  is piecewise continuous (and measurable); we also use  $\mathcal{U} := \mathcal{U}^{(0,\infty)}$ . Further, given any initial condition  $x \in \mathcal{X}$  and control input  $u \in \mathcal{U}^{(0,t]}$ , we use  $\phi(t, x, u)$  to denote the state of system (1) at time  $t$ .

Let  $S_0 \subseteq \mathcal{X}$  be a compact set of admissible initial states, and let  $S_{\text{tgt}} \subseteq \mathcal{X}$  be the prescribed target set.

**Problem 1** (Target Reachability). *Given system (1), an initial state  $x_0 \in S_0$ , and a target set  $S_{\text{tgt}}$ , determine a control signal  $u \in \mathcal{U}^{(0,t]}$  and a time  $t > 0$  such that*

$$\phi(t, x_0, u) \in S_{\text{tgt}}.$$

### C. Recurrence on Energy Layers

We next recall the recurrence structure of the zero-input dynamics associated with (1). Since the system is lossless, the Hamiltonian is conserved along zero-input trajectories. Hence, for each energy value  $E \in H(\mathcal{X})$ , the energy layer  $\Sigma_E := \{x \in \mathcal{X} : H(x) = E\}$  is invariant under the zero-input flow  $\phi(t, x, 0)$ . Our interest is in how trajectories repeatedly revisit dynamically relevant regions on each compact invariant energy layer. We first introduce invariant measures, which describe measures preserved by the zero-input flow.

**Definition 3** (Invariant Measure). *A probability measure  $\mu$  on  $M$  is said to be invariant under the flow  $\phi$  if for any measurable set  $A \subseteq M$  and any  $t \geq 0$ ,  $\mu(\phi(t, A)) = \mu(A)$ .*

To further characterize whether trajectories explore the whole invariant set or remain confined to smaller invariant subsets, we next introduce ergodicity.

**Definition 4** (Ergodic Measure). *Let  $\mu$  be an invariant probability measure on a set  $M$ . The measure  $\mu$  is said to be ergodic if for any measurable set  $A \subseteq M$  that is invariant under the flow, i.e.,  $\phi(t, A, 0) \subseteq A$  for all  $t \geq 0$ , it holds that  $\mu(A) \in \{0, 1\}$ .*

Intuitively, ergodicity means that trajectories are not confined to smaller invariant subsets, but instead propagate throughout  $M$ . In particular, for almost every initial condition, trajectories are dense in the support of  $\mu$ . The next theorem [24, Theorem 5.1.3] illustrates how every finite invariant measure admits an ergodic decomposition.

**Theorem 1** (Ergodic Decomposition on an Energy Layer). *Let  $\mu_E$  be an invariant measure on  $\Sigma_E$ . Then there exists a measurable family of ergodic measures  $\{\mu_\alpha^E\}_{\alpha \in \mathcal{A}_E}$  and a probability measure  $\nu_E$  on the index set  $\mathcal{A}_E$  such that  $\mu_E = \int_{\mathcal{A}_E} \mu_\alpha^E d\nu_E(\alpha)$ .*

For each  $\alpha \in \mathcal{A}_E$ , define the corresponding support of the ergodic component by

$$K_\alpha^E := \text{supp}(\mu_\alpha^E) \subseteq \Sigma_E,$$

where  $\text{supp}(\mu) := \{x \in \Sigma_E : \mu(\mathcal{B}_r(x)) > 0, \forall r > 0\}$  is the smallest closed subset of  $\Sigma_E$  that has full  $\mu$ -measure. In Theorem 1, each measure  $\mu_\alpha^E$  represents an ergodic component of the invariant dynamics on  $\Sigma_E$ , and  $K_\alpha^E$  is the closed region where the corresponding ergodic dynamics take place. On each  $K_\alpha^E$ , typical trajectories are dense [24, Proposition 4.3.5].

**Proposition 1** (Density of Typical Trajectories in an Ergodic Support). *For  $\mu_\alpha^E$ -almost every  $x \in K_\alpha^E$ , the forward orbit of  $x$  is dense in  $K_\alpha^E$ , namely*

$$\overline{\{\phi(t, x, 0) : t \geq 0\}} = K_\alpha^E.$$

*Hence, for  $\mu_\alpha^E$ -almost every initial condition in  $K_\alpha^E$ , the zero-input trajectory visits every neighborhood of every point in  $K_\alpha^E$  infinitely often.*

#### D. Chain Policies

Motivated by the nonparametric chain-policy idea in [18], we now specialize the policy construction to the reachability problem for the system (1). The basic idea is to build a finite library of demonstrated control snippets and then select among them according to the current state.

Suppose we are given a finite set of expert demonstrations  $\mathcal{D} := \{(x_j, u_j(\cdot), \tau_j)\}_{j=1}^M$ , where each  $u_j : (0, \tau_j] \rightarrow U$  is a piecewise continuous control signal, and  $x_j$  is the corresponding initial states of the expert demonstrations' trajectories of (1). In addition, let  $u_0 : (0, \tau_0] \rightarrow U$  be a prescribed default control signal, where  $\tau_0 > 0$ .

**Definition 5** (Control Alphabet). *A **control alphabet** is a finite collection of control signals*

$$\mathcal{A} := \{u_i : (0, \tau_i] \rightarrow U\}_{i=0}^M,$$

where each  $u_i$  is piecewise continuous and  $\tau_i > 0$ .

The control alphabet provides a library of candidate control snippets. To determine where each snippet should be applied in the state space, we introduce an assignment set.

**Definition 6** (Assignment Set). *An **assignment set** is a finite collection of verification triples*

$$\mathcal{K} := \{(x_i, r_i, u_i)\}_{i=1}^N \subseteq \mathbb{R}^n \times \mathcal{R}_{>0} \times \mathcal{A},$$

where  $x_i \in \mathbb{R}^n$  is the center state point,  $u_i \in \mathcal{A}$  is the control signal assigned to that region, and  $r_i > 0$  is its

effective radius. The support of  $\mathcal{K}$  is

$$\text{Supp}(\mathcal{K}) := \bigcup_{i=1}^N \mathcal{B}_{r_i}(x_i),$$

where  $N := |\mathcal{K}|$  is the size of the assignment set.

While an assignment set specifies regions that the control is effective, it does not by itself resolve which control to apply when balls overlap, nor what to do when a state lies outside  $\text{Supp}(\mathcal{K})$ . Based on assignment set, we introduce a normalized nearest-neighbor selection rule with a default fall-back option. For each  $x \in \mathcal{X}$ , define  $\rho_{\mathcal{K}}(x) := \min_{1 \leq i \leq N} \frac{\|x - x_i\|}{r_i}$ . The associated index map  $\iota_{\mathcal{K}} : \mathcal{X} \rightarrow \{0, 1, \dots, N\}$  is given by

$$\iota_{\mathcal{K}}(x) := \begin{cases} \arg \min_{1 \leq i \leq N} \frac{\|x - x_i\|}{r_i}, & \rho_{\mathcal{K}}(x) \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, if  $x \in \text{Supp}(\mathcal{K})$ , the rule selects the assignment whose normalized distance is minimal; otherwise, it selects the default control  $u_0$ . In the present analysis, we take  $u_0$  to be the zero input, so that when the state lies outside  $\text{Supp}(\mathcal{K})$ , the system follows the zero-input dynamics until it re-enters the support of the assignment set. Building on this rule, we now formalize the induced nonparametric policy.

**Definition 7** (Nonparametric Chain Policy). *Given an assignment set  $\mathcal{K}$  and a default control  $u_0$ , the nonparametric chain policy (NCP) is the map*

$$\pi_{\mathcal{K}} : \mathcal{X} \rightarrow \mathcal{A}$$

defined by  $\pi_{\mathcal{K}}(x) = u_{\iota_{\mathcal{K}}(x)}$ .

**Remark 2** (Execution of the Nonparametric Chain Policy). *Given an initial state  $x_0 = x$ , the policy  $\pi_{\mathcal{K}}$  induces an infinite-horizon control signal by concatenation. For each  $n \geq 0$ , define recursively  $u_n := \pi_{\mathcal{K}}(x_n)$ ,  $T_n := \tau(u_n)$ ,  $x_{n+1} := \phi(T_n, x_n, u_n)$ , where  $\tau(u_n)$  denotes the duration of the selected control snippet  $u_n$ . Let  $s_0 := 0$  and  $s_{n+1} := s_n + T_n$ . Then the induced control signal  $u_{\mathcal{K},x} : (0, \infty) \rightarrow U$  is defined by*

$$u_{\mathcal{K},x}(t) := u_n(t - s_n), \quad t \in [s_n, s_{n+1}).$$

### III. REACHABILITY IN HAMILTONIAN SYSTEMS

We are now ready to present the main results of this paper. We establish target reachability by combining two ingredients: local control actions that reduce an energy-based distance to the target, and recurrence of the zero-input Hamiltonian flow, which returns trajectories to regions where those actions can be reused. Repeating this interplay allows the trajectory to reach the target energy band and, ultimately, the target set. This separation between controlled energy reduction and passive recurrence underlies all results in this section.

### A. Target Reachability via Chain Policies

As mentioned above, our strategy for reachability is energy-based. We aim to design control actions that drive the system toward the energy levels associated with the target set. Since  $S_{\text{tgt}}$  is compact and connected, its image under the Hamiltonian is an interval,

$$H(S_{\text{tgt}}) = [H_{\min}, H_{\max}],$$

where  $H_{\min} := \min_{x \in S_{\text{tgt}}} H(x)$  and  $H_{\max} := \max_{x \in S_{\text{tgt}}} H(x)$ .

Once the trajectory reaches this energy band, the zero-input dynamics preserve energy, and the ergodic structure of each energy layer can be used to reach the target set, provided that the target is not dynamically isolated within the layer. This motivates the following assumption.

**Assumption 3** (Ergodic Component Coverage). *For every  $E \in [H_{\min}, H_{\max}]$  and every ergodic component  $K_\alpha^E \subseteq \Sigma_E$ ,*

$$S_{\text{tgt}} \cap K_\alpha^E \neq \emptyset.$$

**Remark 3.** *Assumption 3 is done for ease of exposition. Violation of this assumption would require a more sophisticated strategy that in philosophy does not depart from the presented here and is left for the journal version of this paper.*

To quantify progress toward the target set, we introduce an energy-based distance that measures how far a state lies from the target energy interval. This quantity will serve as a Barrier-like function that we aim to decrease through control actions.

**Definition 8** (Energy Signed Distance to  $S_{\text{tgt}}$ ). *Let  $H(S_{\text{tgt}}) = [H_{\min}, H_{\max}]$ , and define*

$$H_+^* := \frac{H_{\max} + H_{\min}}{2}, \quad H_-^* := \frac{H_{\max} - H_{\min}}{2}.$$

*The energy distance to the target set is defined as*

$$\Delta H(x) := |H(x) - H_+^*| - H_-^*.$$

*In particular,  $\Delta H(x) \leq 0$  if and only if  $H(x) \in H(S_{\text{tgt}})$ .*

We are therefore interested in finding controls that bring the system toward the set

$$H_{\text{tgt}} := \{x \in \mathcal{X} : \Delta H(x) \leq 0\}.$$

However, in order to provide guarantees, we will require our demonstrations to reach a slightly smaller set. Thus, for any  $0 < \epsilon < H_-^*$ , we define

$$H_{\text{tgt}}^\epsilon := \{x \in \mathcal{X} : \Delta H(x) \leq -\epsilon\}.$$

This leads to the following assumption.

**Assumption 4** (Reachability of  $H_{\text{tgt}}^\epsilon$ ). *For all  $x \in \mathcal{X} \setminus H_{\text{tgt}}^\epsilon$ , there exist  $T > 0$  and  $u \in \mathcal{U}^{(0, T]}$  such that*

$$\phi(T, x, u) \in H_{\text{tgt}}^\epsilon.$$

To quantify how quickly the system can be driven toward the target energy band, we introduce the corresponding first

hitting time. For any  $x \in \mathcal{X}$  and control signal  $u \in \mathcal{U}$ , define

$$T_\epsilon(x, u) := \inf\{t > 0 : \phi(t, x, u) \in H_{\text{tgt}}^\epsilon\}.$$

Thus, using the energy distance  $\Delta H(x)$ , we can compute the average decrease rate as

$$v_\epsilon(x, u) := \frac{\Delta H(x) - \Delta H(\phi(T_\epsilon(x, u), x, u))}{T_\epsilon(x, u)}. \quad (3)$$

By definition  $\phi(T_\epsilon(x, u), x, u) \in \partial H_{\text{tgt}}^\epsilon$ , thus

$$v_\epsilon(x, u) = \frac{\Delta H(x) + \epsilon}{T_\epsilon(x, u)}.$$

Thus the best achievable decrease rate at  $x$  is given by

$$v_\epsilon(x) := \sup_{u \in \mathcal{U}} v_\epsilon(x, u) = \frac{\Delta H(x) + \epsilon}{T_\epsilon^*(x)}, \quad (4)$$

where  $T_\epsilon^*(x)$  is the optimal hitting time maximized (4), i.e.,  $T_\epsilon^*(x) := \inf_{u \in \mathcal{U}} T_\epsilon(x, u)$ . This leads to our final requirement.

**Assumption 5** (Uniform Positive Energy Decrease Rate). *There exists  $\epsilon > 0$  such that*

$$\underline{v}_\epsilon := \inf_{x \in \mathcal{X} \setminus H_{\text{tgt}}^\epsilon} v_\epsilon(x) > 0.$$

We will use the above assumptions to ensure uniform energy decrease toward the target energy band and repeated opportunities to apply control through recurrence of the zero-input dynamics. Once the trajectory reaches this energy band, the ergodic structure of the Hamiltonian flow ensures eventual arrival to the target set.

**Theorem 2** (Target Reachability). *Consider system (1) under Assumptions 1–5. Let  $\mathcal{K} = \{(x_i, r_i, u_i)\}_{i=1}^N$  be a finite assignment set. Assume that  $\mathcal{K}$  satisfies the following:*

1) **Local energy decrease:** *For each  $(x_i, r_i, u_i) \in \mathcal{K}$ ,*

$$\Delta H(x_i(\tau_i)) + v_0 \tau_i + L_H r_i e^{L \tau_i} \leq \Delta H(x_i) - L_H r_i,$$

*where  $x_i(\tau_i) := \phi(\tau_i, x_i, u_i)$ ,  $\tau_{\min} = \min_i \tau_i$ , and  $v_0 > 0$ .*

2) **Energy coverage:** *Let  $c := \sup_{x \in S_0} \Delta H(x)$ . Then*

$$\Delta H(x) \leq c \implies H(x) \in H(\text{Supp}(\mathcal{K})).$$

3) **Ergodic coverage:** *For all  $E \in H(\text{Supp}(\mathcal{K}))$  and all ergodic components  $K_\alpha^E \subseteq \Sigma_E$ ,*

$$\text{int}(\text{Supp}(\mathcal{K})) \cap K_\alpha^E \neq \emptyset.$$

*Then, the chain policy  $\pi_{\mathcal{K}}$  ensures that for almost every  $x_0 \in S_0$ , there exists  $t < \infty$  such that*

$$\phi(t, x_0, \pi_{\mathcal{K}}) \in S_{\text{tgt}}.$$

*Proof.* We first establish three key points: (i) each control segment decreases the energy distance on its associated support ball, (ii) whenever the trajectory leaves the support, the zero-input dynamics return it to the support in finite time for almost every initial condition, and (iii) once the trajectory enters the target energy band, it reaches the target set in finite time for almost every initial condition. We then combine these three arguments to conclude the theorem.

**Step 1: Energy decrease on the support.** Let  $y \in \text{Supp}(\mathcal{K})$ . Then there exists  $i$  such that  $y \in \mathcal{B}_{r_i}(x_i)$ , i.e.,  $\|y - x_i\| \leq r_i$ . By Lipschitz continuity of the dynamics and Grönwall's inequality,

$$\|\phi(t, x_i, u_i) - \phi(t, y, u_i)\| \leq r_i e^{Lt}.$$

Since  $\|\nabla H(x)\| \leq L_H$ , the function  $\Delta H$  is Lipschitz with constant  $L_H$ , and therefore

$$\Delta H(\phi(\tau_i, y, u_i)) \leq \Delta H(\phi(\tau_i, x_i, u_i)) + L_H r_i e^{L\tau_i}, \quad (5)$$

$$\Delta H(x_i) \leq \Delta H(y) + L_H r_i. \quad (6)$$

Combining (5)–(6) with Condition 1 yields

$$\Delta H(\phi(\tau_i, y, u_i)) + v_0 \tau_i \leq \Delta H(y). \quad (7)$$

Thus, whenever the state lies in  $\text{Supp}(\mathcal{K})$ , the corresponding control segment strictly decreases the energy distance to the target set.

**Step 2: Return to the support.** Suppose that after applying a control segment from some  $y \in \text{Supp}(\mathcal{K})$ , the state

$$y' := \phi(\tau_i, y, u_i)$$

lies outside  $\text{Supp}(\mathcal{K})$ . From (7), we have

$$\Delta H(y') \leq \Delta H(y).$$

Since the trajectory starts from  $S_0$  and  $\Delta H$  decreases along each controlled execution, it follows that

$$\Delta H(y') \leq c := \sup_{x \in S_0} \Delta H(x).$$

By Condition 2, this implies that  $H(y') \in H(\text{Supp}(\mathcal{K}))$ . Notably, for a point  $y'$  s.t.  $H(y') \in H(\text{Supp}(\mathcal{K}))$ , the chain policy applies  $u \equiv 0$ , so the Hamiltonian is preserved and the trajectory remains on the energy layer

$$\Sigma_E, \quad E := H(y').$$

By Condition 3,  $\text{int}(\text{Supp}(\mathcal{K}))$  intersects every ergodic component  $K_\alpha^E \subseteq \Sigma_E$ . Hence, by Theorem 1 and Proposition 1, for almost every initial condition  $y' \in \Sigma_E$ , the zero-input trajectory  $\phi(t, y', 0)$  is dense in its ergodic component. Therefore, for almost every  $y'$  s.t.  $H(y') \in H(\text{Supp}(\mathcal{K}))$ , there exists a finite time  $T > 0$  such that

$$\phi(T, y', 0) \in \text{Supp}(\mathcal{K}).$$

**Step 3: Reachability inside the target energy band.** Suppose now that  $y \in H_{\text{tgt}}$ , i.e.,  $H(y) \in [H_{\min}, H_{\max}]$ . By Assumption 3, the target set  $S_{\text{tgt}}$  intersects every ergodic component of the energy layer  $\Sigma_{H(y)}$ . Since the zero-input dynamics preserve the Hamiltonian, the trajectory remains on  $\Sigma_{H(y)}$ . Therefore, by Theorem 1 and Proposition 1, for almost every initial condition  $y \in H_{\text{tgt}}$ , the zero-input trajectory is dense in its ergodic component and hence intersects  $S_{\text{tgt}}$  in finite time. That is, for almost every  $y \in H_{\text{tgt}}$ , there exists  $T' > 0$  such that  $\phi(T', y, 0) \in S_{\text{tgt}}$ .

**Conclusion.** Starting from any  $x_0 \in S_0$ , the chain policy alternates between controlled segments and zero-input evo-

lution. By Step 1, each controlled execution decreases  $\Delta H$  by at least  $v_0 \tau_{\min} > 0$ , and thus after at most  $N_* = \left\lceil \frac{c}{v_0 \tau_{\min}} \right\rceil$  executions the trajectory enters  $H_{\text{tgt}}$ . By Steps 2 and 3, each excursion outside the support returns in finite time, and once in  $H_{\text{tgt}}$  the trajectory reaches  $S_{\text{tgt}}$  in finite time for almost every initial condition.

Since only finitely many such events occur and the flow maps are continuous, the union of all exceptional null sets remains null after finitely many concatenation of controls. Therefore, for almost every  $x_0 \in S_0$ , there exists  $t < \infty$  such that  $\phi(t, x_0, \pi_{\mathcal{K}}) \in S_{\text{tgt}}$ .  $\square$

**Remark 4.** *Theorem 2 shows that reachability does not require coverage of the full state space. Instead, it is sufficient to cover (i) the one-dimensional energy interval connecting  $S_0$  to  $S_{\text{tgt}}$ , and (ii) the ergodic components within each corresponding energy layer. This reduces the coverage requirement from the full  $n$ -dimensional state space to a structure parameterized by energy and the ergodic index  $\alpha$ . In particular, reachability can be achieved by covering a set whose effective dimension is that of  $\alpha$  plus one, accounting for energy, without requiring demonstrations throughout the state space.*

## B. Existence of the Chain Policy

Theorem 2 provides conditions on the assignment set  $\mathcal{K}$  under which the chain policy guarantees target reachability. However, it is not a priori clear whether such conditions can be satisfied using a finite set of control segments. To address this question, we first derive upper and lower bounds on the quantities that govern energy decrease and recurrence, which will allow us to establish existence and sample complexity guarantees for  $\mathcal{K}$ .

**Lemma 1** (Velocity and Hitting-Time Bounds). *Under Assumption 5, the energy decrease rate and the first hitting time to  $H_{\text{tgt}}^\epsilon$  satisfy, for all  $x \in \mathcal{X} \setminus H_{\text{tgt}}$ ,*

$$v_\epsilon(x) \leq L_H C_f, \quad \text{and} \quad T_\epsilon^*(x) \geq \frac{\epsilon}{L_H C_f}.$$

*Proof.* From (2), we have

$$\begin{aligned} |\Delta H(x) - \Delta H(\phi(T, x, u))| &\leq L_H \|x - \phi(T, x, u)\| \\ &= L_H \left\| \int_0^T f(\phi(t, x, u), u(t)) dt \right\| \leq L_H C_f T. \end{aligned}$$

Combining the above with (3), it follows that

$$v_\epsilon(x, u) = \frac{\Delta H(x) - \Delta H(\phi(T_\epsilon(x, u), x, u))}{T_\epsilon(x, u)} \leq L_H C_f.$$

Hence  $v_\epsilon(x) \leq L_H C_f$ , and

$$v_\epsilon(x) = \frac{\Delta H(x) + \epsilon}{T_\epsilon^*(x)} \leq L_H C_f, \quad \Rightarrow \quad T_\epsilon^*(x) \geq \frac{\Delta H(x) + \epsilon}{L_H C_f}.$$

Since  $\Delta H(x) > 0$  for all  $x \in \mathcal{X} \setminus H_{\text{tgt}}$ , we obtain

$$T_\epsilon(x, u) \geq T_\epsilon^*(x) \geq \frac{\epsilon}{L_H C_f}, \quad \forall x \in \mathcal{X} \setminus H_{\text{tgt}}, \quad u \in \mathcal{U}. \quad \square$$

To establish the existence of the assignment set  $\mathcal{K}$ , we reduce the problem to a covering argument. In particular, constructing  $\mathcal{K}$  requires (i) controlling the complexity of the dynamics within each energy layer, and (ii) relating spatial coverage of  $\mathcal{X}$  to coverage of the corresponding energy values. The following two assumptions address these requirements.

**Assumption 6** (Ergodicity of Energy Layers). *For every  $E \in H(\text{Supp}(\mathcal{K}))$ , the energy layer  $\Sigma_E$  is ergodic, i.e.,*

$$\forall E \in H(\text{Supp}(\mathcal{K})), \quad K_\alpha^E = \Sigma_E.$$

**Assumption 7** (Strong Convexity of the Hamiltonian). *The Hamiltonian function  $H$  is strongly convex on  $\mathcal{X}$ , i.e., there exists  $\mu_H > 0$  such that for all  $x, y \in \mathcal{X}$ ,*

$$H(y) \geq H(x) + \nabla H(x)^\top (y - x) + \frac{\mu_H}{2} \|y - x\|^2.$$

**Remark 5.** *Assumption 6 reduces each energy layer to a single dynamically connected region, so that the relevant coverage is effectively one-dimensional and parameterized by energy. This corresponds to a simple setting, which includes, for example, Anosov energy surface and Axiom A systems [25]–[27]. In the numerical section, we also consider examples where this assumption is not satisfied. Extending the analysis to more general ergodic decompositions is left for future work.*

**Remark 6.** *Assumption 6 ensures that each energy layer behaves as a single dynamically connected region, eliminating the need to cover multiple ergodic components. Assumption 7 provides a regular relationship between distance in state space and variation in energy, allowing geometric coverings of  $\mathcal{X}$  to translate into coverage of the corresponding energy values. Together, these assumptions enable finite coverings that lead to the construction of  $\mathcal{K}$ .*

**Theorem 3** (Existence of Chain Policy). *Consider system (1) satisfying Assumptions 4–Assumption 7. Let  $v_0 \in (0, v_\epsilon)$ . Then there exists a nonparametric chain policy  $\pi_{\mathcal{K}}$ , constructed from a finite assignment set  $\mathcal{K} = \{(x_i, r_i, u_i)\}_{i=1}^N$ , where*

$$r_i = \frac{(v_\epsilon(x_i) - v_0)T_\epsilon^*(x_i)}{L_H(1 + e^{LT_\epsilon^*(x_i)})} > 0,$$

with  $u_i : (0, T_\epsilon^*(x_i)) \rightarrow U$  being the optimal control for reaching  $\partial H_{\text{tgt}}^\epsilon$  from  $x_i$ ,  $T_\epsilon^*(x_i)$  denoting its hitting time, such that the following holds:

- 1) **Reachability:** *For almost every  $x_0 \in S_0$ , there exists  $t < \infty$  such that  $\phi(t, x_0, \pi_{\mathcal{K}}) \in S_{\text{tgt}}$ .*
- 2) **Sample complexity:** *Let  $H(\text{Supp}(\mathcal{K})) = [H_1, H_2]$ . Then*

$$N \leq (H_2 - H_1) \frac{16L_H^2}{\mu_H(1 - \frac{v_0}{v_\epsilon})^2} \frac{\exp(\frac{2L(L_H D_{\mathcal{X}} + \epsilon)}{v_\epsilon})}{\epsilon^2}.$$

*Proof.* We construct a canonical assignment set and then verify the three conditions of Theorem 2.

**Step 1: Canonical local construction.** Let

$$\mathcal{X}_c := \{x \in \mathcal{X} : \Delta H(x) \leq c\} \setminus H_{\text{tgt}}^\epsilon, \quad c > 0$$

For every  $x \in \mathcal{X}_c$ , by Assumptions 4 and 5, choose an optimal control  $u_x : (0, T_\epsilon^*(x)) \rightarrow U$  such that

$$\phi(T_\epsilon^*(x), x, u_x) \in \partial H_{\text{tgt}}^\epsilon, \quad v_\epsilon(x) = \frac{\Delta H(x) + \epsilon}{T_\epsilon^*(x)}$$

where  $T_\epsilon^*(x)$  is the shortest reaching time from  $x$  to  $\partial H_{\text{tgt}}^\epsilon$ . Then we could define the certified radius

$$r(x) := \frac{(v_\epsilon(x) - v_0)T_\epsilon^*(x)}{L_H(1 + e^{LT_\epsilon^*(x)})}.$$

Since  $v_0 < v_\epsilon \leq v_\epsilon(x)$ , we have  $r(x) > 0$ . Moreover,

$$\begin{aligned} \Delta H(\phi(T_\epsilon^*(x), x, u_x)) + v_0 T_\epsilon^*(x) + L_H r(x) e^{LT_\epsilon^*(x)} \\ = \Delta H(x) - L_H r(x), \end{aligned}$$

so each triple  $(x, r(x), u_x)$  satisfies Condition 1 of Theorem 2.

**Step 2: Uniform lower bound on the certified radii.**

According to Assumption 5:  $v_\epsilon(x) \geq \frac{v_\epsilon}{v_\epsilon} > 0$ ,  $v_\epsilon(x) = \frac{\Delta H(x) + \epsilon}{T_\epsilon^*(x)}$ , we have  $T_\epsilon^*(x) \leq \frac{\Delta H(x) + \epsilon}{v_\epsilon}$ . Now recall that  $r = \frac{\Delta H(x) + \epsilon - v_0 T_\epsilon^*(x)}{L_H(1 + e^{LT_\epsilon^*(x)})}$ . Then

$$\Delta H(x) + \epsilon - v_0 T_\epsilon^*(x) \geq (\Delta H(x) + \epsilon) \left(1 - \frac{v_0}{v_\epsilon}\right) > 0,$$

because of  $0 < v_0 < v_\epsilon$ .

Therefore, for each fixed  $x \in \mathcal{X}_c$ , we obtain the lower bound of radius  $r(x)$

$$\begin{aligned} r(x) &\geq \frac{(\Delta H(x) + \epsilon) \left(1 - \frac{v_0}{v_\epsilon}\right)}{L_H \left(1 + \exp\left(\frac{L(\Delta H(x) + \epsilon)}{v_\epsilon}\right)\right)} \\ &\geq \frac{\epsilon \left(1 - \frac{v_0}{v_\epsilon}\right)}{2L_H \exp\left(\frac{L(L_H D_{\mathcal{X}} + \epsilon)}{v_\epsilon}\right)}, \end{aligned} \quad (8)$$

where equation (8) holds from the fact that  $H(x)$  is  $L_H$ -continuous.

**Step 3: Each certified ball covers a nontrivial energy interval.**

Fix  $x \in \mathcal{X}_c$  and consider the ball  $\mathcal{B}_{r(x)}(x)$ . By Assumption 7, for any  $\|v\| = 1$ , we have

$$H(x + rv) + H(x - rv) \geq 2H(x) + \mu_H r^2,$$

so at least one of  $x \pm rv$  is larger than the average value, that is to say,

$$\max\{H(x + rv), H(x - rv)\} \geq H(x) + \frac{\mu_H}{2} r^2. \quad (9)$$

For any  $y_1, y_2$  and  $y \in \mathcal{B}_{r(x)}(x)$ , according to (9) it follows

that

$$\begin{aligned} & \max_{y_1, y_2 \in \mathcal{B}_{r(x)}(x)} (H(y_1) - H(y_2)) \\ & \geq \max_{y \in \mathcal{B}_{r(x)}(x)} (H(y) - H(x)) \\ & \geq \frac{\mu_H}{2} r(x)^2. \end{aligned}$$

If  $H_+^* \notin H(\mathcal{B}_{r(x)}(x))$ , we have:

$$\begin{aligned} & \max_{y_1, y_2 \in \mathcal{B}_{r(x)}(x)} |\Delta H(y_1) - \Delta H(y_2)| \\ & = \max_{y_1, y_2 \in \mathcal{B}_{r(x)}(x)} |H(y_1) - H(y_2)| \\ & \geq \frac{\mu_H}{2} r(x)^2. \end{aligned}$$

And if  $H_+^* \in H(\mathcal{B}_{r(x)}(x))$ , since (9) holds, there exists at least one  $y^* \in \mathcal{B}_{r(x)}(x)$  such that  $H(x) + \frac{\mu_H}{2} r(x)^2 \leq H(y^*) \leq \max H(\mathcal{B}_{r(x)}(x))$ . Thus  $[H(x), H(x) + \frac{\mu_H}{2} r(x)^2] \subseteq H(\mathcal{B}_{r(x)}(x))$ , and therefore we have:

$$\max\{|H(x) - H_+^*|, |H(x) + \frac{\mu_H}{2} r(x)^2 - H_+^*|\} \geq \frac{\mu_H}{4} r(x)^2.$$

Consequently, for any  $y_1, y_2 \in \mathcal{B}_{r(x)}(x)$ , we have

$$\begin{aligned} & \max_{y_1, y_2 \in \mathcal{B}_{r(x)}(x)} |\Delta H(y_1) - \Delta H(y_2)| \\ & \geq \max_{y_1 \in \mathcal{B}_{r(x)}(x)} |H(y_1) - H_+^*| \\ & \geq \frac{\mu_H}{4} r(x)^2, \end{aligned} \quad (10)$$

where inequality (10) holds since we can chose  $y_2 \in \mathcal{B}_{r(x)}(x)$  such that  $H(y_2) = H_+^*$ .

Above all, we know that  $\forall x \in \mathcal{X}_c$ , we have

$$\begin{aligned} & \max_{y_1, y_2 \in \mathcal{B}_{r(x)}(x)} |\Delta H(y_1) - \Delta H(y_2)| \geq \frac{\mu_H}{4} r(x)^2 \\ & \geq \frac{\mu_H(1 - \frac{v_0}{v_\epsilon})^2}{16L_H^2} \frac{\epsilon^2}{\exp(\frac{2L(L_H D_{\mathcal{X}} + \epsilon)}{v_\epsilon})} \end{aligned} \quad (11)$$

**Step 4: Finite covering of the relevant energy interval.** Let

$$[H_1, H_2] := H(\mathcal{X}_c) \cup H(H_{\text{tgt}}^\epsilon).$$

For every  $E \in [H_1, H_2]$ , choose any  $x \in \mathcal{X}_c$  with  $H(x) = E$ . Then  $E \in H(\mathcal{B}_{r(x)}(x))$ , so the family  $\{H(\mathcal{B}_{r(x)}(x))\}_{x \in \mathcal{X}_c}$  covers  $[H_1, H_2]$ . Since  $[H_1, H_2]$  is compact and lower bound (11) holds, there exists a finite subcover

$$[H_1, H_2] \subseteq \bigcup_{i=1}^N H(\mathcal{B}_{r_i}(x_i)).$$

For each selected center  $x_i$ , define  $r_i := r(x_i)$ ,  $u_i := u_{x_i}$ , and let

$$\mathcal{K} := \{(x_i, r_i, u_i)\}_{i=1}^N,$$

where this finite selection implies

$$[H_1, H_2] \subseteq H(\text{Supp}(\mathcal{K})),$$

which verifies Condition 2 of Theorem 2.

**Step 5: Ergodic Coverage.** Moreover, under Assumption 6,

each energy layer in the relevant range is itself a unique ergodic component. Since  $\text{Supp}(\mathcal{K})$  intersects every energy level in  $[H_1, H_2]$ , Condition 3 of Theorem 2 also holds.

Therefore all three conditions of Theorem 2 are satisfied, and the resulting chain policy  $\pi_{\mathcal{K}}$  guarantees that for almost every  $x_0 \in S_0$ , there exists  $t < \infty$  such that

$$\phi(t, x_0, \pi_{\mathcal{K}}) \in S_{\text{tgt}}.$$

**Step 6: Sample complexity bound.** By (11), every selected interval  $\mathcal{B}_{r_i}(x_i)$  has length at least  $\eta$ . Hence a greedy interval-covering argument on  $[H_1, H_2]$  gives

$$\begin{aligned} N & \leq \frac{H_2 - H_1}{\frac{\mu_H(1 - \frac{v_0}{v_\epsilon})^2}{16L_H^2} \frac{\epsilon^2}{\exp(\frac{2L(L_H D_{\mathcal{X}} + \epsilon)}{v_\epsilon})}} \\ & = (H_2 - H_1) \frac{16L_H^2}{\mu_H} \frac{\exp(\frac{2L(L_H D_{\mathcal{X}} + \epsilon)}{v_\epsilon})}{\left(1 - \frac{v_0}{v_\epsilon}\right)^2 \epsilon^2}. \end{aligned}$$

This completes the proof.  $\square$

### C. Finite Time Reachability

The previous subsection establishes a general reachability theorem and the existence of a chain policy that realizes it. We now refine this qualitative result into a finite-time guarantee by deriving a uniform upper bound on the time required for the chain policy to drive the system to the target set. The key additional assumption is the uniform bounds on the uncontrolled hitting times:

**Assumption 8** (Upper Bound of Hitting Time). *Assume that there are upper bounds of the return time:*

- 1) *Return time to the support set  $\text{Supp}(\mathcal{K})$ :  $\forall x$  satisfies  $H(x) \in [H_1, H_2]$ , there exists a finite time  $T_1 > 0$ , s.t.  $\min_{t \in (0, T_1]} d(\phi(t, x, 0), \text{Supp}(\mathcal{K})) = 0$*
- 2) *Reaching time to the target set  $S_{\text{tgt}}$ :  $\forall x$  satisfies  $H(x) \in [H_{\min}, H_{\max}]$ , there exists a finite  $T_2 > 0$ , s.t.  $\min_{t \in (0, T_2]} d(\phi(t, x, 0), S_{\text{tgt}}) = 0$*

Under this assumption, the maximal time needed to reach the target set can be bounded explicitly.

**Theorem 4** (Finite-Time Reachability). *Let  $\mathcal{K}$  be an assignment set satisfying the conditions of Theorem 2, and define  $\tau_{\min} := \min_i \tau_i$ . Under Assumption 8, the time required to reach the target set from any initial state  $x \in S_0$  is uniformly bounded by*

$$T_{\max} \leq \frac{L_H D_{\mathcal{X}}}{v_0} \left(1 + \frac{T_1}{\tau_{\min}}\right) + T_2.$$

*Proof.* For any initial state  $x_0 \in S_0$ , Theorem 2 ensures that each time the trajectory leaves the support set  $\text{Supp}(\mathcal{K})$ , it returns to the support within time at most  $T_1$ .

Accordingly, we construct two sequences of states  $\{x_i\}_{i=0, \dots, N}$  and  $\{y_i\}_{i=0, \dots, N}$  as follows. Starting from  $x_i$ , there exists a time  $0 \leq t_{1,i} \leq T_1$  such that

$$y_i = \phi(t_{1,i}, x_i, 0) \in \text{Supp}(\mathcal{K}).$$

From  $y_i$ , applying the control  $u_i \in \mathcal{U}^{(0, \tau_i]}$  for time  $t_{2,i} := \tau_i$  yields

$$x_{i+1} = \phi(t_{2,i}, y_i, u_i),$$

with the energy decrease condition

$$\Delta H(x_{i+1}) + v_0 t_{2,i} \leq \Delta H(y_i),$$

for all  $i = 0, \dots, N$ . Thus the policy, iteratively induces the sequence  $x_i \xrightarrow{t_{1,i}} y_i \xrightarrow{t_{2,i}} x_{i+1}$ .

Now, let  $N$  be the smallest index such that  $x_N \in H_{\text{tgt}}^\epsilon$ . Then

$$N\tau_{\min} \leq \sum_{i=1}^N t_{2,i} \leq \frac{\Delta H(x_0)}{v_0},$$

which implies

$$N \leq \left\lfloor \frac{\Delta H(x_0)}{v_0 \tau_{\min}} \right\rfloor.$$

Since  $t_{1,i} \leq T_1$ , the total time to reach  $H_{\text{tgt}}$ , i.e.  $\bar{T}$  s.t.

$$\forall x_0 \in S_0, \exists T \leq \bar{T} \text{ s.t. } \phi(T, x_0, \pi_{\mathcal{K}}) \in H_{\text{tgt}},$$

satisfies

$$\bar{T} \leq \sum_{i=1}^N t_{2,i} + NT_1 \quad (12)$$

$$\leq \frac{\Delta H(x_0)}{v_0} + \left\lfloor \frac{\Delta H(x_0)}{v_0 \tau_{\min}} \right\rfloor T_1 \quad (13)$$

$$\leq \frac{L_H D_{\mathcal{X}}}{v_0} \left(1 + \frac{T_1}{\tau_{\min}}\right). \quad (14)$$

Finally, by Assumption 8, the maximum time  $T_{\max}$  to reach  $S_{\text{tgt}}$  from  $S_0$  satisfies

$$T_{\max} \leq \bar{T} + T_2 \leq \frac{L_H D_{\mathcal{X}}}{v_0} \left(1 + \frac{T_1}{\tau_{\min}}\right) + T_2. \quad \square$$

#### D. From Expert Demonstrations to NCPs

This subsection explains how the assignment set is constructed from expert demonstrations and how it induces the nonparametric chain policy introduced in Section II-D.

Consider a data set  $\mathcal{D} = \{(x_j, u_j(\cdot), T_j)\}_{j=1}^M$  consisting of  $M$  expert trajectories, where for each tuple  $(x_j, u_j(\cdot), T_j) \in \mathcal{D}$ , the trajectory  $\phi(t, x_j, u_j)$ ,  $t \in (0, T_j]$  satisfies,  $x_j \in S_0$  and  $\phi(T_j, x_j, u_j) \in S_{\text{tgt}}^\delta$ ,  $\forall j = 1, \dots, M$ , where  $S_{\text{tgt}}^\delta \subseteq H_{\text{tgt}}^\epsilon \cap S_{\text{tgt}}$ . The assignment set is obtained by extracting local control snippets and their certified radii along this trajectory. As Algorithm 1 shows, starting from an anchor time  $s \in [0, T_j]$ , define  $x_i := \phi(s, x_j, u_j)$ . Given a small  $v_0$ , for each candidate duration  $t \in (0, T_j - s]$ , let  $u_{i,t}$  be the restriction of  $u_j$  to  $(s, s+t]$ , so that  $\phi(t, x_i, u_{i,t}) = \phi(s+t, x_j, u_j)$ . Its certified radius is

$$r_i(t) = \frac{\Delta H(x_i) - \Delta H(\phi(t, x_i, u_{i,t})) - v_0 t}{L_H + L_H e^{L t}}.$$

If  $r_i(t) > 0$ , then  $u_{i,t}$  is valid on  $\mathcal{B}_{r_i(t)}(x_i)$ . We choose  $t_i \in \arg \max_{t \in (0, T_j - s], r_i(t) > 0} r_i(t)$ , set  $\tau_i := t_i$ ,  $u_i := u_{i,t_i}$ , and  $r_i := r_i(t_i)$ , and add  $(x_i, r_i, u_i)$  to  $\mathcal{K}$ . Then let  $\sigma_i := \inf\{\delta \in$

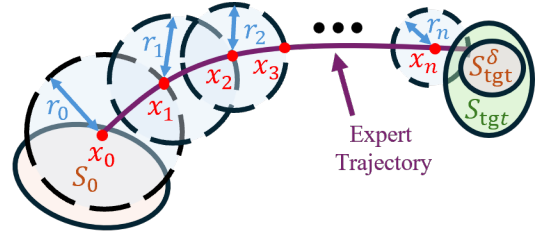


Fig. 1: Assignment set construction.

$(0, \tau_i] \mid \phi(s + \delta, x_j, u_j) \in \partial \mathcal{B}_{r_i}(x_i)\}$ , with  $\sigma_i := \tau_i$  if the set is empty. The next anchor is  $x_{i+1} := \phi(s + \sigma_i, x_j, u_j)$ , which is displayed in Fig 1. Repeating this procedure until the trajectory reaches  $S_{\text{tgt}}$ , and then over all demonstrations in  $\mathcal{D}$ , yields  $\mathcal{K} = \{(x_i, r_i, u_i)\}_{i=1}^N$ . After the assignment set is constructed, the NCP follows Remark 2 for  $\forall x \in \mathcal{X}$ .

---

#### Algorithm 1 Assignment-Set Construction

---

- 1: **Input:**  $\mathcal{D} = \{(x_j, u_j(\cdot), T_j)\}_{j=1}^M$
  - 2: **Output:**  $\mathcal{K}$
  - 3:  $\mathcal{K} \leftarrow \emptyset$
  - 4: **for**  $j = 1, \dots, M$  **do**
  - 5:    $s \leftarrow 0$
  - 6:   **while**  $s < T_j$  and  $\phi(s, x_j, u_j) \notin S_{\text{tgt}}$  **do**
  - 7:      $x_i \leftarrow \phi(s, x_j, u_j)$
  - 8:     choose  $t_i \in \arg \max_{t \in (0, T_j - s], r_i(t) > 0} r_i(t)$
  - 9:     **if** no such  $t_i$  exists **then**
  - 10:       **break**
  - 11:     **end if**
  - 12:      $\tau_i \leftarrow t_i$ ,  $u_i \leftarrow u_{i,t_i}$ ,  $r_i \leftarrow r_i(t_i)$
  - 13:      $\mathcal{K} \leftarrow \mathcal{K} \cup \{(x_i, r_i, u_i)\}$
  - 14:      $\sigma_i \leftarrow \inf\{\delta : \|\phi(s + \delta, x_j, u_j) - x_i\| = r_i\}$
  - 15:     **if** the set is empty **then**
  - 16:        $\sigma_i \leftarrow \tau_i$
  - 17:     **end if**
  - 18:      $s \leftarrow s + \sigma_i$
  - 19:   **end while**
  - 20: **end for**
- 

## IV. NUMERICAL SIMULATION

We evaluate the proposed NCP on two systems: a spring-mass system and a pendulum, and compare them with a vanilla Behavior Cloning (BC) baseline. The state is written as  $x = [q^\top, p^\top]^\top$ , where  $q$  and  $p$  denote the generalized coordinates and momenta. The goal is to drive the state to a small neighborhood of a target state  $x^*$ , namely  $S_{\text{tgt}} = \{x : \|x - x^*\| \leq \epsilon\}$ , where  $\epsilon = 0.1$ . The BC policy is a three-layer multilayer perceptron with hidden sizes (24, 24, 16), trained with Adam on the mean-squared imitation loss using learning rate  $1.2 \times 10^{-3}$ , weight decay  $5 \times 10^{-4}$ , and 40 epochs. Expert demonstrations are generated offline by a nonlinear model predictive controller (NMPC) implemented in Python using CasADi and do-mpc.

For each system, we uniformly sample 500 initial states from the set of states whose total energy is no greater than  $\bar{H}$

for testing. For each number  $M$  of expert demonstrations, we construct the proposed chain policy and train the BC baseline using the same  $M$  trajectories. We report the success rate and the average reach time, where unsuccessful trajectories are assigned the simulation horizon. The horizon is 20 s for the spring-mass system and 150 s for the single pendulum. All experiments are conducted on a 3.2 GHz AMD Ryzen 7 7735HS CPU with 16 GB RAM.

1) *Spring-Mass*: The spring-mass dynamics are given by

$$x = \begin{bmatrix} q \\ p \end{bmatrix} = \begin{bmatrix} q \\ m\dot{q} \end{bmatrix}, H(x) = \frac{p^2}{2m} + \frac{k}{2}q^2,$$

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \nabla H(x) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

Here,  $m = 1$  is the mass, and  $k = 1$  is the spring stiffness. In the simulations, the control input is constrained to  $u \in [-20, 20]$  and  $x^* = [0, 0]^\top$ .

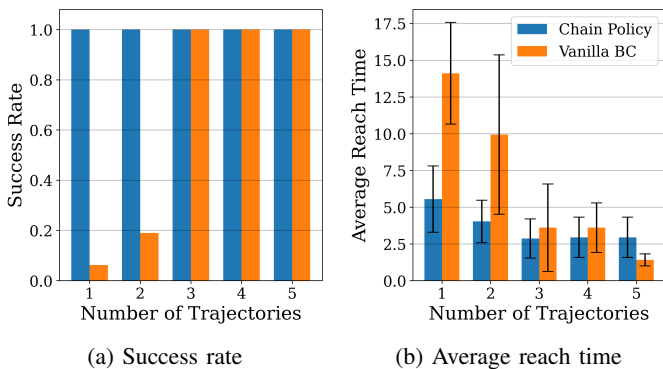


Fig. 2: Spring-mass system results.

As shown in Fig. 2a, the proposed chain policy achieves a success rate of 1.0 for all tested numbers of expert trajectories, indicating that the task can be solved reliably with very limited demonstration data. In contrast, vanilla BC performs poorly when only a small number of trajectories are available, with success rates of only 0.062 and 0.19 for  $M = 1$  and  $M = 2$ , respectively. Fig. 2b further shows that the average reach time of the chain policy decreases from 5.54s at  $M = 1$  to about 2.94s when more demonstrations are provided. The largest improvement occurs between  $M = 1$  and  $M = 3$ , after which the performance becomes nearly saturated. This suggests that adding a small number of expert trajectories is sufficient to substantially improve the efficiency of the learned assignment set in this relatively simple system.

2) *Single Pendulum*: The single pendulum dynamics are

$$x = \begin{bmatrix} q \\ p \end{bmatrix} = \begin{bmatrix} \theta \\ m\ell^2\dot{\theta} \end{bmatrix}, H(x) = \frac{p^2}{2m\ell^2} + mg\ell(1 - \cos q),$$

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \nabla H(x) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u.$$

Here,  $m = 1$  is the pendulum mass,  $\ell = 2$  is the pendulum length, and  $g = 9.81\text{m/s}^2$  is the gravitational acceleration. In the simulations, the control torque is constrained to  $u \in [-20, 20]$  and  $x^* = [\pi, 0]^\top$ .

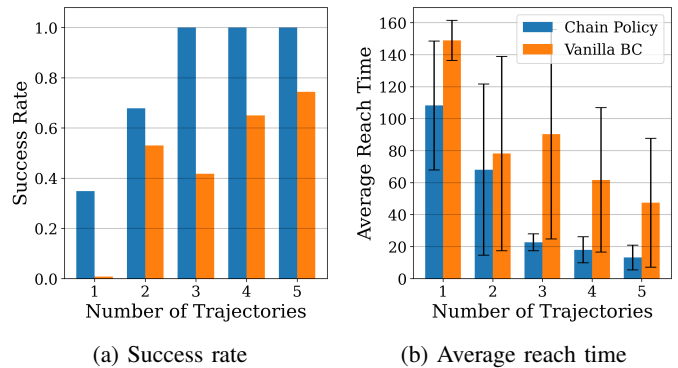


Fig. 3: Single pendulum results.

In this task, as Fig 3a shows, the success rate of the proposed chain policy increases from 0.348 at  $M = 1$  to 0.678 at  $M = 2$ , and reaches 1.0 for all  $M \geq 3$ . However, vanilla BC remains consistently inferior, with success rates 0.008, 0.53, 0.418, 0.65, and 0.744 for  $M = 1, \dots, 5$ , respectively. Besides, the average reach time shown in Fig. 3b also decreases significantly as the number of expert trajectories grows, from 114.71s at  $M = 1$  to 13.16s at  $M = 5$ . This trend indicates that enlarging the assignment set with additional expert trajectories substantially improves the ability of the chain policy to find effective local control snippets and steer the system to the target set more efficiently.

## V. CONCLUSIONS AND FUTURE WORK

This paper proposed a data-driven control framework for Hamiltonian systems that leverages physical structure as an inductive bias. Rather than learning policies over the entire state space, we exploit energy conservation and recurrence to construct control laws from a finite collection of locally verified trajectory segments. By combining controlled energy reduction with recurrence on invariant energy layers, we establish target-set reachability and derive finite-time guarantees under suitable conditions. Numerical examples demonstrate the effectiveness of the approach.

Future work will focus on extending these ideas to more general settings, including systems with multiple ergodic components and partial observability, as well as robust formulations that account for model uncertainty and energy dissipation [28]. Another direction is to develop principled strategies for data selection and augmentation that improve coverage of the relevant low-dimensional structures while mitigating compounding errors [29].

## REFERENCES

- [1] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
- [2] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [4] S. Oymak and N. Ozay, “Non-asymptotic identification of linear dynamical systems from a single trajectory,” *arXiv preprint arXiv:1806.05722*, 2018.

- [5] Y. Zheng and N. Li, “Non-asymptotic identification of linear dynamical systems using multiple trajectories,” *IEEE Control Systems Letters*, vol. 5, no. 5, pp. 1693–1698, 2021.
- [6] Y. Hu, A. Wierman, and G. Qu, “On the sample complexity of stabilizing lti systems on a single trajectory,” in *Advances in Neural Information Processing Systems*, vol. 35, 2022, pp. 16 989–17 002.
- [7] S. W. Werner and B. Peherstorfer, “On the sample complexity of stabilizing linear dynamical systems from data,” *Foundations of Computational Mathematics*, vol. 24, no. 3, pp. 955–987, 2024.
- [8] L. F. Toso, L. Ye, and J. Anderson, “Learning stabilizing policies via an unstable subspace representation,” in *IEEE Conference on Decision and Control (CDC)*, IEEE, 2025, pp. 7543–7550.
- [9] S. Dean, H. Mania, N. Matni, B. Recht, and S. Tu, “On the sample complexity of the linear quadratic regulator,” *Foundations of Computational Mathematics*, vol. 20, no. 4, pp. 633–679, 2020.
- [10] C. De Persis and P. Tesi, “Formulas for data-driven control: Stabilization, optimality, and robustness,” *IEEE Transactions on Automatic Control*, vol. 65, no. 3, pp. 909–924, 2020.
- [11] J. Coulson, J. Lygeros, and F. Dörfler, “Data-enabled predictive control: In the shallows of the deep,” in *European Control Conference (ECC)*, IEEE, 2019, pp. 307–312.
- [12] J. Berberich, J. Köhler, M. A. Müller, and F. Allgöwer, “Data-driven model predictive control with stability and robustness guarantees,” *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1702–1717, 2021.
- [13] T. Dai and M. Sznaier, “A semi-algebraic optimization approach to data-driven control of continuous-time nonlinear systems,” *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 487–492, 2020.
- [14] M. Guo, C. De Persis, and P. Tesi, “Data-driven stabilization of nonlinear polynomial systems with noisy data,” *IEEE Transactions on Automatic Control*, vol. 67, no. 8, pp. 4210–4217, 2021.
- [15] R. Strässer, J. Berberich, and F. Allgöwer, “Data-driven control of nonlinear systems: Beyond polynomial dynamics,” in *IEEE Conference on Decision and Control (CDC)*, IEEE, 2021, pp. 4344–4351.
- [16] N. Monshizadeh, C. De Persis, and P. Tesi, “A versatile framework for data-driven control of nonlinear systems,” *IEEE Transactions on Automatic Control*, 2025, to appear.
- [17] N. M. Boffi, S. Tu, N. Matni, J.-J. Slotine, and V. Sindhvani, “Learning stability certificates from data,” *arXiv preprint arXiv:2008.05952*, 2020.
- [18] R. Siegelmann and E. Mallada, “Data-driven practical stabilization of nonlinear systems via chain policies: Sample complexity and incremental learning,” *arXiv preprint arXiv:2510.03982*, 2025.
- [19] T. Lew, L. Janson, R. Bonalli, and M. Pavone, “A simple and efficient sampling-based algorithm for general reachability analysis,” in *Learning for Dynamics and Control Conference*, PMLR, 2022, pp. 1086–1099.
- [20] R. Siegelmann, Y. Shen, F. Paganini, and E. Mallada, “A recurrence-based direct method for stability analysis and gpu-based verification of non-monotonic lyapunov functions,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*, IEEE, 2023, pp. 6665–6672.
- [21] H. Sibai and E. Mallada, “Recurrence of nonlinear control systems: Entropy, bit rates, and finite alphabet controllers,” *Nonlinear Analysis: Hybrid Systems*, vol. 59, p. 101 649, 2026.
- [22] J. Liu and E. Mallada, “Recurrent control barrier functions: A path towards nonparametric safety verification,” in *2025 IEEE 64th Conference on Decision and Control (CDC)*, IEEE, 2025, pp. 7721–7727.
- [23] J. Liu and E. Mallada, “Safety-critical control via recurrent tracking functions,” *arXiv preprint arXiv:2510.01147*, 2025.
- [24] M. Viana and K. Oliveira, *Foundations of ergodic theory*. Cambridge University Press, 2016.
- [25] J. F. Plante, “Anosov flows,” *American Journal of Mathematics*, vol. 94, no. 3, pp. 729–754, 1972.
- [26] E. Hopf, “Ergodic theory and the geodesic flow on surfaces of constant negative curvature,” 1971.
- [27] R. Bowen and D. Ruelle, “The ergodic theory of axiom a flows,” in *The theory of chaotic attractors*, Springer, 1975, pp. 55–76.
- [28] J. G. Romero, “A robust adaptive velocity observer for mechanical systems transformed in cascade form,” *Automatica*, vol. 165, p. 111 671, 2024.
- [29] T. T. Zhang, D. Pfrommer, C. Pan, N. Matni, and M. Simchowitz, “Action chunking and data augmentation yield exponential improvements in behavior cloning for continuous spaces,” in *International Conference on Learning Representations (ICLR)*, 2026.