On the Imitation of Non-Markovian Demonstrations: From Low-Level Stability to High-Level Planning

Anonymous Authors[1]

Abstract

We propose a theoretical framework for studying the imitation of stochastic, non-Markovian, potentially multi-modal expert demonstrations in nonlinear dynamical systems. Our framework invokes low-level controllers - either learned or implicit in position-command control - to stabilize imitation policies around expert demonstrations. We show that with (a) a suitable low-level stability guarantee and (b) a stochastic continuity property of the learned policy we call "total variation continuity" (TVC), an imitator that accurately estimates actions on the demonstrator's state distribution closely matches the demonstrator's distribution over entire trajectories. We then show that TVC can be ensured with minimal degradation of accuracy by combining a popular data-augmentation regimen with a novel algorithmic trick: adding augmentation noise at execution time. We instantiate our guarantees for policies parameterized by diffusion models and prove that if the learner accurately estimates the score of the (noise-augmented) expert policy, then the distribution of imitator trajectories is close to the demonstrator distribution in a natural optimal transport distance. Our analysis constructs intricate couplings between noise-augmented trajectories, a technique that may be of independent interest. We conclude by empirically validating our algorithmic recommendations.

1. Introduction

Training dynamic agents from datasets of expert examples, known as *imitation learning*, promises to take advantage of the plentiful demonstrations available in the modern data environment, in an analogous manner to the recent successes of language models conducting unsupervised learning on enormous corpora of text (Thoppilan et al., 2022; Vaswani et al., 2017). Imitation learning is especially exciting in robotics, where mass stores of pre-recorded demonstrations on Youtube (Abu-El-Haija et al., 2016) or cheaply collected simulated trajectories (Mandlekar et al., 2021; Dasari et al., 2019) can be converted into learned robotic policies.

An outstanding challenge for imitation learning is that demonstrator policies correlate with past actions in sophisticated ways. Multi-modal trajectories present a key example. Consider a robot navigating around an obstacle; because there is no difference between navigating around the object to the right and around to the left, the dataset of expert trajectories may include examples of both options. This bifurcation of good trajectories can make it difficult for the agent to effectively choose which direction to go, possibly even causing the robot to oscillate between directions and run into the object instead of going around it (Chi et al., 2023). Crucially, human demonstrators correlate current actions with the past in order to *commit* to either a right or left path, which makes even formulating the idea of an "expert *policy*" a conceptually challenging one.

In this paper, we develop a theory of imitation learning flexible enough to imitate non-Markovian (e.g. multi-modal or bifurcated as in the above example) demonstrations in smooth, nonlinear control systems. As in previous work, we formalize imitation learning in two stages: at *train-time*, we learn a map from observations to distributions over actions, supervised by (state, action)-pairs from expert demonstrations, while at *test-time*, the learned map, or *policy*, is executed on random initial states (distributed identically to initial training states). What makes imitation learning more challenging than supervised learning is the problem of compounding errors, which may bring the agent to regions of state space not seen during training. Unless one is permitted to collect data adaptively (Laskey et al., 2017; Ross et al., 2011), it is understood that some form of "stability" is required so that the agent navigates back from deviations (Tu et al., 2022; Havens & Hu, 2021).

Contributions. We propose a hierarchical formulation of stability to analyze imitation learning. During training, the

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the Workshop on New Frontiers in Learning, Control, and Dynamical Systems at the International Conference on Machine Learning (ICML). Do not distribute.

learner synthesizes sequences of *primitive controllers* - time-varying affine control policies which locally stabilize around each demonstration trajectory. We break these {demonstrator trajectory, primitive controller} pairs into sub-trajectories we call "chunks." Building on (Chi et al., 2023), we use DDPMs to estimate the conditional distribution of primitive controller chunks conditioned on recent states from the previous chunk. We also adopt a popular data-augmentation technique that corrupts trajectories (but not supervising actions) with a small amount of Gaussian noise (Ke et al., 2021; Laskey et al., 2017; Ross et al., 2011). Unlike prior work, we propose adding noise *back into the policies at inference time*, a technique which is both both provably indispensable in our analysis, and which our simulations suggest yields considerable benefit over the conventional approach of not adding noise at inference time.

We prove that the learner can approximate the expert's trajectory distribution provided three conditions hold: along each expert trajectory, (a) the dynamics are sufficiently smooth; (b) one can synthesize primitive controllers that stabilize the Jacobian-linearized dynamics; and (c) one can approximately sample from conditional distributions over sequences of primitive controllers. For concreteness, we formulate part (c) in the language of Denoising Diffusion Probabilistic Models (DDPMs), although our results hold for arbitrary generative models. Our notion of trajectory approximation is a natural optimal transport metric, which considers a Wassertstein-like distance between the *marginal* distributions of visited states, which is strong enough to ensure closeness of Lipschitz trajectory costs which decompose across time-steps.

Our analysis reformulates our setting as imitation in a composite MDP, where composite states s_h corresponds to trajectory chunks, and composite-actions a_h correspond to sub-sequences of primitive controllers. A learner's policy maps composite-states to distributions over composite-actions, and a marginalization trick lets us represent non-Markovian demonstrator trajectories in the same format. The primitive controller sequences a_h provide the requisite stability, and we show that noising the learner policy at inference time ensures continuity in the total variation distance (TVC). Our proof is inspired by the notion of replica symmetry in statistical physics (Mezard & Montanari, 2009): we show that by noising at inference time, we consistently estimate a "replica" policy, which, up to the stability of controllers, has marginals over states and actions close to those of the expert policy. The proof constructs a sophisticated coupling between the learned policy, replica policy, and other interpolating sequences; this construction is enabled by subtle measure-theoretic arguments demonstrating consistency of our couplings. We also establish stability guarantees for sequences of primitive controllers in non-linear control systems, which may be of independent interest. Finally, we empirically validate the benefits of our proposed augmentation strategy in simulated robotic manipulation tasks.

Abridged Related Work. Due to space, we defer a full comparison to past work to Appendix B. DDPMs, proposed in (Ho et al., 2020; Sohl-Dickstein et al., 2015), along with their relatives have seen success in image generation (Song & Ermon, 2019; Ramesh et al., 2022), along with imitation learning (without data augmentation) (Janner et al., 2022; Chi et al., 2023; Pearce et al., 2023), which is the starting point of our work. Data augmentation is ubiquitous in modern imitation learning (Laskey et al., 2017) and our approach corresponds to that of (Ke et al., 2021) but with noise added at inference time. Despite the benefits of adaptive data collection (Ross et al., 2011; Laskey et al., 2017), adaptive demonstrations are more expensive to collect. Previous analyses of imitation learning without adaptive data collection have focused on classical control-theoretic notions of stability, notably incremental stability, (Tu et al., 2022; Havens & Hu, 2021; Pfrommer et al., 2022), which require continuity, Markovianity, and often determinism, and preclude the bifurcations permitted in our setting.

Organization. In Section 2 we formally introduce our setting as well as some preliminary notation and our main desideratum. We then state our assumptions and our proposed algorithm, TODA before giving our main guarantee (Theorem 1) in Section 3. In Section 4 we describe our proof techniques and provide a high level overview before concluding with some experiments in Section 5. The organization of our many appendices is given in Appendix A.

2. Setting

Notation and Preliminaries. Appendix A gives a full review of notation. Bold lower-case (resp. upper-case) denote vectors (resp. matrices). We abbreviate the concatenation of sequences via $\mathbf{z}_{1:n} = (\mathbf{z}_1, \dots, \mathbf{z}_n)$. Norms $\|\cdot\|$ are Euclidean for vectors and operator norms for matrices unless otherwise noted. Rigorous probability-theoretic preliminaries are provided in Appendix C. In short, all random variables take values in Polish spaces $\mathcal{X}$ (which include real vector spaces), the space of Borel distributions on $\mathcal{X}$ is denoted $\Delta(\mathcal{X})$. We rely heavily on ***coupling***s from optimal transport theory: given measures $X \sim \mathsf{P}$ and $X' \sim \mathsf{P}'$ on $\mathcal{X}$ and $\mathcal{X}'$ respectively, $\mathscr{C}(\mathsf{P}, \mathsf{P}')$ denotes the space of joint distributions $\mu \in \Delta(\mathcal{X} \times \mathcal{X}')$ called "couplings" such that $(X, X') \sim \mu$ has marginals $X \sim \mathsf{P}$ and $X' \sim \mathsf{P}$. $\Delta(\mathcal{X} \mid \mathcal{Y})$ denotes the space of ***kernels*** $\mathsf{Q} : \mathcal{Y} \to \Delta(\mathcal{X})$; Appendix C rigorously justifies that, in our setting, all conditional distributions can be expressed as kernels (which we do throughout the paper without comment).

Dynamics and Demonstrations. We consider a discrete-time, control system with states $\mathbf{x}_t \in \mathcal{X} := \mathbb{R}^{d_x}$, and inputs

$\mathbf{u}_t \in \mathcal{U} := \mathbb{R}^{d_u}$, obeying the following nonlinear dynamics

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t), \quad t \geq 1. \tag{2.1}$$

Given length $T \in \mathbb{N}$, we call sequences $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T}) \in \mathscr{P}_T := \mathcal{X}^{T+1} \times \mathcal{U}^T$ ***trajectories***. For simplicity, we assume that (2.1) deterministic and address stochastic dynamics in Appendix J. Though the dynamics are Markov and deterministic, we consider a stochastic and possibly *non-Markovian* demonstrator, which allows for the multi-modality described in the Section 1.

Definition 2.1 (Expert Distribution)**.** Let $\mathcal{D}_{\mathrm{exp}} \in \Delta(\mathscr{P}_T)$ denote an ***expert distribution*** over trajectories to be imitated. $\mathcal{D}_{\mathbf{x}_1}$ denotes the distribution of $\mathbf{x}_1$ under $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T}) \sim \mathcal{D}_{\mathrm{exp}}$.

Primitive Controllers and Synthesis Oracle. Let $\mathcal{K}$ denote the space of affine mappings $\mathcal{X} \to \mathcal{U}$ (redundantly) parameterized as $\mathbf{x} \mapsto \bar{\mathbf{u}} + \bar{\mathbf{K}}(\mathbf{x} - \bar{\mathbf{x}})$; we call these ***primitive controllers***. We say $\kappa_{1:T} \in \mathcal{K}^T$ is *consistent with* a trajectory $\boldsymbol{\rho} = (\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T}) \in \mathscr{P}_T$ if $\bar{\mathbf{x}}_t = \mathbf{x}_t$ and $\bar{\mathbf{u}}_t = \mathbf{u}_t$ for all $t \in [T]$; note that this implies that $\kappa_t(\mathbf{x}_t) = \mathbf{u}_t$ for all t. A ***synthesis oracle*** $\mathtt{synth}$ maps $\mathscr{P}_T \to \mathcal{K}^T$ such that, for all $\boldsymbol{\rho}_T \in \mathscr{P}_T$, $\kappa_{1:T} = \mathtt{synth}(\boldsymbol{\rho}_T)$ is consistent with $\boldsymbol{\rho}_T$. For our theory, we assume access to a synthesis oracle at training time, and assume the ability to estimate conditional distributions over joint sequences of primitive controllers; Appendix G explains how this can be implemented by solving Ricatti equations if dynamics are known (e.g. in a simulator), smooth, and stabilizable. In our experimental environment, control inputs are desired robot configurations, which the simulated robot executes by applying feedback gains.

Chunking Policies and Indices. The expert distribution $\mathcal{D}_{\mathrm{exp}}$ may involve non-Markovian sequences of actions. We imititate these sequences via ***chunking policies***. Fix a ***chunk length*** $\tau_{\mathrm{c}} \in \mathbb{N}$ and ***memory length*** $\tau_{\mathrm{m}} \leq \tau_{\mathrm{c}}$, and define time indices $t_h = (h-1)\tau_{\mathrm{c}} + 1$. For simplicity, we assume τ_{c} divides T, and set $H = T/\tau_{\mathrm{c}}$. Given a $\boldsymbol{\rho}_T \in \mathscr{P}_T$, define the ***trajectory-chunks*** $\boldsymbol{\rho}_{\mathrm{c},h} := (\mathbf{x}_{t_{h-1}:t_h}, \mathbf{u}_{t_{h-1}:t_h-1}) \in \mathscr{P}_{\tau_{\mathrm{c}}}$ and ***memory-chunks*** $\boldsymbol{\rho}_{\mathrm{m},h} := (\mathbf{x}_{t_h-\tau_{\mathrm{m}}+1:t_h}, \mathbf{u}_{t_h-\tau_{\mathrm{m}}+1:t_h-1}) \in \mathscr{P}_{\tau_{\mathrm{m}}-1}$ for $h > 1$, and $\boldsymbol{\rho}_{\mathrm{c},1} = \boldsymbol{\rho}_{\mathrm{m},1} = \mathbf{x}_1$. We call τ_{c}-length sequences of primitive controllers ***composite actions*** $\mathsf{a}_h = \kappa_{t_h:t_{h-1}} \in \mathcal{A} := \mathcal{K}^{\tau_{\mathrm{c}}}$. A ***chunking policy*** $\pi = (\pi_h)$ consists of functions π_h mapping memory chunks $\boldsymbol{\rho}_{\mathrm{m},h}$ to distributions $\Delta(\mathcal{A})$ over composite actions and interacting with the dynamics (2.1) by $\mathsf{a}_h = \kappa_{t_h:t_{h-1}} \sim \pi_h(\boldsymbol{\rho}_{\mathrm{m},h})$, and executing $\mathbf{u}_t = \kappa_t(\mathbf{x}_t)$. The chunking scheme is represented in Figure 1 in Section 4, alongside the abstraction we use in our analysis.

Desideratum. The quality of imitation of a deterministic policy is naturally measured in terms of step-wise closeness of state and action (Tu et al., 2022; Pfrommer et al., 2022). In stochastic settings, however, two rollouts of even the same policy can visit different states. We propose measuring *distributional closeness* via *couplings* introduced in the preliminaries above. We define the following losses:

Definition 2.2. Given $\varepsilon > 0$ and a (chunking) policy π, the imitation loss $\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi)$ is defined to be

$$\max_{t \in [T]} \inf_{\mu} \left\{ \mathbb{P}_\mu \left[\|\mathbf{x}_{t+1}^{\mathrm{exp}} - \mathbf{x}_{t+1}^{\pi}\| > \varepsilon \right], \mathbb{P}_\mu \left[\|\mathbf{u}_{t}^{\mathrm{exp}} - \mathbf{u}_{t}^{\pi}\| > \varepsilon \right] \right\}$$

where the infimum is over all couplings μ between the distribution of $\boldsymbol{\rho}_T$ under $\mathcal{D}_{\mathrm{exp}}$ and that induced by the policy π as described above, such that $\mathbb{P}_\mu[\mathbf{x}_1^{\mathrm{exp}} = \mathbf{x}_1^{\pi}] = 1$. Also define $\mathcal{L}_{\mathrm{fin},\varepsilon}(\pi) := \inf_\mu \mathbb{P}_\mu \left[\|\mathbf{x}_{T+1}^{\mathrm{exp}} - \mathbf{x}_{T+1}^{\pi}\| > \varepsilon \right]$, the loss restricted to the final states under each distribution.

Under stronger conditions (whose necessity we establish), we can also imitate joint distributions over actions (Appendix I). Observe that $\mathcal{L}_{\mathrm{fin},\varepsilon} \leq \mathcal{L}_{\mathrm{marg},\varepsilon}$, and that both losses are equivalent to Wasserstein-type metrics on bounded domains (and correspond to total variation analogues of shifted Renyi divergences (Altschuler & Talwar, 2022; Altschuler & Chewi, 2023)). While empirically evaluating these infima over couplings is challenging, $\mathcal{L}_{\mathrm{marg},\varepsilon}$ upper bounds the difference in expectation between any bounded and Lipschitz control cost decomposing across time steps, states and inputs, and $\mathcal{L}_{\mathrm{fin},\varepsilon}$ upper bounds differences in bounded, Lipschitz final-state costs; see Appendix I for further discussion.

Diffusion Models. Our analysis provides imitiation guarantees when chunking policies π_h select a_h via a sufficiently accurate generative model. Given their recent success, we adopt the popular Denoising Diffusion Probabilistic Models (DDPM) framework (Chen et al., 2022; Lee et al., 2023) that allows the learner to sample from a density $q \in \Delta(\mathbb{R}^d)$ assuming that the *score* $\nabla \log q$ is known to the learner. More precisely, suppose the learner is given an observation $\boldsymbol{\rho}_{\mathrm{m},h}$ and wishes to sample $\mathsf{a}_h \sim q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h})$ for some family of probability kernels $q(\cdot|\cdot)$. A DDPM starts with some a_h^0 sampled from a standard Gaussian noise and iteratively "denoises" for each DDPM-time step $0 \leq j < J$:

$$\mathsf{a}_h^j = \mathsf{a}_h^{j-1} - \alpha \cdot \mathbf{s}_{\theta,h}(\mathsf{a}_h^{j-1}, \boldsymbol{\rho}_{\mathrm{m},h}, j) + 2 \cdot \mathcal{N}(0, \alpha^2 \mathbf{I}), \tag{2.2}$$

where $\mathbf{s}_{\theta,h}(\mathsf{a}_h^j, \boldsymbol{\rho}_{\mathrm{m},h}, j)$ estimates the true score $\mathbf{s}_{\star,h}(\mathsf{a}_h, \boldsymbol{\rho}_{\mathrm{m},h}, \alpha j)$, formally defined for any continuous argument $t \leq J\alpha$ to be $\mathbf{s}_{\star,h}(\mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},h}, t) := \nabla_{\mathsf{a}} \log q_{[t]}(\mathsf{a} \mid \boldsymbol{\rho}_{\mathrm{m},h})$, where $q_{[t]}(\cdot|\boldsymbol{\rho}_{\mathrm{m},h})$ is the distribution of $e^{-t}\mathsf{a}_h^{(0)} + \sqrt{1 - e^{-2t}}\boldsymbol{\gamma}$ with $\mathsf{a}_h^{(0)} \sim q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h})$ and $\boldsymbol{\gamma} \sim \mathcal{N}(0, \mathbf{I})$ is a standard Gaussian. We will denote by $\mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h})$ the law of a_h^J sampled according to the DDPM using $\mathbf{s}_\theta(\cdot, \boldsymbol{\rho}_{\mathrm{m},h}, \cdot)$ as a score estimator. Preliminaries on DPPMs are detailed in Appendix H.

3. Algorithm and Results

We show that trajectories of the form given in Definition 2.1 can be efficiently imitated if (a) we are given a synthesis oracle that locally stabilizes chunks of the trajectory with primitive controllers and (b) the score of the following conditional distributions (whose existence is guaranteed by Appendix C) lies in a class Θ of bounded statistical complexity.

Formal Assumptions. We say trajectory $\boldsymbol{\rho}_\tau = (\mathbf{x}_{1:\tau+1}, \mathbf{u}_{1+\tau}) \in \mathscr{P}_\tau$ is ***feasible*** if it obeys the dynamics in (2.1). We assume that the transition map f takes the form of an Euler-like discretization

$$f(\mathbf{x}_t, \mathbf{u}_t) = \mathbf{x}_t + \eta f_\eta(\mathbf{x}_t, \mathbf{u}_t)$$

for a small step size $\eta > 0$ and say $\boldsymbol{\rho}_\tau$ is $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-*regular* if, for any $t \in [\tau]$ and $(\mathbf{x}'_t, \mathbf{u}'_t) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_u}$ such that $\|\mathbf{x}'_t - \mathbf{x}_t\| \vee \|\mathbf{u}_t - \mathbf{u}'_t\| \le R_{\mathrm{dyn}}$, it holds that $\|\nabla f_\eta(\mathbf{x}'_t, \mathbf{u}'_t)\|_{\mathrm{op}} \le L_{\mathrm{dyn}}$ and $\|\nabla^2 f_\eta(\mathbf{x}'_k, \mathbf{u}'_t)\|_{\mathrm{op}} \le M_{\mathrm{dyn}}$.[1] The ***Jacobian linearizations*** along a path $\boldsymbol{\rho}_\tau = (\mathbf{x}_{1:\tau+1}, \mathbf{u}_{1:\tau}) \in \mathscr{P}_\tau$ are matrices $\mathbf{A}_t(\boldsymbol{\rho}_\tau) := \frac{\partial}{\partial x} f(x_t, u_t)$ and $\mathbf{B}_t(\boldsymbol{\rho}_\tau) := \frac{\partial}{\partial u} f(x_t, u_t)$ for $t \in [\tau]$. Given $\boldsymbol{\rho}_\tau \in \mathscr{P}_\tau$ and primitive controllers $\kappa_{1:\tau}$, expressed as $\kappa_t(\mathbf{x}) = \bar{\mathbf{K}}_t(\mathbf{x} - \bar{\mathbf{x}}_t) + \bar{\mathbf{u}}_t(x)$, we say $(\boldsymbol{\rho}_\tau, \kappa_{1:\tau})$ are $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-***Jacobian stable*** if (a) $\kappa_{1:\tau}$ is consistent with $\boldsymbol{\rho}_\tau$ (b) $\max_{t\in[\tau]} \|\bar{\mathbf{K}}_t\| \vee \|\bar{\mathbf{x}}_t\| \vee \|\bar{\mathbf{u}}_t\| \le R_{\mathrm{stab}}$, and (c) the linearized closed-loop transition operator has exponential decay:

$$\|\boldsymbol{\Phi}_{\mathrm{cl},k,j}\|_{\mathrm{op}} \le B_{\mathrm{stab}}\left(1 - \tfrac{\eta}{L_{\mathrm{stab}}}\right)^{k-j}$$
$$\boldsymbol{\Phi}_{\mathrm{cl},k,j} := (\mathbf{I} + \eta\mathbf{A}_{\mathrm{cl},k-1}) \cdot (\mathbf{I} + \eta\mathbf{A}_{\mathrm{cl},k-2}) \cdots (\mathbf{I} + \eta\mathbf{A}_{\mathrm{cl},j}),$$

where above $\mathbf{A}_{\mathrm{cl},k} = \mathbf{A}_k(\boldsymbol{\rho}_\tau) + \mathbf{B}_{k-1}(\boldsymbol{\rho}_\tau)\mathbf{K}_{k-1}$. Our first two assumptions are as follows.

Assumption 3.1. The $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$ is feasible and $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular with probability 1.

Assumption 3.2. With probability 1 over $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$ and $\kappa_{1:T} = \mathtt{synth}(\boldsymbol{\rho}_T)$, the chunk-action pairs $(\boldsymbol{\rho}_{\mathrm{c},h+1}, \mathsf{a}_h)$ are $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian Stable for $1 \le h \le H$.

Assumption 3.1 enforces smoothness of the dynamics, but *not* smoothness or continuity of the underlying policy. Assumption 3.2 generalizes popular quantifications of stability (e.g. strong stability (Cohen et al., 2019)), and is satisfied when primitive controllers are synthesized via Ricatti equations of dynamics with stabilizable linearizations (Appendix G). Finally, we require access to a class of score functions rich enough to represent the deconvolution conditionals, defined as follows.

Definition 3.1 (Deconvolution Conditionals)**.** For $h \in [H]$, let $\pi^\star_{\mathrm{dec},h} \in \Delta(\mathcal{A}|\mathscr{P}_{\tau_{\mathrm{m}}-1})$ denote a conditional distribution of $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1} \mid \tilde{\boldsymbol{\rho}}_{\mathrm{m},h}$, where $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$, $\kappa_{1:T} = \mathtt{synth}(\boldsymbol{\rho}_T)$, and $\boldsymbol{\rho}_{\mathrm{m},h}$ is the memory chunk of $\boldsymbol{\rho}_T$ at step h, and $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} \sim \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2\mathbf{I})$ augments $\boldsymbol{\rho}_{\mathrm{m},h}$ with noise.

Assumption 3.3. For $h \in [H]$ let $\pi^\star_{\mathrm{dec},h,[t]} \in \Delta(\mathcal{A}|\mathscr{P}_{\tau_{\mathrm{m}}-1})$ denote $q_{[t]}$ as defined below (2.2) for $q = \pi^\star_{\mathrm{dec},h}$ the deconvolution policy defined above. For fixed $\sigma, \alpha > 0$ and $j \in \mathbb{N}$, let $\mathbf{s}_{\star,h,\sigma,[j]}$ denote the score function of $\pi^\star_{\mathrm{dec},h,[\alpha j]}$. We suppose that for any $J \in \mathbb{N}$ and $\alpha, \sigma > 0$, we are given a class of scores $\Theta = \Theta(\tau_{\mathrm{c}}, \tau_{\mathrm{m}}, \sigma) = \{\mathbf{s}_{\theta,1:H}\} = \bigcup_{j\in[J]} \Theta_j$ such that (a) for all $1 \le j \le J$, $\mathbf{s}_{\star,h,\sigma,[\alpha j]} \in \Theta_j$ and (b) a Rademacher-like complexity of Θ_j, $\mathcal{R}_n(\Theta_j)$ (defined in Appendix H) has polynomial decay $\mathcal{R}_n(\Theta_j) \le C_\Theta(1/\alpha)^\nu n^{-1/\nu}$ for some $\nu \ge 1$ and $C_\Theta = C_\Theta(\sigma, \tau_{\mathrm{c}}, \tau_{\mathrm{m}})$.

As justified in Appendix H, the above assumption is a natural for statistical learning, the decay condition on $\mathcal{R}_n(\Theta)$ holds for most common function classes (often with $\nu \le 2$ and even more benign dependence on J, α), and our results extend to approximate realizability. $\mathcal{R}_n(\Theta)$ depends implicitly on chunk and memory lengths $\tau_{\mathrm{c}}, \tau_{\mathrm{m}} > 0$ and problem dimension through the specification of $\mathbf{s}_{\star,h,\sigma,[\alpha j]}$. Realizability is motivated by the approximation power of deep neural networks (Bartlett et al., 2021).

Algorithm. Our proposed algorithm, TODA (Algorithm 1) combines DDPM-learning of chunked policies as in (Chi et al., 2023) with a popular form of data-augmentation (Ke et al., 2021). We collect N_{exp} expert trajectories, synthesize gains, and segment trajectories into memory chunks $\boldsymbol{\rho}_{\mathrm{m},h}$ and composite actions a_h as described in Section 2. We perturb each $\boldsymbol{\rho}_{\mathrm{m},h}$ to form N_{aug} chunks $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}$, as well as horizon indices $j \in [J]$ and inference noises $\boldsymbol{\gamma} \sim \mathcal{N}(\mathsf{a}_h, (\alpha j_h)^2\mathbf{I})$, and add these tuples $(\mathsf{a}_h, \tilde{\boldsymbol{\rho}}_{\mathrm{m},h}, j_h, \boldsymbol{\gamma}_h, h)$ to our data $\mathfrak{D}$. We end the training phase by minimizing the standard DDPM loss (Song & Ermon, 2019) $\mathcal{L}_{\mathrm{DDPM}}(\theta, \mathfrak{D})$:

$$\sum \left\| \boldsymbol{\gamma}_h - \mathbf{s}_{\theta,h}\left(e^{-\alpha j}\mathsf{a}_h + \sqrt{1 - e^{-2\alpha j}}\boldsymbol{\gamma}_h, \tilde{\boldsymbol{\rho}}_{\mathrm{m},h}, j_h\right) \right\|^2, \tag{3.1}$$

where the sum is over $(\mathsf{a}_h, \tilde{\boldsymbol{\rho}}_{\mathrm{c},h}, j_h, \boldsymbol{\gamma}_h, h) \in \mathfrak{D}$. Our algorithm differs subtly from past work in Line 8: we add augmentation noise *back in* at test time. Here, the notation $\mathtt{DDPM}(\mathbf{s}_{\theta,h}, \cdot) \circ \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2\mathbf{I})$ means, given $\boldsymbol{\rho}_{\mathrm{m},h}$, we perturb it to $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} \sim \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2\mathbf{I})$, and sample $\mathsf{a}_h \sim \mathtt{DDPM}(\mathbf{s}_{\theta,h}, \tilde{\boldsymbol{\rho}}_{\mathrm{m},h})$. The motivation for this is that adding noise at inference time *removes distribution shift* coming from training on augmented data; this simple observation is crucial for our theoretical guarantees.

Theoretical Guarantee. We now state our main theorem, which bounds the imitation losses of TODA trained on expert demonstrations. Let $d = \tau_{\mathrm{c}}(d_x + d_u + d_x d_u)$, and let $c_1, \ldots, c_5$ denote terms given in Appendix G that are polynomial in the parameters in Assumptions 3.1 and 3.2.

[1] Here, $\|\nabla^2 f_\eta(\mathbf{x}'_t, \mathbf{u}'_t)\|_{\mathrm{op}}$ denotes the operator-norm of a three-tensor.

Algorithm 1 Trajectory Optimization with Data Augmentation (TODA)

1: **Initialize** Synthesis oracle `synth`, sample sizes $N_{\mathrm{exp}}, N_{\mathrm{aug}} \in \mathbb{N}$, $\sigma \geq 0$, DDPM step size $\alpha > 0$, DDPM horizon J, function class $\{\mathbf{s}_\theta\}_{\theta \in \Theta}$, gain magnitude $R > 0$, empty data buffer $\mathfrak{D} \leftarrow \emptyset$.
  % For no augmentation, set $\sigma = 0$ **and** $N_{\mathrm{aug}} = 1$

2: **for** $n = 1, 2, \ldots N_{\mathrm{exp}}$ **do**

3: Sample $\boldsymbol{\rho}_T = (x_{1:T+1}, u_{1:T}) \sim \mathcal{D}_{\mathrm{exp}}$ and set $\kappa_{1:T} = \mathtt{synth}(\boldsymbol{\rho})$
  % Segment $\boldsymbol{\rho}_{\mathrm{m},1:H}$ **from** $\boldsymbol{\rho}_T$ **and** $\mathsf{a}_{1:H}$ **from** $\kappa_{1:T}$

4: **for** $i = 1, 2, \ldots, N_{\mathrm{aug}}$ and $h = 1, 2, \ldots, H$ **do**

5: Sample $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} \sim \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2 \mathbf{I})$, $j_h \sim \mathrm{Unif}([J])$ and $\boldsymbol{\gamma}_h \sim \mathcal{N}(\mathsf{a}_h, (j_h \alpha)^2 \mathbf{I})$.

6: $\mathfrak{D} \leftarrow \mathfrak{D}.\mathrm{append}\left(\{(\mathsf{a}_h, \tilde{\boldsymbol{\rho}}_{\mathrm{c},h}, j_h, \boldsymbol{\gamma}_h, h)\}\right)$

7: Fit $\theta \in \arg\min_{\theta \in \Theta} \mathcal{L}_{\mathrm{DDPM}}(\theta, \mathfrak{D})$

8: **return** $\hat{\pi}_\sigma = (\hat{\pi}_{1:H})$, where $\hat{\pi}_{h,\sigma}(\boldsymbol{\rho}_{\mathrm{m},h}) = \mathtt{DDPM}(\mathbf{s}_{\theta,h}, \cdot) \circ \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2 \mathbf{I})$.

Theorem 1. *Consider running* TODA *for* $\sigma > 0$ *with parameters* J, α *polynomial in the parameters given in Assumptions 3.1 and 3.2 specified in Appendix H. Suppose that Assumptions 3.1 to 3.3 hold and further suppose the chunk length satisfies* $\tau_{\mathrm{c}} \geq c_3/\eta$. *Given* $\sigma, \delta > 0$, *select any* $\varepsilon > 0$ *for which* $5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right) \leq c_4^2/(16\sigma^2)$. *If* $N_{\mathrm{exp}} \geq \mathrm{poly}\left(C_\Theta, \varepsilon/\sigma, R_{\mathrm{stab}}, d, \log(H/\delta)\right)^\nu$, *then for* $\hat{\pi}_\sigma$ *the policy output by* TODA, *it holds with probability* $1 - \delta$ *over the training data that both* $\mathcal{L}_{\mathrm{marg},\varepsilon_1}(\hat{\pi}_\sigma)$ *and* $\mathcal{L}_{\mathrm{fin},\varepsilon_2}(\hat{\pi}_\sigma)$ *are upper bounded by*

$$H\left(\frac{3\varepsilon}{\sigma} + 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}e^{-\frac{\eta(\tau_{\mathrm{c}} - \tau_{\mathrm{m}})}{L_{\mathrm{stab}}}}\right) \tag{3.2}$$

where $\varepsilon_1 = \varepsilon + 4c_5\sigma \cdot (5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right))^{1/2}$ *and* $\varepsilon_2 = \varepsilon + 4c_5 e^{-\eta\tau_{\mathrm{c}}/L_{\mathrm{stab}}}\sigma \cdot (5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right))^{1/2}$.

Theorem 1 guarantees imitation of the distribution of marginals and final states of $\mathcal{D}_{\mathrm{exp}}$. Each term in (3.2) can be made small by decreasing the amount of noise σ in the augmentation, increasing the number of trajectories, and increasing the chunk length τ_{c}. Increasing τ_{c} comes at the (implicit) expense of requiring a more expressive score class Θ (requiring greater N_{exp}); similarly, as expressed in Appendix H, the scores $\mathsf{s}_{\star,h,\sigma,[\alpha j]}$ may become harder to learn σ decreases. Note that the contribution of the additive σ-term in ε_2, used for the final-state loss $\mathcal{L}_{\mathrm{fin},\varepsilon}$, is exponentially-in-$\tau_{\mathrm{c}}$ smaller than that in ε_1. Interestingly, our theory suggest no benefit to increasing τ_{m} (corroborated empirically in (Chi et al., 2023)). Appendix I gives guarantees for imitating *joint* trajectories under the further assumptions that (a) the demonstrator has memory (or, more generally, a mixing time) of at most τ_{m}, and (b) *either* the demonstrator distribution happens to satisfy a certain continuity property, *or* $\sigma = 0$ and instead the learned $\hat{\pi}$ satisfies that same property.

Theorem 1 leverages statistical learning guarantees for DPPMs to show our learned policy approximately samples from $\pi^\star_{\mathrm{dec},h}$ in a truncated Wasserstein distance (Appendix H). These steps are combined with a general template for imitation learning developed in Section 4, with the final proof deferred to Appendix I . In Appendix F we show that this framework is essentially tight and thus suboptimality in Theorem 1 comes from looseness in conditional sampling guarantees. If we were above to approximately sample from $\pi^\star_{\mathrm{dec},h}$ in *total variation*, rather than a truncated Wasserstein distance, the imitation learning problem would be trivialized (Appendix I). Appendix H explains that the needed assumptions for this stronger sense of approximate sampling do not hold in our setting, because expert distributions over actions typically lie on low-dimensional manifolds.

Stability, limitations, and future work. We never explicitly model bifurcations; rather, we allow expert demonstrations to be sufficiently rich as to permit them. Eschewing global stability, τ_{c} ensures that trajectories are long enough for the *strictly local* stability assumptions in Assumption 3.2 to provide benefit. Thus, non-Markovianity is challenging only insofar as it relates to the difficulty of local stabilization. A key limitation of our work is that, to take advantage of local stability, we rely on either synthesized primitive controllers (in our analysis) or low-level stabilizing controllers built into problem environments (in our experiments). Developing a more comprehensive approach to stability (perhaps one that does not require explicit gain synthesis, and extends to non-smooth systems) is an exciting direction for future work. Appendix B compares our hierarchical approach to stability to more standard notions, which we show rule out the possibility for bifurcated demonstrations.

4. Analysis

Our analysis abstracts away the vector-valued dynamics into a deterministic MDP with ***composite-states*** $\mathsf{s} \in \mathcal{S}$ and ***composite-actions*** $\mathsf{a} \in \mathcal{A}$, with dynamics

$$\mathsf{s}_{h+1} = F_h(\mathsf{s}_h, \mathsf{a}_h), \quad h \in \{1, 2, \ldots, H\} \tag{4.1}$$

A ***composite-policy*** π is a sequence of kernels $\pi_1, \pi_2, \ldots, \pi_H : \mathcal{S} \to \Delta(\mathcal{A})$. We let $\mathsf{P}_{\mathrm{init}}$ denote the distribution of initial state s_1, and D_π denote the distribution of $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ subject to $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\mathsf{a}_h \mid \mathsf{s}_{1:h}, \mathsf{a}_{1:h-1} \sim \pi_h(\mathsf{s}_h)$, and the composite-dynamics (4.1). We assume that we have an optimal policy $\pi^\star$ to be imitated, and define $\mathsf{P}^\star_h$ as the marginal distribution of s_h under $\mathsf{D}_{\pi^\star}$.

Structure of the proof. We begin by explaining key objects, stability and continuity properties required in the composite MDP. Then, Section 4.1 relates the composite MDP to our original setting by taking composite-states $\mathsf{s}_h = \boldsymbol{\rho}_{\mathrm{c},h}$ as chunks, and taking composite actions as sequences of primitive controllers $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$ as in Section 2. We also explain why relevant stability and continuity conditions are met. Finally, we derive Theorem 1 from a generic guarantee for smoothed imitiation learning in the composite MDP, Theorem 2, and from sampling guarantees in Appendix H.

We consider two pseudometrics on the space $\mathcal{S}$: $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\textsc{tvc}} : \mathcal{S}^2 \to \mathbb{R}_{\geq 0}$, and a function $\mathsf{d}_{\mathcal{A}} : \mathcal{A}^2 \to \mathbb{R}_{\geq 0}$. For convenience, *do not require* $\mathsf{d}_{\mathcal{A}}$ *to satisfy the axioms of a pseudometric.* We use $\mathsf{d}_{\mathcal{S}}$ and $\mathsf{d}_{\mathcal{A}}$ to measure error between states and actions, respectively, and $\mathsf{d}_{\textsc{tvc}}(\cdot,\cdot)$ for a probabilistic continuity property described below. In terms of $\mathsf{d}_{\mathcal{S}}$ and $\mathsf{d}_{\mathcal{A}}$, we consider three measures of imitation error: error on the (i) joint distribution of trajectories ($\Gamma_{\mathrm{joint},\varepsilon}$) (ii) marginal distribution of trajectories ($\Gamma_{\mathrm{marg},\varepsilon}$) and (iii) one-step error in actions ($\mathsf{d}_{\mathrm{os},\varepsilon}$). Formally:

Definition 4.1 (Imitation Errors)**.** Given an error parameter $\varepsilon > 0$, define the ***joint-error*** $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star) := \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\max_{h\in[H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}^\star_h, \hat{\mathsf{a}}_h)\} > \varepsilon\right]$, where the first infimum is over trajectory couplings $((\hat{\mathsf{s}}_{1:H+1}, \hat{\mathsf{a}}_{1:H}), (\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H})) \sim \mu_1 \in \mathscr{C}(\mathsf{D}_{\hat{\pi}}, \mathsf{D}_{\pi^\star})$ satisfying $\mathbb{P}_{\mu_1}[\hat{\mathsf{s}}_1 = \mathsf{s}^\star_1] = 1$. Define the ***marginal error*** $\Gamma_{\mathrm{marg},\varepsilon}(\hat{\pi} \parallel \pi^\star) := \max_{h\in[H]}\{\inf_{\mu_1} \mathbb{P}_{\mu_1}[\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1}, \hat{\mathsf{s}}_{h+1}) > \varepsilon], \inf_{\mu_1} \mathbb{P}_{\mu_1}[\mathsf{d}_{\mathcal{A}}(\mathsf{a}^\star_h, \hat{\mathsf{a}}_h) > \varepsilon]\}$ to be the same as the to joint-gap, with the "max" outside the probability and inf over couplings. Lastly, define the ***one-step error*** $\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}) \parallel \pi^\star_h(\mathsf{s})) := \inf_{\mu_2} \mathbb{P}_{\mu_2}[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) \leq \varepsilon]$, where the infimum is over $(\mathsf{a}^\star_h, \hat{\mathsf{a}}_h) \sim \mu_2 \in \mathscr{C}(\hat{\boldsymbol{\pi}}_h(\mathsf{s}), \boldsymbol{\pi}^\star_h(\mathsf{s}))$.

Stability. Our hierarchical approach offloads stability of stochastic $\pi^\star$ onto that of its composite-actions a_h, instantiated as *primitive controllers* (not raw inputs!). This allows us to circumvent more challenging incremental senses of stability (see Appendix B for further discussion).

Definition 4.2 (Input-Stability)**.** A trajectory $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ is ***input-stable*** if all sequences $\mathsf{s}'_1 = \mathsf{s}_1$ and $\mathsf{s}'_{h+1} = F_h(\mathsf{s}'_h, \mathsf{a}'_h)$ satisfy $\mathsf{d}_{\mathcal{S}}(\mathsf{s}'_{h+1}, \mathsf{s}_{h+1}) \vee \mathsf{d}_{\textsc{tvc}}(\mathsf{s}'_{h+1}, \mathsf{s}_{h+1}) \leq \max_{1\leq j\leq h} \mathsf{d}_{\mathcal{A}}\left(\mathsf{a}'_j, \mathsf{a}_j\right)$, $\forall h \in [H]$. A policy π is ***input-stable*** if $(\mathsf{s}_{1:H}, \mathsf{a}_{1:H}) \sim \mathsf{D}_\pi$ is ***input-stable*** almost surely.

TVC. Continuity of probability kernels and policies in TV distance are measured in terms of $\mathsf{d}_{\textsc{tvc}}$.

Definition 4.3. For a measure-space $\mathcal{X}$ and non-decreasing $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, we call a probability kernel $\mathsf{W} : \mathcal{S} \to \Delta(\mathcal{X})$ *γ-**total variation continuous** (γ-**TVC**)* if, for all $\mathsf{s}, \mathsf{s}' \in \mathcal{S}$, $\mathrm{TV}(\mathsf{W}(\mathsf{s}), \mathsf{W}(\mathsf{s}')) \leq \gamma(\mathsf{d}_{\textsc{tvc}}(\mathsf{s}, \mathsf{s}'))$. A policy π is γ-***TVC*** if $\pi_h : \mathcal{S} \to \Delta(\mathcal{A})$ is γ-TVC $\forall h \in [H]$.

Smoothing. In Appendix D, we show that under the strong condition that the learned policy $\hat{\pi}$ is γ-TVC, then TODA with no data augmentation ($\sigma = 0$) learns the distribution. Frequently, however, $\hat{\pi}$ may not satisfy this condition, such as when the ground truth $\pi^\star$ is not also TVC. We circumvent this by introducing a *smoothing kernel* $\mathsf{W}_\sigma : \mathcal{S} \to \Delta(\mathcal{S})$ that corresponds to the data augmentation; in TODA we let the kernel be a Gaussian, sending $\boldsymbol{\rho}_{\mathrm{m},h}$ to $\mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2\mathbf{I}) \in \Delta(\mathscr{P}_{\boldsymbol{\rho}_{\mathrm{m},h}})$. We will thus be able to replace TVC of $\hat{\pi}$ with TVC of W_σ. We now introduce a few key objects.

Definition 4.4. Given a policy π, we define its ***smoothed policy*** $\pi \circ \mathsf{W}_\sigma$ via components $(\pi \circ \mathsf{W}_\sigma)_h = \pi_h \circ \mathsf{W}_\sigma : \mathcal{S} \to \Delta(\mathcal{A})$. For $\pi^\star$ fixed, define the ***augmented distibution*** $\mathsf{P}^\star_{\mathrm{aug},h}$ as the joint distribution over $(\mathsf{s}^\star_h \sim \mathsf{P}^\star_h, \mathsf{a}^\star_h \sim \pi^\star_h(\mathsf{s}^\star_h), \tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h))$, with $\mathsf{a}^\star_h \perp \tilde{\mathsf{s}}^\star_h \mid \mathsf{s}^\star_h$. The ***deconvolution policy*** $\pi^\star_{\mathrm{dec}}$ is defined by letting $\pi^\star_{\mathrm{dec},h}(\mathsf{s})$ denote the distribution of $\mathsf{a}^\star_h \mid \tilde{\mathsf{s}}^\star_h = \mathsf{s}_h$, where $\mathsf{a}^\star_h, \tilde{\mathsf{s}}^\star_h$ are drawn from $\mathsf{P}^\star_{\mathrm{aug},h}$. Finally, the ***replica policy*** is $\pi^\star_{\circlearrowright} = \pi^\star_{\mathrm{dec}} \circ \mathsf{W}_\sigma$.

The operator $\pi \circ \mathsf{W}_\sigma$ composes π with the smoothing kernel. The deconvolution policy $\pi^\star_{\mathrm{dec}}$ captures the distribution of actions under $\pi^\star$ given an augmented state, and corresponds to $\pi^\star_{\mathrm{dec}} = (\pi^\star_{\mathrm{dec},h})_{h=1}^H$. We argue that if a policy $\hat{\pi}$ approximates $\pi^\star_{\mathrm{dec}}$ at each step, then $\hat{\pi} \circ \mathsf{W}_\sigma$ imitates $\pi^\star_{\circlearrowright\sigma} = \pi^\star_{\mathrm{dec}} \circ \mathsf{W}_\sigma$. We explain the "replica policy", and importance of imitating it, after we state our main theorem. First, we define a notion of stability to smoothing, taking $\mathsf{d}_{\textsc{tvc}}, \mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathcal{A}}$ as given.

Definition 4.5. For a non-decreasing maps $\gamma_{\textsc{ips},1}, \gamma_{\textsc{ips},2} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ a pseudometric $\mathsf{d}_{\textsc{ips}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ (possibly other than $\mathsf{d}_{\mathcal{S}}$ or $\mathsf{d}_{\textsc{tvc}}$), and $r_{\textsc{ips}} > 0$, we say a policy π is $(\gamma_{\textsc{ips},1}, \gamma_{\textsc{ips},2}, \mathsf{d}_{\textsc{ips}}, r_{\textsc{ips}})$*-input-&-process stable (IPS)* if the following holds for any $r \in [0, r_{\textsc{ips}}]$. Consider any sequence of kernels $\mathsf{W}_1, \ldots, \mathsf{W}_H : \mathcal{S} \to \Delta(\mathcal{S})$ satisfying $\max_{h,\mathsf{s}\in\mathcal{S}} \mathbb{P}_{\tilde{\mathsf{s}}\sim\mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\textsc{ips}}(\tilde{\mathsf{s}}, \mathsf{s}) \leq r] = 1$, and define a process $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h), \mathsf{a}_h \sim \pi_h(\tilde{\mathsf{s}}_h)$, and $\mathsf{s}_{h+1} := F_h(\mathsf{s}_h, \mathsf{a}_h)$. Then, almost surely, (a) the sequence $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ is input-stable w.r.t $(\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathcal{A}})$ (b) $\max_{h\in[H]} \mathsf{d}_{\textsc{tvc}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \leq \gamma_{\textsc{ips},1}(r)$ and (c) $\max_{h\in[H]} \mathsf{d}_{\mathcal{S}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \leq \gamma_{\textsc{ips},2}(r)$.

Condition (a) means that the policy $\tilde{\pi}$ defined by $\tilde{\pi}_h = \pi_h \circ \mathsf{W}_h$ is input-stable. In the appendix, we instantiate $\mathsf{W}_{1:H}$ not as W_σ, but as (a truncation of) *replica kernels* $\mathsf{W}^\star_{\circlearrowright,h}$ for which $\pi^\star_{\circlearrowright\sigma,h} = \pi^\star_h \circ \mathsf{W}^\star_{\circlearrowright,h}$. We show that the replica kernel inherits any concentration satisfied by W_σ, ensuring (via truncation) that $\mathbb{P}_{\tilde{\mathsf{s}}\sim\mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\textsc{ips}}(\tilde{\mathsf{s}}, \mathsf{s})] \leq r$. Conditions (b & c) merely require that one-step dynamics are robust to small changes in state, measured in terms of both $\mathsf{d}_{\textsc{tvc}}$ and $\mathsf{d}_{\mathcal{S}}$.

4.1. Instantiation for control

Here we explain the mapping from the control setting of interest to the composite MDP; in so doing we distinguish between the case $h > 1$ and $h = 1$ with reference to composite-states. In the former case, $\mathsf{s}_h = (\mathbf{x}_{t_{h-1}:t_h}, \mathbf{u}_{t_{h-1}:t_h-1}) \in \mathscr{P}_{\tau_{\mathrm{m}}}$, and $\mathsf{a}_h =$

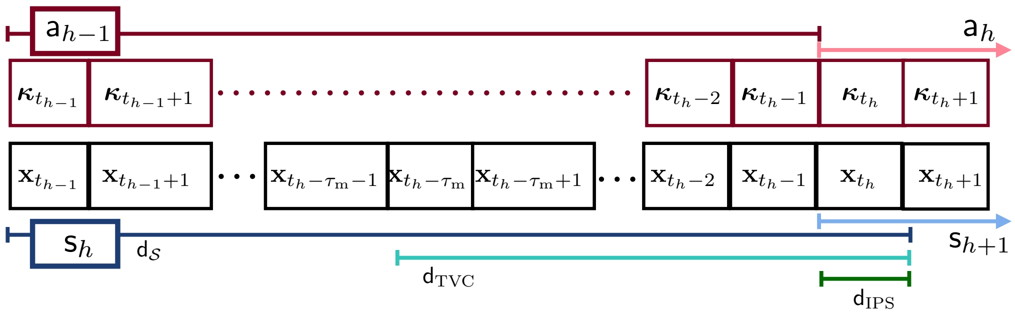


Figure 1. Schematic depicting the composite MDP. States $\mathbf{x}$ and stabilizing gains κ are chunked into composite states s and composite actions a (control inputs $\mathbf{u}$ not depicted). The distance labels correspond to the domain over which each distance is evaluated. Note that a_h begins at the same time that s_{h+1} does, an indexing convention that we adopt to make the notation in the proofs simpler.

$\kappa_{t_h:t_{h+1}-1}$ (as in Section 2). Importantly, a_h ***are primitive controllers, which allows us to meet the strong stability condition in Definition 4.2***. Figure 1 provides a visual aid for the subtle indexing. For $\mathsf{s}_h, \mathsf{s}'_h$, we define $\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_{t\in[t_{h-1}:t_h]} \|\mathbf{x}_t - \mathbf{x}'_t\| \vee \max_{t\in[t_{h-1}:t_h-1]} \|\mathbf{u}_t - \mathbf{u}'_t\|$, which measures distance on the full subtrajectory, $\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_{t\in[t_h-\tau_{\mathrm{m}}:t_h]} \|\mathbf{x}_t - \mathbf{x}'_t\| \vee \max_{t\in[t_h-\tau_{\mathrm{m}}:t_h-1]} \|\mathbf{u}_t - \mathbf{u}'_t\|$, which measures distance on the last τ_{m} steps, and $\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) = \|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\|$, which is only on the last step. In the latter case, when $h = 1$, we let $\mathsf{s}_1 = \mathbf{x}_1 \in \mathcal{X}$, and we let $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ all denote the Euclidean distance on $\mathcal{X}$. In all cases, the transition dynamics F_h are induced by the dynamics (2.1) with $\mathbf{u}_t = \kappa_t(\mathbf{x}_t)$. Finally, for $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{x}}_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{K}}_{1:\tau_{\mathrm{c}}})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{x}}'_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{K}}'_{1:\tau_{\mathrm{c}}})$, we choose a $\mathsf{d}_{\mathcal{A}}$ that takes value ∞ when primitive controllers are too far apart as $\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}')$ defined to be

$$c_1 \max_{k\in[\tau_{\mathrm{c}}]} (\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|) \quad (4.2)$$
$$+ \mathbf{I}_{0,\infty}\{\mathcal{E}\},$$

where we define $\mathcal{E} := \{\max_{1\le k\le\tau_{\mathrm{c}}} \max\{\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\|, \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\|, \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|\} \le c_2\}$, $\mathbf{I}_{0,\infty}$ is the indicator taking infinite value when the event fails to hold, and c_1 and c_2 are constants depending polynomially on all problem parameters, given in Appendix G.

We let the expert policy $\pi^\star$ be the concatenation of policies $\pi^\star_h$, each of which is defined to be the distribution of a_h conditioned on $\boldsymbol{\rho}_{\mathrm{m},h}$ under $\mathcal{D}_{\mathrm{exp}}$ (see Appendix I for a rigorous definition). As noted above, we take the smoothing kernel W_σ to map $\boldsymbol{\rho}_{\mathrm{m},h}$ to a $\mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2\mathbf{I}) \in \Delta(\mathscr{P}_{\boldsymbol{\rho}_{\mathrm{m},h}})$, which that same appendix shows is $\frac{1}{2\sigma}$-TVC (w.r.t. $\mathsf{d}_{\mathrm{TVC}}$ defined above). We note that under these substitutions, the deconvolution policy $\pi^\star_{\mathrm{dec}} = (\pi^\star_{\mathrm{dec},h})_{h=1}^H$ is ***precisely as defined in Definition 3.1***.

Appendix G shows that Assumptions 3.1 and 3.2 imply that $\pi^\star$ enjoys the IPS property in the composite MDP thus instantiated, along with many more granular stability guarantees for time-varying affine feedback in nonlinear control systems, which may be of independent interest.

Proposition 4.1. *Let $c_3, c_4, c_5 > 0$ be as given in Appendix G (and polynomial in relevant quantities). Suppose $\tau_{\mathrm{c}} \ge c_3/\eta$, and let $r_{\mathrm{IPS}} = c_4$, $\gamma_{\mathrm{IPS},1}(u) = c_5 u \exp(-\eta(\tau_{\mathrm{c}} - \tau_{\mathrm{m}})/L_{\mathrm{stab}})$, $\gamma_{\mathrm{IPS},2}(u) = c_5 u$. Then, for $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ as above, we have that $\pi^\star$ is $(\gamma_{\mathrm{IPS},1}, \gamma_{\mathrm{IPS},2}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-IPS.*

4.2. A Guarantee in the Composite MDP Stability, and the derivation of Theorem 1

With the substitutions in Section 4.1, it suffices to prove an imitation guarantee in the composite MDP, assuming $\pi^\star$ is IPS, and $\hat{\pi}$ is close to $\pi^\star_{\mathrm{dec}}$ in the appropriate sense.

Theorem 2. *Suppose $\pi^\star$ is $(\gamma_{\mathrm{IPS},1}, \gamma_{\mathrm{IPS},2}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-IPS and W_σ is γ_σ-TVC. Let $\varepsilon > 0$, $r \in (0, \frac{1}{2} r_{\mathrm{IPS}}]$; define $p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}', \mathsf{s}) > r]$ and $\varepsilon' := \varepsilon + \gamma_{\mathrm{IPS},2}(2r)$. Then, for any policy $\hat{\pi}$, both $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\circlearrowleft})$ and $\Gamma_{\mathrm{marg},\varepsilon'}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ are upper bounded by*

$$H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\mathrm{IPS},1}(2r)\})\right)$$
$$+ \textstyle\sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h\sim\mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h\sim\mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h)).$$

Deriving Theorem 1 from Theorem 2. A full proof is given in Appendix I, using the subtley that $\pi^\star$ as described above yields trajectories with the same marginals (but possibly different joint distributions) as $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$; thus, to bound losses in Definition 2.2, it suffices to bound the imitation gaps in Definition 4.1 w.r.t. $\pi^\star$. Using the analysis in Appendix H, we show that our DDPM training precisely ensures that $\hat{\pi}_\sigma = \hat{\pi} \circ \mathsf{W}_\sigma$ in TODA minimizes (an upper bound on) the term $\sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h\sim\mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h\sim\mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h))$. Finally, we combine the guarantees of Proposition 4.1, the aforementioned TVC-bound on W_σ, and Gaus-

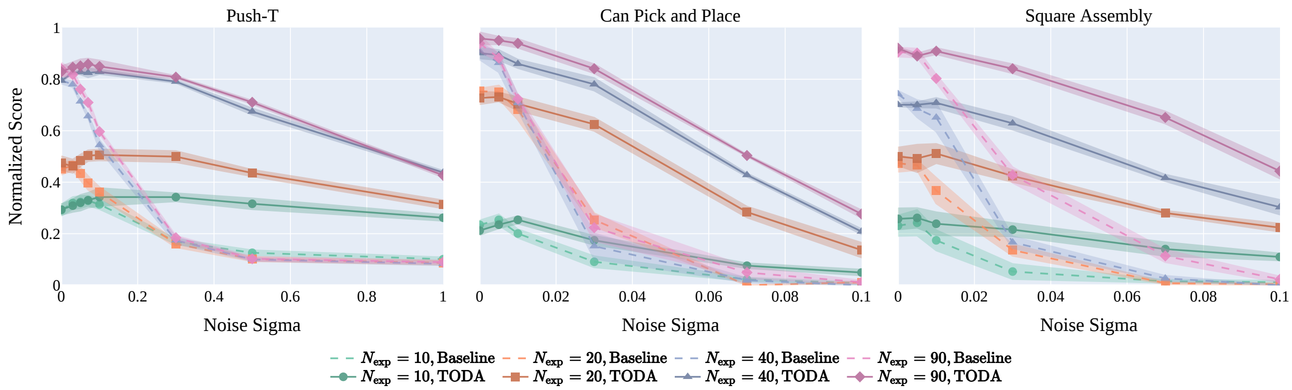


Figure 2. Performance of baseline $\hat{\pi}$ and noise-injected $\hat{\pi} \circ \mathsf{W}_\sigma$ TODA policy for different σ. We use 4 training seeds with 50 and 22 test trajectories per seed for PushT and Can and Square Environments respectively. Mean and standard deviation of the test performance on the 3 best checkpoints across the 4 seeds are plotted. The σ values correspond to noise in the normalized $[-1, 1]$ range.

sian concentration to bound p_r with the bound in Theorem 2 to conclude. $\square$

Proof Sketch of Theorem 2. The proof draws inspiration from the notion of replica symmetry in statistical physics (hence, the name replica) (Mezard & Montanari, 2009). We construct a coupling between a trajectory over $(\mathsf{s}_h^{\circlearrowleft}, \mathsf{a}_h^{\circlearrowleft})$ sampled using the replica policy $\pi^\star_{\circlearrowleft}$, and a trajectory $(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h)$ sampled from $\hat{\pi}_\sigma$. We introduce *teleporting* trajectories $\mathsf{s}_{h+1}^{\circlearrowleft} = F_h(\mathsf{s}_h^{\circlearrowleft}, \mathsf{a}_h^{\circlearrowleft})$, and $\mathsf{s}_{h+1}^{\text{tel}} = F_h(\tilde{\mathsf{s}}_h^{\text{tel}}, \mathsf{a}_h^{\text{tel}})$, where $\tilde{\mathsf{s}}_h^{\text{tel}}$ is sampled from the replica distribution of $\mathsf{s}_h^{\text{tel}}$ and $\mathsf{a}_h^{\text{tel}} \sim \pi_h^\star(\tilde{\mathsf{s}}_h^{\text{tel}})$; in words, $\mathsf{s}_h^{\text{tel}}$ *teleports* to an independent and identically distributed copy conditional on the noise agumentation, and draws an action from the replica policy evaluated on the new state.

The key fact of the replica distribution is that it preserves marginals, meaning that all $\mathsf{s}_h^{\text{tel}}$ and $\tilde{\mathsf{s}}_h^{\text{tel}}$ both have marginals according to $\mathsf{P}_h^\star$. We show that $\mathsf{s}_h^{\circlearrowleft}$ tracks the teleporting trajectories, up to the IPS terms $\gamma_{\text{IPS},i}$ and concentration of the kernel, due to total variation continuity of W_σ. Because the marginals of $\mathsf{s}_h^{\text{tel}}$ are distributed according to $\mathsf{P}_h^\star$, we can argue that a (fictitious) action $\hat{\mathsf{a}}_h^{\text{tel,inter}} \sim \hat{\pi}_\sigma(\mathsf{s}_h^{\text{tel}})$ is close to $\mathsf{a}_h^{\text{tel}}$ (by the data processing inequality, it is bounded by the closeness of $\hat{\pi}_h$ and $\pi^\star_{\text{dec},h}$ on $\tilde{\mathsf{s}}_h^{\text{tel}} \sim \mathsf{W}_\sigma(\mathsf{s}_h^{\text{tel}})$, $\mathsf{s}_h^{\text{tel}} \sim \mathsf{P}_h^\star$). We then use total variation continuity to relate to another fictious action $\hat{\mathsf{a}}_h^{\circlearrowleft,\text{inter}}$ to $\mathsf{a}_h^{\circlearrowleft}$. Finally, we use input-stability and TVC again, to relate $\hat{\mathsf{a}}_h^{\circlearrowleft,\text{inter}}$ to actions $\hat{\mathsf{a}}_h \sim \hat{\pi}_\sigma(\hat{\mathsf{s}}_h)$. Our couplings are summarized in the following diagram:

$$\underbrace{(\mathsf{a}^{\circlearrowleft} \leftrightarrow \mathsf{a}^{\text{tel}})}_{\gamma_{\text{TVC}}\text{ and induction}} \rightarrow \underbrace{(\mathsf{a}^{\text{tel}} \leftrightarrow \hat{\mathsf{a}}^{\text{tel,inter}})}_{\text{learning and sampling}}$$
$$\rightarrow \underbrace{(\hat{\mathsf{a}}^{\text{tel,inter}} \leftrightarrow \hat{\mathsf{a}}^{\circlearrowleft,\text{inter}})}_{\gamma_{\text{TVC}}\text{ and induction}} \rightarrow \underbrace{(\hat{\mathsf{a}}^{\circlearrowleft,\text{inter}} \leftrightarrow \hat{\mathsf{a}})}_{\gamma_{\text{TVC}}\text{ and induction}}.$$

We construct conditional couplings between pairs of the aforementioned trajectories, each of which corresponds to a certain optimal transport cost. That past trajectories can be associated to optimal couplings measurably is non-trivial, and proven in Proposition C.3. To conclude, we apply a measure theoretic result (Lemma C.2) to "glue" the pairwise couplings together and establish the main result. The full proof is given in Appendix E, relying on measure-theoretic details in Appendix C. $\square$

5. Simulation Study of Test-Time Noise-Injection

We empirically evaluate the effect on policy performance of our proposal to inject noise back into the dynamics at inference time. We consider three challenging robotic manipulation tasks studied in prior work: PushT block-pushing (Chi et al., 2023); Robomimic Can Pick-and-Place and Square Nut Assembly (Mandlekar et al., 2021). We explain the environments in greater detail, along with all training and computational details in Appendix K. The learned diffusion policy generates state trajectories over a $\tau_c = 8$ chunking horizon using fixed feedback gains provided by the `synth` oracle to perform position-tracking of the DDPM model output. We direct the reader to Chi et al. (2023) for an extensive empirical investigation into the performance of diffusion policies in the noiseless $\sigma = 0$ setting. We display the results of our experiments in Figure 2. Observe that the performance degredation of the replica policy from the unsmoothed $\sigma = 0$ variant is minimal across all environments and even leads to a slight but noticeable improvement in the small-noise regime for PushT (and low-data Can Pick and Place). In the presence of non-negligible noise TODA significantly outperforms the conventional policy $\hat{\pi}$ (obtained by adding augmentation at training but not test time), as predicted by our theory.

References

Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., and Vijayanarasimhan, S. Youtube-8m: A large-scale video classification benchmark. *arXiv preprint arXiv:1609.08675*, 2016.

Ajay, A., Gupta, A., Ghosh, D., Levine, S., and Agrawal, P. Distributionally adaptive meta reinforcement learning. *arXiv preprint arXiv:2210.03104*, 2022.

Altschuler, J. and Talwar, K. Privacy of noisy stochastic gradient descent: More iterations without more privacy loss. *Advances in Neural Information Processing Systems*, 35:3788–3800, 2022.

Altschuler, J. M. and Chewi, S. Faster high-accuracy log-concave sampling via algorithmic warm starts. *arXiv preprint arXiv:2302.10249*, 2023.

Anderson, B. D. and Moore, J. B. *Optimal control: linear quadratic methods*. Courier Corporation, 2007.

Angel, O. and Spinka, Y. Pairwise optimal coupling of multiple random variables. *arXiv preprint arXiv:1903.00632*, 2019.

Bansal, M., Krizhevsky, A., and Ogale, A. Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst. *arXiv preprint arXiv:1812.03079*, 2018.

Bartlett, P. L., Montanari, A., and Rakhlin, A. Deep learning: a statistical viewpoint. *Acta numerica*, 30:87–201, 2021.

Block, A., Mroueh, Y., and Rakhlin, A. Generative modeling with denoising auto-encoders and langevin sampling. *arXiv preprint arXiv:2002.00107*, 2020a.

Block, A., Mroueh, Y., Rakhlin, A., and Ross, J. Fast mixing of multi-scale langevin dynamics under the manifold hypothesis. *arXiv preprint arXiv:2006.11166*, 2020b.

Bojarski, M., Del Testa, D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

Chang, J. D., Uehara, M., Sreenivas, D., Kidambi, R., and Sun, W. Mitigating covariate shift in imitation learning via offline data without great coverage. *arXiv preprint arXiv:2106.03207*, 2021.

Chen, S., Chewi, S., Li, J., Li, Y., Salim, A., and Zhang, A. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *NeurIPS 2022 Workshop on Score-Based Methods*, 2022.

Chi, C., Feng, S., Du, Y., Xu, Z., Cousineau, E., Burchfiel, B., and Song, S. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137*, 2023.

Cohen, A., Koren, T., and Mansour, Y. Learning linear-quadratic regulators efficiently with only $\sqrt{T}$ regret. In *International Conference on Machine Learning*, pp. 1300–1309. PMLR, 2019.

Dasari, S., Ebert, F., Tian, S., Nair, S., Bucher, B., Schmeckpeper, K., Singh, S., Levine, S., and Finn, C. Robonet: Large-scale multi-robot learning. *arXiv preprint arXiv:1910.11215*, 2019.

De Haan, P., Jayaraman, D., and Levine, S. Causal confusion in imitation learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Durrett, R. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

Finn, C., Yu, T., Zhang, T., Abbeel, P., and Levine, S. One-shot visual imitation learning via meta-learning. In *Conference on robot learning*, pp. 357–368. PMLR, 2017.

Hagood, J. W. and Thomson, B. S. Recovering a function from a dini derivative. *The American Mathematical Monthly*, 113(1):34–46, 2006.

Havens, A. and Hu, B. On imitation learning of linear control policies: Enforcing stability and robustness constraints via lmi conditions. In *2021 American Control Conference (ACC)*, pp. 882–887. IEEE, 2021.

Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., Desai, R., Zhu, T., Parajuli, S., Guo, M., et al. Jacob steinhardt et justin gilmer. the many faces of robustness: A critical analysis of out-of-distribution generalization. *arXiv preprint arXiv:2006.16241*, 2020.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

Hussein, A., Gaber, M. M., Elyan, E., and Jayne, C. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)*, 50(2):1–35, 2017.

Hussein, A., Elyan, E., Gaber, M. M., and Jayne, C. Deep imitation learning for 3d navigation tasks. *Neural computing and applications*, 29:389–404, 2018.

Hyvärinen, A. and Dayan, P. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005.

Janner, M., Du, Y., Tenenbaum, J. B., and Levine, S. Planning with diffusion for flexible behavior synthesis. *arXiv preprint arXiv:2205.09991*, 2022.

Jin, C., Netrapalli, P., Ge, R., Kakade, S. M., and Jordan, M. I. A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736*, 2019.

Ke, L., Wang, J., Bhattacharjee, T., Boots, B., and Srinivasa, S. Grasping with chopsticks: Combating covariate shift in model-free imitation learning for fine manipulation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6185–6191. IEEE, 2021.

Kelly, M., Sidrane, C., Driggs-Campbell, K., and Kochenderfer, M. J. Hg-dagger: Interactive imitation learning with human experts. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8077–8083. IEEE, 2019.

Kostrikov, I., Yarats, D., and Fergus, R. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. *arXiv preprint arXiv:2004.13649*, 2020.

Laskey, M., Lee, J., Fox, R., Dragan, A., and Goldberg, K. Dart: Noise injection for robust imitation learning. In *Conference on robot learning*, pp. 143–156. PMLR, 2017.

Lee, H., Lu, J., and Tan, Y. Convergence of score-based generative modeling for general data distributions. In *International Conference on Algorithmic Learning Theory*, pp. 946–985. PMLR, 2023.

Mandlekar, A., Xu, D., Wong, J., Nasiriany, S., Wang, C., Kulkarni, R., Fei-Fei, L., Savarese, S., Zhu, Y., and Martín-Martín, R. What matters in learning from offline human demonstrations for robot manipulation. *arXiv preprint arXiv:2108.03298*, 2021.

Maurer, A. A vector-contraction inequality for rademacher complexities. In *Algorithmic Learning Theory: 27th International Conference, ALT 2016, Bari, Italy, October 19-21, 2016, Proceedings 27*, pp. 3–17. Springer, 2016.

Mezard, M. and Montanari, A. *Information, physics, and computation*. Oxford University Press, 2009.

Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681*, 2019.

Nichol, A. Q. and Dhariwal, P. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pp. 8162–8171. PMLR, 2021.

Pearce, T., Rashid, T., Kanervisto, A., Bignell, D., Sun, M., Georgescu, R., Macua, S. V., Tan, S. Z., Momennejad, I., Hofmann, K., et al. Imitating human behaviour with diffusion models. *arXiv preprint arXiv:2301.10677*, 2023.

Perez, E., Strub, F., De Vries, H., Dumoulin, V., and Courville, A. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Pfrommer, D., Zhang, T., Tu, S., and Matni, N. Tasil: Taylor series imitation learning. *Advances in Neural Information Processing Systems*, 35:20162–20174, 2022.

Pfrommer, D., Simchowitz, M., Westenbroek, T., Matni, N., and Tu, S. The power of learned locally linear models for nonlinear policy optimization. *arXiv preprint arXiv:2305.09619*, 2023.

Polyanskiy, Y. and Wu, Y. *Information Theory: From Coding to Learning*. Cambridge University Press, 2022+.

Raginsky, M., Rakhlin, A., and Telgarsky, M. Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. In *Conference on Learning Theory*, pp. 1674–1703. PMLR, 2017.

Rakhlin, A., Sridharan, K., and Tsybakov, A. B. Empirical entropy, minimax regret and minimax risk. *Bernoulli Society for Mathematical Statistics and Probability*, 2017.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.

Ronneberger, O., Fischer, P., and Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer, 2015.

Ross, S. and Bagnell, D. Efficient reductions for imitation learning. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 661–668. JMLR Workshop and Conference Proceedings, 2010.

Ross, S., Gordon, G., and Bagnell, D. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 627–635. JMLR Workshop and Conference Proceedings, 2011.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pp. 2256–2265. PMLR, 2015.

Song, J., Meng, C., and Ermon, S. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.

Song, Y. and Ermon, S. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

Song, Y., Garg, S., Shi, J., and Ermon, S. Sliced score matching: A scalable approach to density and score estimation. In *Uncertainty in Artificial Intelligence*, pp. 574–584. PMLR, 2020b.

Sun, X., Yang, S., and Mangharam, R. Mega-dagger: Imitation learning with multiple imperfect experts. *arXiv preprint arXiv:2303.00638*, 2023.

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., et al. Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*, 2022.

Tu, S., Robey, A., Zhang, T., and Matni, N. On the sample complexity of stability constrained imitation learning. In *Learning for Dynamics and Control Conference*, pp. 180–191. PMLR, 2022.

Van Handel, R. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Villani, C. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Villani, C. et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

Vincent, P. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Wainwright, M. J. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

Westenbroek, T., Simchowitz, M., Jordan, M. I., and Sastry, S. S. On the stability of nonlinear receding horizon control: a geometric perspective. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 742–749. IEEE, 2021.

Wu, Y. and He, K. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

Zhang, T., McCarthy, Z., Jow, O., Lee, D., Chen, X., Goldberg, K., and Abbeel, P. Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5628–5635. IEEE, 2018.

Contents

A. Notation, Organization of Appendix, and Full Related Work

In this appendix, we collect the notation we use throughout the paper, as well as providing a high level organization of the appendices.

A.1. Notation Summary

In this section, we summarize some of the notation used throughout the work, divided by subject.

Measure Theory We always let $\mathcal{X}$ denote a Polish space, $\mathscr{B}(\mathcal{X})$ the Borel-algebra on $\mathcal{X}$, and $\Delta(\mathcal{X})$ the set of borel probability measures on $\mathcal{X}$. For a random variable X on $\mathcal{X}$, we let P_X denote the law of X. For random variables X, Y, we let $\mathscr{C}(\mathsf{P}_X, \mathsf{P}_Y)$ denote the set of couplings of these measures and for laws $\mathsf{P}_1, \mathsf{P}_2$. We write $\mathsf{P}_1 \otimes \mathsf{P}_2$ for the product measure. We will generally reserve P to denote measure, Q and W for probability kernels, and μ for a joint measure on several random variables.

When $\mathsf{P}_1, \mathsf{P}_2 \in \Delta(\mathcal{X})$ are laws on the sampe space, we let $\mathrm{TV}(\mathsf{P}_1, \mathsf{P}_2)$ denote the total variation distance. We write $\mathsf{P}_1 \ll \mathsf{P}_2$ if P_1 is absolutely continuous with respect to P_2. Given a Polish space $\mathcal{X}$ and element $x \in \mathcal{X}$, we let $\delta_x \in \Delta(\mathcal{X})$ denote the dirac-delta measure supported on the set $\{x\} \in \mathscr{B}(\mathcal{X})$ (note that, in a Polish space, the singleton $\{x\}$ set is closed, and therefore Borel).

Norms and linear algebra notation. We use bold lower case vector $\mathbf{z}$ to denote vectors, and bold upper case $\mathbf{Z}$ to denote matrices. We let $\mathbf{z}_{1:K} = (\mathbf{z}_1, \dots, \mathbf{Z})$ and $\mathbf{Z}_{1:K} = (\mathbf{Z}_1, \dots, \mathbf{Z}_K)$ denote concatenations. The norms $\|\cdot\|$ denote Euclidean norms on vectors and operator norms on matrices. We identify the spaces $\mathscr{P}_k$ with Euclidean vectors in the standard sence. Given a Euclidean vector $\mathbf{z} \in \mathbb{R}^d$, $\mathcal{N}(\mathbf{z}, \sigma^2\mathbf{I})$ denote the multivariate normal distribution on $\mathbb{R}^d$ with covariance $\sigma^2\mathbf{I}$.

Control notation. We let $\mathbf{x}_t \in \mathbb{R}^{d_x}$ denote control states, $\mathbf{u}_t \in \mathbb{R}^{d_u}$ denote control inputs, and $\boldsymbol{\rho}_\tau \in \mathscr{P}_\tau$ denotes trajectories $(\mathbf{x}_{1:\tau+1}, \mathbf{u}_{1:\tau})$. T denotes the time horizon of imitation, so $\boldsymbol{\rho}_T \sim \mathscr{P}_T$. Our dynamics are $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t)$; for our main results (Section 3), we suppose $f(\mathbf{x}, \mathbf{u}) = \mathbf{x} + \eta f_\eta(\mathbf{x}, \mathbf{u})$, parametrizing dynamics in the form of an Euler discretization with step $\eta > 0$.

Recall that primitive controllers κ take the form $\kappa(\mathbf{x}) = \bar{\mathbf{K}}(\mathbf{x} - \bar{\mathbf{x}}) + \bar{\mathbf{u}}$, where terms with $\bar{(\cdot)}$, $\bar{\mathbf{K}}, \bar{\mathbf{x}}, \bar{\mathbf{u}}$, denote parameters of the primitive controller. The space of these is $\mathcal{K}$.

We also recall the chunk-length τ_{c} and memory length τ_{m} satisfying $0 \le \tau_{\mathrm{m}} \le \tau_{\mathrm{c}}$. We recall the definition of the trajectory-chunk $\boldsymbol{\rho}_{\mathrm{c},h}$ and memory-chunk in $\boldsymbol{\rho}_{\mathrm{m},h}$ in Section 2, which introduced the indexing h, such that $t_h = (h-1)\tau_{\mathrm{c}} + 1$. Recall also the composite actions $\mathsf{a}_h = (\kappa_{t_h:t_{h+1}-1}) \in \mathcal{A} = \mathcal{K}^{\tau_{\mathrm{c}}}$ as the concatenation of τ_{c} primitive controllers.

Abstractions in the composite MDP. The composite MDP is a deterministic MDP with composite-states $\mathsf{s} \in \mathcal{S}$ and composite-actions $\mathsf{a} \in \mathcal{A}$, and (possibly time-varying) deterministic transition dynamics $F_h : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ for $1 \le h \le H$. The goal is to imitate a policy $\pi^\star = (\pi^\star_h)_{1 \le h \le H}$, in terms of imitation gaps $\Gamma_{\mathrm{joint},\varepsilon}$ and $\Gamma_{\mathrm{marg},\varepsilon}$ defined in Definition 4.1. We refer the reader to Section 4 for the relevant terminology, and to Section 4.1 for its instantiation in our original control setting.

A.2. Organization of the Appendix

We now describe the organization of our many appendices. In Appendix B, we expand on our abbreviated discussion of related work in the body as well as provide a more detailed comparison of our notion of stability Definition 4.5 with those found in prior work.

After the preliminaries on organization, notation, and related work, we divide our appendices into two parts. In the first part, we expand on and provide rigorous proofs of statements and results pertaining to the composite MDP as considered in Section 4. We begin by providing a detailed background in Appendix C on the requisite measure theory we use to make our arguments rigorous. In particular, we provide definitions of probability kernels and couplings, as well as measurability properties of optimal transport couplings. In Appendix D, we provide a warmup to the proof of Theorem 2. In particular, the argument substantially simplifies if we consider the case of no added augmentation (when $\sigma = 0$ in TODA) and we present a coupling construction that implies the analogous bound in the presence of an additional assumption. The heart of the first part of our appendices is Appendix E, where we rigorously prove a generalization of Theorem 2 by constructing a sophisticated coupling between the imitator and demonstrator trajectories. We conclude the first part of our appendices by proving a number of lower bounds in the composite MDP setting in Appendix F, which demonstrate the tightness of our arguments in Appendix E.

We continue our appendices in the second part, which is concerned with the instantiation of the composite MDP in the control setting of interest. In Appendix G, we provide a detailed proof that the control setting considered in Section 2 satisfies the stability properties required by our analysis of the composite MDP and prove Proposition 4.1. Of particular note are Definition G.7, which provide explicit dependence of the relevant constants in Theorem 1 on the parameters of interest, and Appendix G.8, which explains how to synthesize stabilizing gains, as assumed in Section 2. With the stability properties thus proven, we proceed in Appendix H to instantiate our conditional sampling guarantees with DDPMs. In particular, by applying earlier work, we state and prove Theorem 6, which guarantees that with sufficiently many samples, in our setting we can ensure that the learned DDPM provides samples close in the relevant optimal transport distance to the expert distribution. We also explain in Remark H.5 why stronger total variation guarantees on sampling are unrealistic in our setting. The heart of the second part of our appendices is Appendix I, which provides the final, end-to-end guarantees and the proof of Theorem 1. In that section, we prove a reduction from imitation learning to conditional sampling and derive Theorem 1 as a corollary. We also provide a number of variations on this result, including stronger guarantees on imitation of the joint trajectories (Appendix I.3), guarantees on TODA under the aassumption that sampling is close in total variation (Appendix I.4), and imitation with no augmentation (Appendix I.5). We also show in Proposition I.5 that most natural cost funtions have similar expected values on imitator and demonstrator trajectories assuming that the imitation losses are small.

We provide a number of extensions of our main results in Appendix J, including to the setting of noisy dynamics (Appendix J.1). Finally, in Appendix K, we expand the discussion of our experiments, including training and compute details, environment details, and a link to our code for the purpose of reproducibility.

B. Complete Related Work

Imitation Learning. Over the past few years, there has been a significant surge of interest in utilizing machine learning techniques for the execution of exceedingly intricate manipulation and control tasks. Imitation learning, whereby a policy is trained to mimic expert demonstrations, has emerged as a highly data efficient and effective method in this domain, with application to self-driving vehicles (Hussein et al., 2017; Bojarski et al., 2016; Bansal et al., 2018), visuomotor policies (Finn et al., 2017; Zhang et al., 2018), and navigation tasks (Hussein et al., 2018). A widely acknowledged challenge of imitation learning is distribution shift: since the training and test time distributions are induced by the expert and trained policies respectively, compounding errors in imitating the expert at test-time can lead the trained policy to explore out-of-distribution states (Ross & Bagnell, 2010). This distribution shift has been shown to result in the imitator making incorrect judgements regarding observation-action causality, often with catastrophic consequences (De Haan et al., 2019). Prior work in this domain has predominantly attempted to mitigate this issue in the non-stochastic setting via online data augmentation strategies, sampling new trajectories to mitigate distribution shift (Ross et al., 2011; Ross & Bagnell, 2010; Laskey et al., 2017). Among this class of methods, the DAgger algorithm in particular has seen widespread adoption (Ross & Bagnell, 2010; Sun et al., 2023; Kelly et al., 2019). These approaches have the drawback that sampling new trajectories or performing queries on the expert is often expensive or intractable. Due to these limitations, recent developments have focused on novel algorithms and theoretical guarantees for imitation learning in an offline, non-interactive environment (Chang et al., 2021; Pfrommer et al., 2022). Our work is similarly focused on the offline setting, but is capable of handling stochastic, non-Markovian demonstrators. Unlike (Pfrommer et al., 2022), we do not require our expert demonstrations to be sampled from a stabilizing expert policy, instead utilizing a synthesis oracle to stabilize around the provided demonstrations. This is a significantly weaker requirement and enables the development of high-probability guarantees for human demonstrators, where sampling new trajectories and reasoning about the stability properties is not possible.

Denoising Diffusion Probabilistic Models. Denoising Diffusion Probabilistic Models (DDPMs) (Sohl-Dickstein et al., 2015; Ho et al., 2020) and their variant, Annealed Langevin Sampling (Song & Ermon, 2019), have seen enourmous empirical success in recent years, especially in state-of-the-art image generation (Ramesh et al., 2022; Nichol & Dhariwal, 2021; Song et al., 2020a). More relevant to this paper is their application to imitation learning, where they have seen success even without the proposed data augmentation in Janner et al. (2022); Chi et al. (2023); Pearce et al. (2023). DDPMs rely on learning the score function of the target distribution, which is generally accomplished through some kind of denoised estimation (Hyvärinen & Dayan, 2005; Vincent, 2011; Song et al., 2020b). On the theoretical end, annealed Langevin sampling has been studied with score estimators under a variety of assumptions including the manifold hypothesis and some form of dissapitivity (Raginsky et al., 2017; Block et al., 2020a;b), although these works have generally suffered from an exponential dependence on ambient dimension, which is unacceptable in our setting. Of greatest relevance to the present paper are the concurrent works of Chen et al. (2022); Lee et al. (2023) that provide polynomial guarantees on the quality of sampling using a DDPM assuming that the score functions are close in an appropriate mean squared error sense. We take advantage of these latter two works in order to provide concrete end-to-end bounds in our setting of interest. To our knowledge, ours is the first work to consider the application of DDPMs to imitation learning under a rigorous theoretical framework, although we emphasize that this does not constitute a strong technical contribution as opposed to an instantiation of earlier work for the sake of completeness and concreteness.

Smoothing Augmentations. Data augmentation with smoothing noise has become such common practice, its adoption is essentially folklore. While augmentation of actions which noise is common practice for exploration (see, e.g. (Laskey et al., 2017)), it is widely accepted that noising actions in the learned policy is not best practice, and thus it is more common to add noise to the *states* at training time, preserving target actions as fixed (Ke et al., 2021). Our work gives an interpretation of this decision as enforcing that the learned policy obey the distributional continuity property we term TVC (Definition 4.3), so that the policy selects similar actions on nearby states. Previous work has interpreted noise augmentation as providing robustness. Data augmentation has been demonstrated to provide more robustness in RL from pixels (Kostrikov et al., 2020), adaptive meta-learning (Ajay et al., 2022), in more traditional supervised learning as well (Hendrycks et al., 2020).

B.1. Comparison to prior notions of Stability.

Prior work in guarantees for imitation learning focuses either on constraining the learned policy to be stable (Havens & Hu, 2021; Tu et al., 2022) or assume the expert policy is suitably stable (Pfrommer et al., 2022).

The principal notion of stability used in these prior works is *incremental-input-to-state* stability of the closed-loop system under a deterministic controller π:

Definition B.1 (Incremental Input-to-State Stability)**.** There exists class $\mathcal{K}$ function γ and class $\mathcal{KL}$ function β such that for any two initial conditions $\xi_1, \xi_2 \in \mathcal{X}$, the closed-loop dynamics under policy $\pi : \mathcal{X} \to \mathcal{U}$ given by $f_{\mathrm{cl}}(x_t, \Delta_t) = f(x_t, \pi(x_t) + \Delta_t)$ satisfies:

$$\|x_t(\xi_1; \{\Delta_s\}_{s=0}^t) - x_t(\xi_2; \{0\}_{s=0}^t)\| \le \beta(\|\xi_1 - \xi_2\|) + \gamma\left(\max_{0 \le s \le t-1} \|\Delta_s\|\right),$$

where $x_t(\xi; \{\Delta_s\}_{s=0}^{t-1})$ is the state at time t under f_{cl} with $x_0 = \xi$ and input perturbations $\{\Delta_s\}_{s=0}^{t-1}$.

This notion of stability is quite restrictive, as the β-term necessitates that the dynamics converge irrespective of initial condition. Without time-varying dynamics this can only be achieved by a policy which stabilizes to an equilibrium point, as a policy which tracks a reference trajectory is unable to "forget" the initial condition. Constraining learned policies such that they satisfy this notion of stability is also challenging. Tu et. al. (Tu et al., 2022) attempt to do so through regularization while Haven et. a. (Havens & Hu, 2021) use matrix inequalities to satisfy this stability property under linear dynamics. Pfrommer et. at. (Pfrommer et al., 2022) avoid this difficulty by relaxing the incremental stability to a local variant of stability:

Definition B.2 (η-Local Incremental Input-to-State Stability)**.** There exists class $\mathcal{K}$ function γ such that for any $\xi \in \mathcal{X}$, the closed-loop dynamics under policy $\pi : \mathcal{X} \to \mathcal{U}$ given by $f_{\mathrm{cl}}(x_t, \Delta_t) = f(x_t, \pi(x_t) + \Delta_t)$ satisfies:

$$\|x_t(\xi; \{\Delta_s\}_{s=0}^t) - x_t(\xi; \{0\}_{s=0}^t)\| \le \gamma\left(\max_{0 \le s \le t-1} \|\Delta_s\|\right),$$

for all $\{\Delta_s\}_{s=0}^t$ where $\max_{0 \le s \le t} \|\Delta_s\| \le \eta$.

This weaker notion of incremental stability simply postulates the existence of a (local) input-perturbation to state-perturbation gain function γ. Since this stability property does not necessitate convergence across with different initial conditions and only under input perturbations of magnitude $\le \eta$, this only necessitates that the expert policy can correct from small input perturbations.

We further weaken this assumption, which we formalize in Assumption 3.2 and abstract to the composite MDP through Definition G.4, by only requiring that a locally stabilizing controller can be synthesized per-demonstration. Through the introduction of a synthesis oracle which can generate locally stabilizing primitive controllers, we decouple the stability

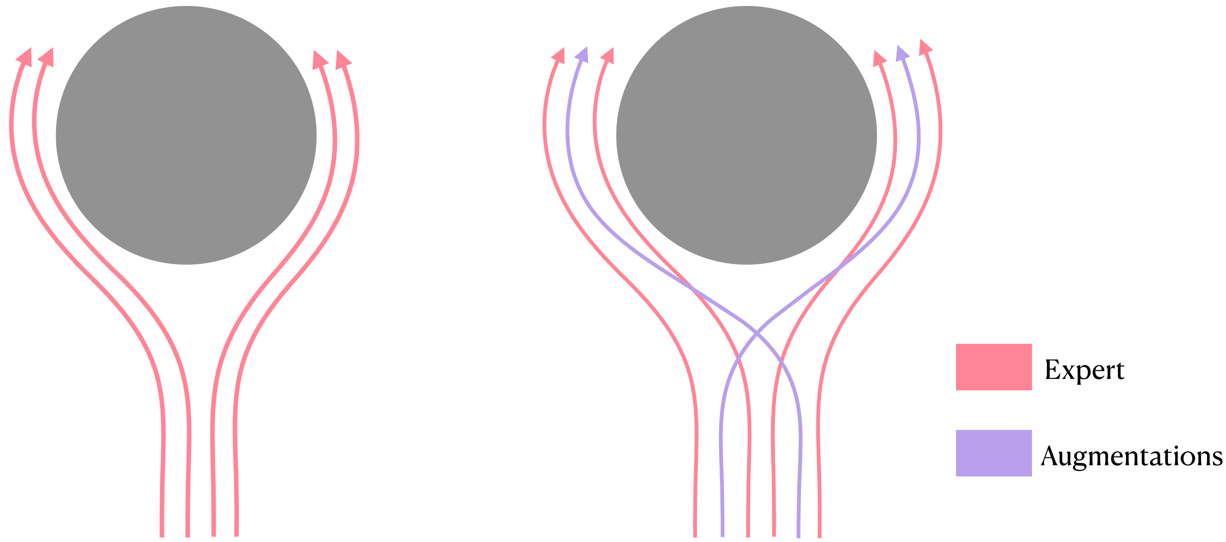


Figure 3. Instance of bifurcation, where augmentation is necessary for stability. The example on the left has an expert demonstrator bifurcating around a circular obstacle. The example on the right demonstrates the utility of augmentations, allowing for trajectories that navigate around the object in the direction farther from their starting point.

properties of the expert from the stabilizability of the underlying dynamical system. This allows for reasoning about generalization in the presence of bifurcations or conflicting demonstrations, which is precluded by Definition B.2 since an expert policy cannot simultaneously stabilize to multiple branches of a bifurcation. For a concrete example, consider Figure 3. Indeed, continuity is the *sine qua non* of stability and the example given demonstrates the necessity of augmentation to enforce the former. In detail, the figure illustrates an example where an agent is navigating around an obstacle, providing a bifurcation. Without augmentation, the demonstrator trajectories always navigate around the obstacle in the direction closer to their starting point, leading to a sharp discontinuity along a bisector of the obstacle. On the other hand, the data augmentations allow for the policy to have some probability of navigating around the obstacle in the "wrong" direction, which leads to the notion of continuity we consider: total variation continuity.

Because our notion of stability is applied in chunks, our theory is sufficiently flexible so as to allow for the learned policy to switch between expert demonstrations in a manner preserving the marginal distributions but not consistent with the joint distribution across the entire trajectory. This flexibility is illustrated in Figure 4, where we suppose that the demonstrator distribution consists both of trajectories traversing a figure "8" consistently in either a clockwise or counter-clockwise manner, with both orientations represented in the data set. Due to the multi-modality at the critical point in the trajectory, there is ambiguity about which loop to traverse next; specifically, there may exist a policy that randomly select which loop to traverse each time the critical point is visited in such a way that the marginal distributions on states and actions is the same as that induced by the demonstrator. Such a policy will, by definition, preserve the correct *marginal* distributions across states and actions; at the same time, this policy has a different *joint* distribution across all time steps from the demonstrator due to the possibility of traversing the same loop twice in a row.

Part I

Composite MDP

C. Measure-Theoretic Background

In this section, we introduce the prerequisite notions from probability theory that we use to formally construct the couplings in Appendices D and E. We begin by introducing general preliminaries, followed by kernels, regular conditional probabilities

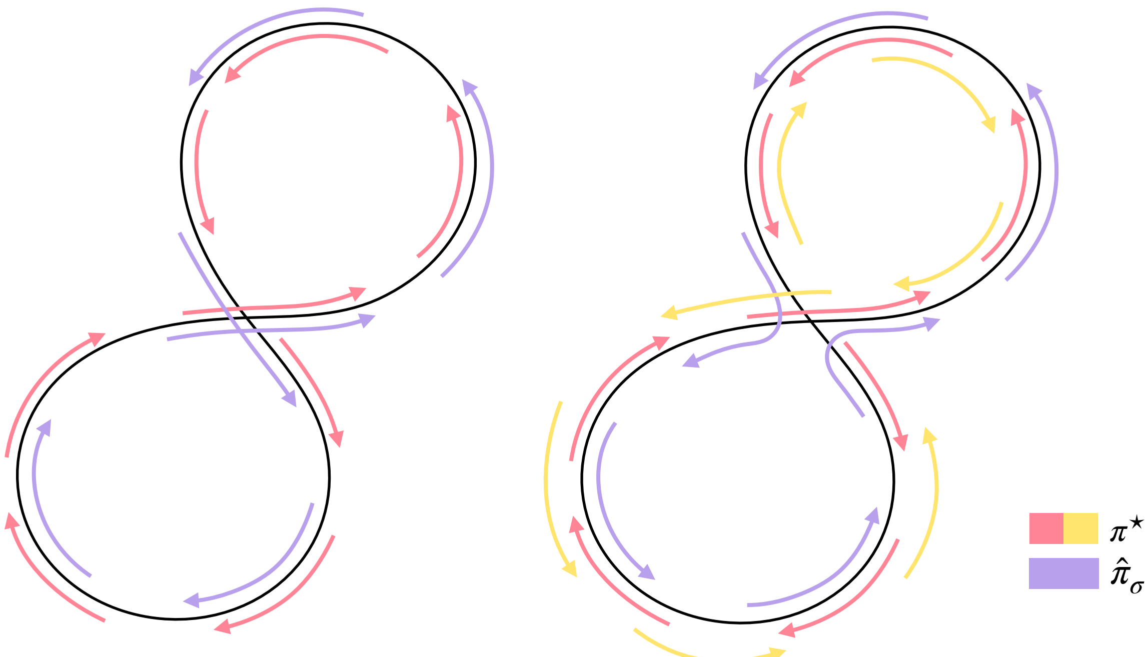


Figure 4. Instance where $\hat{\pi}_\sigma$ and $\pi^\star$ induce the same marginals and joint distributions (left), but in the presence of expert demonstration trajectories that traverse the figure eight both clockwise and counterclockwise directions, $\hat{\pi}_\sigma$ may switch with some probability between demonstrations where they overlap.

and a "gluing" lemma in Appendix C.1. We then show that optimal transport costs commute in an appropriate sense with conditional probabilities (Proposition C.3 in Appendix C.2). We use the preliminaries in the previous sections to derive certain optimal-transport and data processing inequalities in Appendix C.3. We prove Proposition C.3 in Appendix C.4. Finally, we state a simple union bound lemma (Lemma C.11 in Appendix C.5) of use in later appendices.

General preliminaries. We rely extensively on the exposition in Durrett (2019) and refer the reader there for a more thorough introduction. Throughout, we assume there is a Polish space Ω such that all random variables of interest are mappings $X : \Omega \to \mathcal{X}$, where $\mathcal{X}$ is also Polish. Here, the σ-algebras are always the Borel algebras (the σ-algebra generated by open subsets), denoted $\mathscr{B}(\Omega)$ and $\mathscr{B}(\mathcal{X})$.

The space of (Borel) probablity distributions on $\mathcal{X}$ is denoted $\Delta(\mathcal{X})$, and measurability is meant in the Borel sense. Given a measure μ on a space $\mathcal{X} \times \mathcal{Y}$, we say that $X \sim \mathsf{P}_X$ under μ if, for all $A \in \mathscr{B}(\mathcal{X})$, $\mu(X \in A) = \mathsf{P}_X(A)$.

We adopt standard information theoretic notation to denote joint, marginal, and conditional distributions on vectors of random variables. In particular, if random variables X, Y are distributed according to P, we denote by P_X as the marginal over X, $\mathsf{P}_{X|Y}$ as the conditional of $X|Y$ under P, and $\mathsf{P}_{X,Y}$ as the joint distribution when this needs to be empasized.

Definition C.1 (Couplings)**.** Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces and let $\mathsf{P}_X \in \Delta(\mathcal{X})$ and $\mathsf{P}_Y \in \Delta(\mathcal{Y})$. The set of couplings $\mathscr{C}(\mathsf{P}_X, \mathsf{P}_Y)$ denotes the set of measure $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$ such that, $(X, Y) \sim \mu$ has marginals $X \sim \mathsf{P}_X$ and $Y \sim \mathsf{P}_Y$.[2] We let $\mathsf{P}_X \otimes \mathsf{P}_Y \in \mathscr{C}(\mathsf{P}_X, \mathsf{P}_Y)$ denote the *indepent coupling* under which X and Y are independent.

It is standard that $\mathsf{P}_X \otimes \mathsf{P}_Y$ is always a valid coupling, and hence $\mathscr{C}(\mathsf{P}_X, \mathsf{P}_Y)$ is nonempty. Couplings have the advantage that they can be used to design many probability-theoretic distances. Through the paper, we use the total variation distance.

Definition C.2 (Total Variation Distance)**.** Let $\mathsf{P}_1, \mathsf{P}_2 \in \Delta(\mathcal{X})$. We define the total variation distance $\mathrm{TV}(\mathsf{P}_1, \mathsf{P}_2) := \sup_{A \subset \mathscr{B}(\mathcal{X})} |\mathsf{P}_1(A) - \mathsf{P}_2(A)|$

The total variation distance can be expressed in terms of couplings as follows (Polyanskiy & Wu, 2022+).

[2]More pedantically, for all Borel sets $A_1 \in \mathscr{B}(\mathcal{X})$, $\mu(A_1 \times \mathcal{Y}) = \mathsf{P}_X(A_1)$ all Borel sets $A_2 \in \mathscr{B}(\mathcal{X})$, $\mu(\mathcal{X} \times A_2) = \mathsf{P}_2(A_2)$.

Lemma C.1. *Let* $\mathsf{P}_1, \mathsf{P}_2 \in \Delta(\mathcal{X})$. *Then,*

$$\mathrm{TV}(\mathsf{P}_1, \mathsf{P}_2) = \inf_{\mu \in \mathscr{C}(\mathsf{P}_1, \mathsf{P}_2)} \mathbb{P}_{(X_1, X_2) \sim \mu}\{X_1 \neq X_2\}.$$

Moreover, there exists a coupling $\mu_\star$ *attaining the infinum.*

Support and absolute continuity. We will also require the definition of the support of a measure.

Definition C.3. Given a measure μ on a Borel space $(\Omega, \mathcal{F})$, we define the *support* $\mathrm{supp}(\mu)$ to be the closure in the topology given by the metric of the set $\{\omega \in \Omega | \mu(\mathcal{U}) > 0 \text{ for all open } U \ni \omega\}$.

In addition, we require the definition of absolute continuinty.

Definition C.4 (Absolute Continuity)**.** We say that $\mathsf{P} \in \Delta(\mathcal{X})$ is absolutely continuous with respect to law $\mathsf{P}' \in \Delta(\mathcal{X})$, written $\mathsf{P} \ll \mathsf{P}'$, if for $A \in \mathscr{B}(\mathcal{X})$, $\mathsf{P}'(A) = 0$ implies $\mathsf{P}(A) = 0$.

We now go into greater detail on the kinds of couplings that we consider.

C.1. Kernels, Regular Conditional Probabilities and Gluing

One key technical challenge in proving results in the sequel is the fact that we need to "glue" together multiple different couplings. Specifically, while it may be the case that there exist pairwise couplings which satisfy desired properties, there exists a coupling such that the probability of the relevant event is small, it is not obvious that there exists a *single* coupling such that all of these probabilities are small *simultaneously*. There are two natural ways to due this gluing: the first, using regular conditional probabilities we provide here. The second, involving a sophisticated construction of Angel & Spinka (2019) requires stronger assumptions on the pseudo-metric, but generalizing beyond Polish spaces, we simply remark can be substituted with a loss of a constant factor.

Kernels. We begin by introducing the notion of a kernel.

Definition C.5 (Kernels)**.** Let $(\Omega, \mathbb{P})$ be a probability space and let X denote a random variable on this space. For a given σ-algebra $\mathcal{G}$, and map $Q : \Omega \times \mathcal{G} \to [0, 1]$, we say that Q is a probability kernel if the following two conditions are satisfied:

1. For all measurable events A, the map $\omega \mapsto Q(\omega, A)$ is measurable.
2. For almost every $\omega \in \Omega$, the map $A \mapsto Q(\omega, A)$ is a probability measure.

We can combine a probability kernel with a probabilty measure on $\mathcal{Y}$ to yield joint distributions over $\mathcal{X} \times \mathcal{Y}$.

Definition C.6. Given an $\mathsf{P}_Y \in \Delta(\mathcal{Y})$, we define the probability measure $\mathrm{law}(\mathsf{Q}_{X|Y}; \mathsf{P}_Y) \in \Delta(\mathcal{X} \times \mathcal{Y})$ such that $\mu = \mathrm{law}(\mathsf{Q}_{X|Y}; \mathsf{P}_Y)$ satisfies[3]

$$\mu(A \times B) = \mathbb{E}_{Y \sim \mathsf{P}_Y}\left[\mathsf{Q}_{X|Y}(A \mid Y)\mathbf{I}\{Y \in B\}\right], \quad \forall A \in \mathscr{B}(\mathcal{X}), B \in \mathscr{B}(\mathcal{Y}). \tag{C.1}$$

We let $\mathsf{Q}_{X|Y} \circ \mathsf{P}_Y \in \Delta(\mathcal{X})$ denote the measure for which $\mu = \mathsf{Q}_{X|Y} \circ \mathsf{P}_Y$ satisfies

$$\mu(A) = \mathbb{E}_{Y \sim \mathsf{P}_Y}\left[\mathsf{Q}_{X|Y}(A \mid Y)\right], \quad \forall A \in \mathscr{B}(\mathcal{X})$$

From these, we define the space of conditional couplings as follows.

Definition C.7 (Kernel Couplings)**.** Let $\mathsf{P}_Y \in \Delta(\mathcal{Y})$, and $\mathsf{Q}_{X_i|Y} \in \Delta(\mathcal{X} \mid \mathcal{Y})$ for $i \in \{1, 2\}$. We let $\mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_{X_1|Y}, \mathsf{Q}_{X_1|Y})$ denote the space of measures $\mu \in \Delta(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y})$ over random variables (X_1, X_2, Y) such that $(X_i, Y) \sim \mathrm{law}(\mathsf{Q}_{X|Y}; \mathsf{P}_Y)$ for $i \in \{1, 2\}$.

Note that a similar construction to the independent coupling ensures $\mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_{X_1|Y}, \mathsf{Q}_{X_1|Y})$ is nonempty, namely considering the measure $\mu(A_1 \times A_2 \times B_2) = \mathbb{E}_{Y \sim \mathsf{P}_Y}\left[\mathsf{Q}_{X_1|Y}(A_1 \mid Y)\mathsf{Q}_{X_2|Y}(A_2)\mathbf{I}\{Y \in B\}\right]$.

[3]Recall that $\mathscr{B}(\mathcal{X} \times \mathcal{Y})$ is generated by sets $A \times B \in \mathscr{B}(\mathcal{X}) \times \mathscr{B}(\mathcal{Y})$, so (C.1) defines a unique probability measure

Regular Conditional Probabilities. We now recall a standard result that conditional probabilities can be expressed through kernels in our setting.

Theorem 3 (Theorem 5.1.9, Durrett (2019))**.** *If Ω is a Polish space and $\mathbb{P}$ is a probability measure on the Borel sets of Ω, such that random variables $(X, Y) \sim \mathbb{P}$ in spaces $\mathcal{X}$ and $\mathcal{Y}$, then there exists a kernel $\mathsf{Q}(\cdot \mid \cdot) \in \Delta(\mathcal{X} \mid \mathcal{Y})$ such that, for all $A \in \mathscr{B}(\mathcal{X})$ and $\mathbb{P}$-almost every y, the (standard) conditional probability $\mathbb{P}[X \in A \mid Y] = \mathsf{Q}(A \mid y)$. We can $\mathsf{Q}(\cdot \mid \cdot)$ the* regular conditional probability measure.

Regular conditional probabilities allow one to think of conditional probabilities in the most intuitive way, i.e., for two random variables X, Y, the map $Y \mapsto \mathbb{P}(X \in A \mid Y)$ is a probability kernel. This will be the essential property that we use below.

Gluing. Finally, regular conditional probabilities allow us to "glue together" couplings which share a common random variable.

Lemma C.2 (Gluing Lemma)**.** *Suppose that X, Y, Z are random variables taking value in Polish spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$. Let $\mu_1 \in \Delta(\mathcal{X} \times \mathcal{Y}), \mu_2 \in \Delta(\mathcal{Y} \times \mathcal{Z})$ be couplings of (X, Y) and (Y, Z) respectively. Then there exists a coupling $\mu \in \Delta(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ on (X, Y, Z) such that under μ, $(X, Y) \sim \mu_1$ and $(Y, Z) \sim \mu_2$.*

Proof. Let $\mathcal{Q}(\cdot \mid Y)$ be a regular conditional probability for Z given Y under μ_2 (who existence is ensured by Theorem 3). We construct μ by first sampling $(X, Y) \sim \mu_1$ and then sampling $Z \sim \mathsf{Q}(\cdot \mid Y)$; observe that by the second property in Definition C.5, this is a valid construction. It is immediate that under μ, we have $(X, Y) \sim \mu_1$ and thus we must only show that $(Y, Z) \sim \mu_2$ to conclude the proof. Let A, B be two measurable sets and we see that

$$\begin{aligned}
\mathbb{P}_\mu \left((Y, Z) \in A \times B\right) &= \mathbb{E}_{Y \sim \mu} \left[\mathbb{P}_\mu \left((Y, Z) \in A \times B | Y\right)\right] \\
&= \mathbb{E}_{Y \sim \mu} \left[\mathbb{E}_{(Y,Z) \sim \mu} \left[\mathbf{I}[Y \in A] \cdot \mathbf{I}[Z \in B] | Y\right]\right] \\
&= \mathbb{E}_{Y \sim \mu} \left[\mathbf{I}[Y \in A] \cdot \mathbb{E}_\mu \left[\mathbf{I}\left[Z \in B\right] | Y\right]\right] \\
&= \mathbb{E}_{Y \sim \mu} \left[\mathbf{I}[Y \in A] \cdot \mathbb{P}_{\mu_2}(Z \in B | Y)\right] \\
&= \mu_2 \left((Y, Z) \in A \times B\right),
\end{aligned}$$

where the first equality follows from the tower property of expectations, the second follows by definition of conditional probability, the third follows from the definition of conditional expectation, the fourth follows by the first property from Definition C.5, and the last follows from the fact that the marginals of Y under μ and under μ_2 are the same. The result follows. ☐

C.2. Optimal Transport and Kernel Couplings

As shown above for the TV distance, many measures of distributional distance can be quantified in terms of *optimal transport* costs; these are quantities expressed as infima, over all couplings, of the expectation of a certain lower-semicontinuous functions. We show that if the optimal transport costs between two kernels $Y \to \Delta(\mathcal{X}_i)$ are controlled pointwise, then for any $\mathsf{P}_Y \in \Delta(\mathcal{Y})$, is a there exists a joint distribution over (X_1, X_2, Y) which attains the minimal transport cost.

Proposition C.3. *Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$ be Polish spaces, and let $\mathsf{P}_Y \in \Delta(\mathcal{Y})$, and $\mathsf{Q}_i \in \Delta(\mathcal{X}_i \mid \mathcal{Y})$. for $i \in \{1, 2\}$. Finally, let $\phi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbb{R}$ be lower semicontinuous and bounded below. Then, the following function*

$$\psi(y) := \inf_{\mu \in \mathscr{C}(\mathsf{Q}_1(y), \mathsf{Q}_2(y))} \mathbb{E}_{(X_1, X_2) \sim \mu}[\phi(X_1, X_2)]$$

is a measurable function of y and there exists some $\mu_\star \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1, \mathsf{Q}_2)$ such that

$$\mathbb{E}_{(X_1, X_2, Y) \sim \mu_\star}[\phi(X_1, X_2)] = \mathbb{E}_{Y \sim \nu_Y} \psi(Y).$$

In particular it holds $\mu_\star$-almost surely that

$$\mathbb{E}_{\mu_\star}[\phi(X_1, X_2) | Y] = \psi(Y).$$

We prove the above proposition in Appendix C.4. One useful consequence is the following identity for the total variation distance.

Corollary C.1. *Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces, and let $\mathsf{P}_Y \in \Delta(\mathcal{Y})$, and $\mathsf{Q}_i \in \Delta(\mathcal{X} \mid \mathcal{Y})$, for $i \in \{1,2\}$. Then, there exists a coupling $\mu_\star \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1, \mathsf{Q}_2)$ such that*

$$\mathbb{P}_{\mu_\star}[X_1 \neq X_2] = \mathbb{E}_{Y \sim \mathsf{P}_Y} \mathsf{TV}(\mathsf{Q}_1(\cdot \mid Y), \mathsf{Q}_2(\cdot \mid Y)),$$

with the left-hand side integrand being measurable.

Proof. Using Lemma C.1, we can represent total variation as an optimal transport cost with $\phi(x_1, x_2) = \mathbf{I}\{x_1 \neq x_2\}$. Note that $\phi(x_1, x_2)$ is lower semicontinuous, being the indicator of an open set. Thus, the result follows from Proposition C.3 with $\mathcal{X} = \mathcal{X}_1 = \mathcal{X}_2$, and $\phi(x_1, x_2) = \mathbf{I}\{x_1 \neq x_2\}$. □

C.3. Data Processing Inequalities

We now derive two *inequalities*. First, we recall the classical version for the total variation distance, and check that a well-known identity holds in our setting.

Lemma C.4 (Data Processing for Total Variation)**.** *Let $\mathsf{P}_{Y_1}, \mathsf{P}_{Y_2} \in \Delta(\mathcal{Y})$ and let $\mathsf{Q}_X \in \Delta(\mathcal{X} \mid \mathcal{Y})$. Then,*

$$\mathsf{TV}(\mathsf{Q}_X \circ \mathsf{P}_{Y_1}, \mathsf{Q}_X \circ \mathsf{P}_{Y_2}) \leq \mathsf{TV}(\mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_1}), \mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_2})) = \mathsf{TV}(\mathsf{P}_{Y_1}, \mathsf{P}_{Y_2}).$$

Proof. The first inequality is just the data processing inequality (Polyanskiy & Wu, 2022+, Theorem 7.7), which also shows that $\mathsf{TV}(\mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_1}), \mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_2})) \geq \mathsf{TV}(\mathsf{P}_{Y_1}, \mathsf{P}_{Y_2})$. To prove the reverse inequality, we use Lemma C.1 to find a coupling μ_Y such that $(\mathsf{P}_{Y_1}, \mathsf{P}_{Y_2})$ such that $\mathbb{E}[\mathbf{I}\{Y_1 \neq Y_2\}] = \mathsf{TV}(Y_1, Y_2)$.

Define a probability kernel in $\Delta(\mathcal{X} \times \mathcal{X} \mid \mathcal{Y}_1 \times \mathcal{Y}_2)$ via defining the set $B_= \{(x_1, x_2) \in \mathcal{X} \times \mathcal{X} : x_1 = x_2\} \subset \mathcal{X} \times \mathcal{X}$, and define for $A \in \mathscr{B}(\mathcal{X} \times \mathcal{X})$,

$$\mathsf{Q}(A \mid y_1, y_2) = \begin{cases} \mathsf{Q}_X\left(\pi_1\left(A \cap B_=\right) \mid y_1\right) & y_1 = y_2 \\ \mathsf{Q}_X(\cdot \mid y_1) \otimes \mathsf{Q}_X(\cdot \mid y_2)(A) & \text{otherwise} \end{cases}$$

In a Polish space, Lemmas C.6 and C.7 imply that $A \mapsto \mathsf{Q}_X\left(\pi_1\left(A \cap B_=\right) \mid y_1\right)$ for eacy y_1 is a valid measure, and it is standard that the product measures $\mathsf{Q}_X(\cdot \mid y_1) \otimes \mathsf{Q}_X(\cdot \mid y_2)(A)$ are valid. Moreover, this construction ensures that for $\mu = \mathrm{law}(\mathsf{Q}; \mu_Y)$,

$$\mathbb{P}_\mu[\{Y_1 = Y_2\} \text{ and } \{X_1 \neq X_2\}] = 0. \tag{C.2}$$

Lastly, one can check that under $\mu = \mathrm{law}(\mathsf{Q}; \mu_Y)$, that $(X_1, Y_1) \sim \mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_1})$ and $(X_2, Y_2) \sim \mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_2})$. Thus, μ can be regarded as an element of $\mathscr{C}(\mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_1}), \mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_2}))$. Hence, Lemma C.1 implies that

$$\begin{aligned} \mathsf{TV}(\mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_1}), \mathrm{law}(\mathsf{Q}_X; \mathsf{P}_{Y_2})) &\leq \mathsf{TV}(\mathbb{P}_\mu[(X_1, Y_1) \neq (X_2, Y_2)] \\ &= \mathbb{P}_\mu[Y_1 \neq Y_2] + \mathbb{P}_\mu[\{Y_1 = Y_2\} \text{ and } \{X_1 \neq X_2\}] \\ &= \mathbb{P}_{\mu_\star}[Y_1 \neq Y_2] && \text{(Eq.(C.2))} \\ &= \mathbb{P}_{(Y_1, Y_2) \sim \mu_Y}[Y_1 \neq Y_2] \\ &= \mathsf{TV}(\mathsf{P}_{Y_1}, \mathsf{P}_{Y_2}). && \text{(construction of } \mu_Y) \end{aligned}$$

□

Next, we derive a general data processing inequality for optimal costs. This result is a corollary of Proposition C.3.

Lemma C.5 (Another Data Processing Inequality for Optimal Transport)**.** *Let $\mathcal{X}_1, \mathcal{X}_2, \mathcal{Y}$ be Polish spaces, and let $\mathsf{P}_Y \in \Delta(\mathcal{Y})$, and $\mathsf{Q}_i \in \Delta(\mathcal{Y} \mid \mathcal{X}_i)$. for $i \in \{1,2\}$. Denote by $\mathsf{Q}_i \circ \mathsf{P}_Y$ the marginal of X_i under $(X_i, Y) \sim \mathrm{law}(\mathsf{Q}_i; \mathsf{P}_Y)$. Then,*

$$\inf_{\mu \in \mathscr{C}(\mathsf{Q}_1 \circ \mathsf{P}_Y, \mathsf{Q}_2 \circ \mathsf{P}_Y)} \mathbb{E}_{X_1, X_2 \sim \mu} \phi(X_1, X_2) \leq \mathbb{E}_{Y \sim \mu_Y} \left(\inf_{\mu' \in \mathscr{C}(\mathsf{Q}_1(Y) \circ \mathsf{Q}_2(Y))} \mathbb{E}_{X_1, X_2 \sim \mu'} \phi(X_1, X_2) \right).$$

Proof. One can check that any coupling in $\mu \in \mathscr{C}(\mathsf{Q}_1 \circ \mathsf{P}_Y, \mathsf{Q}_2 \circ \mathsf{P}_Y)$ can be obtained by marginalizing Y in a certain coupling of $\mu' \in \mathscr{C}(\mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y), \mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y))$, and any coupling in the latter can be marginalized to a coupling in the former. Hence,

$$\inf_{\mu \in \mathscr{C}(\mathsf{Q}_1 \circ \mathsf{P}_Y, \mathsf{Q}_2 \circ \mathsf{P}_Y)} \mathbb{E}_{X_1, X_2 \sim \mu} \phi(X_1, X_2) = \inf_{\mu \in \mathscr{C}(\mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y), \mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y))} \mathbb{E}_{X_1, X_2, Y_1, Y_2 \sim \mu} \phi(X_1, X_2)$$

Moreover, to every measure $\mu \in \mu_{\mathsf{P}_Y}(\mathsf{Q}_1, \mathsf{Q}_2)$ over (X_1, X_2, Y), Lemma C.8 implies that there exists a coupling $\mu' \in \mathscr{C}(\mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y), \mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y))$ over (X_1, X_2, Y_1, Y_2) such (X_1, X_2) have the same marginals under μ and μ'. Therefore,

$$\inf_{\mu \in \mathscr{C}(\mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y), \mathrm{law}(\mathsf{Q}_1; \mathsf{P}_Y))} \mathbb{E}_{X_1, X_2, Y_1, Y_2 \sim \mu} \phi(X_1, X_2) \leq \inf_{\mu' \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1, \mathsf{Q}_2)} \mathbb{E}_{X_1, X_2, Y \sim \mu} \phi(X_1, X_2).$$

Finally, the right hand side is equal to $\mathbb{E}_{Y \sim \mu_Y} \left(\inf_{\mu' \in \mathscr{C}(\mathsf{Q}_1(Y) \circ \mathsf{Q}_2(Y))} \mathbb{E}_{X_1, X_2 \sim \mu'} \phi(X_1, X_2) \right)$ by Proposition C.3. □

C.3.1. Deferred lemmas for the data processing inequalities

Lemma C.6. *Let $\mathcal{X}$ be a Polish space. Then, the set $\{(x_1, x_2) \in \mathcal{X} \times \mathcal{X} : x_1 \neq x_2\}$ is open in $\mathcal{X} \times \mathcal{X}$.*

Proof. The diagonal is closed in any Polish space by definition of the topology. The result follows. □

Lemma C.7. *Let $\mathcal{X}$ be a Polish space, and let $\pi_1, \pi_2 : \mathcal{X} \times \mathcal{X}$ denote the projection mappings onto each coordinate. Then, for any $A \in \mathscr{B}(\mathcal{X} \times \mathcal{X})$, $\pi_1(A)$ and $\pi_2(A)$ are in $\mathscr{B}(\mathcal{X})$.*

Proof. The projection map is open so the result follows immediately by definition of the Borel algebra. □

Lemma C.8. *Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces, and let $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$. Then, there is a measure $\mu' \in \Delta(\mathcal{X} \times \mathcal{Y} \times \mathcal{Y})$ satisfying*

$$\mu'(A \times \mathcal{Y}) = \mu(A), \quad \forall A \in \mathscr{B}(\mathcal{X} \times \mathcal{Y})$$

and

$$\mu'(\mathcal{X} \times \{(y_1, y_2) : y_1 = y_2\}) = 1$$

Proof. Define the set $B_= = \{(y_1, y_2) : y_1 = y_2\}$. One can check that $\mu'(A \times B) = \mu(A \times \pi_1(B \cap B_=)$, where $\pi_1 : \mathcal{Y} \times \mathcal{Y} \to \mathcal{Y}$ is the projection onto the first coordinate, is a valid measure. □

C.4. Proof of Proposition C.3

In the case that $\phi(\cdot, \cdot)$ is continuous, the result follows from Villani et al. (2009, Corollary 5.22). For general lower-semicontinuous ϕ, our argument adopts the strategy of "Step 3" of the proof of Villani (2021, Theorem 1.3). This shows that there exists a sequence $\phi_n \uparrow \phi$ pointwise, such that each ϕ_n is uniformly bounded. Define

$$\psi_n(y) := \inf_{\mu \in \mathscr{C}(\mathsf{Q}_1(y), \mathsf{Q}_2(y))} \mathbb{E}_{(X_1, X_2) \sim \mu}[\phi_n(X_1, X_2)].$$

Then, for each n, the continuous case implies that there exists a measure $\mu_{\star,n} \in \mathscr{C}_{\nu_Y}(\mathsf{Q}_1, \mathsf{Q}_2)$ such that

$$\mathbb{E}_{Y \sim \nu_Y} \psi_n(Y) = \mathbb{E}_{(X_1, X_2, Y) \sim \mu_{\star,n}}[\phi_n(X_1, X_2)] \tag{C.3}$$

Recall now the definition

$$\psi(y) = \inf_{\mu \in \mathscr{C}(\mathsf{Q}_1(y), \mathsf{Q}_2(y))} \mathbb{E}_{(X_1, X_2) \sim \mu}[\phi(X_1, X_2)].$$

Claim C.9. *$\psi(y)$ is measurable and satisfies $\psi_n(y) \uparrow \psi(y)$ pointwise.*

Proof. We can write

$$\sup_{n\geq 0}\psi_n(y) = \sup_{n\geq 0}\inf_{\mu\in\mathscr{C}(\mathsf{Q}_1(y),\mathsf{Q}_2(y))}\mathbb{E}_{(X_1,X_2)\sim\mu}[\phi_n(X_1,X_2)]$$
$$\overset{(i)}{=}\inf_{\mu\in\mathscr{C}(\mathsf{Q}_1(y),\mathsf{Q}_2(y))}\mathbb{E}_{(X_1,X_2)\sim\mu}[\phi(X_1,X_2)] = \psi(y).$$

Here, (i) follows from the "Step 3" in the proof of Villani (2021, Theorem 1.3), which shows that any optimal transport cost C of a lowersemicontinuous ϕ is equal to a limit of the costs C_n of any bounded continuous $\phi_n \uparrow \phi$. In our case, we fix each y, so $C = \psi(y)$ and $C_n = \psi_n(y)$. It is clear that $\psi_n(y)$ is increasing, so for each y, $\psi_n(y) \uparrow \psi(y)$. As ψ is the pointwise monotone limit of ψ_n, it is measurable. □

Claim C.10. *The set of couplings of $\mathscr{C}_{\mathsf{P}_Y}(X_1, X_2)$ is compact in the weak topology.*

Proof. Recall that $\Delta(\mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2)$ denote the set of Borel measures on $\mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2$. This set is also a Polish space in the weak topology. The subset $\mathscr{C}_{\mathsf{P}_Y}(X_1, X_2) \subset \Delta(\mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2)$ is compact if and only if it is relatively compact and closed.

To show relative compactness, Prokhorov's theorem means that it suffices to show that $\mu_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2)$ is tight, i.e. for all $\varepsilon > 0$, there exists a compact $\mathcal{K}_\varepsilon \subset \mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2$ such that for any $\mu \in \mathscr{C}_{\mathsf{P}_Y}(X_1,X_2)$, $\mathbb{P}_\mu[(Y,X_1,X_2)\in\mathcal{K}_\varepsilon] \geq 1-\varepsilon$. This follows by setting $\mathcal{K} = \mathcal{K}_{Y,\varepsilon}\times\mathcal{K}_{X,1,\varepsilon}\times\mathcal{K}_{X,2,\varepsilon}$, where the sets are such that $\mathbb{P}_{\mathsf{P}_Y}[Y\notin\mathcal{K}_{Y,\varepsilon}] \geq 1-\varepsilon/3$ and $\mathbb{P}_{\mathsf{Q}_i}[X_i\notin\mathcal{K}_{X,i,\varepsilon}] \geq 1-\varepsilon/3$, where Q_i is the marginal of X_i given by $Y\sim\mathsf{P}_Y$, $X_i\sim\mathsf{P}_i(\cdot\mid Y)$ (such sets exist because $\mathcal{X}_1,\mathcal{X}_2,\mathcal{Y}$ are Polish).

To check that $\mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2) \subset \Delta(\mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2)$ is closed, it suffices to show that it is sequentially closed (as $\Delta(\mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2)$ is Polish). To this end, consider any sequence $\mu_n \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2)$ such that $\mu_n \overset{\text{weak}}{\to} \mu \in \Delta(\mathcal{Y}\times\mathcal{X}_1\times\mathcal{X}_2)$ in the weak topology. By definition, this means that for any $i\in\{1,2\}$ and any continuous and bounded $f_i : \mathcal{Y}\times\mathcal{X}_i \to \mathbb{R}$,

$$\lim_{n\to\infty}\mathbb{E}_{\mu_n} f_i(Y,X_i) = \mathbb{E}_\mu f_i(Y,X_i).$$

For all $\mu_n \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2)$, $\mathbb{E}_{\mu_n} f_i(Y,X_i) = \mathbb{E}_{Y\sim\nu_Y}\mathbb{E}_{X_i\sim\nu_i(\cdot|Y_i)} f_i(Y,X_i)$. Thus,

$$\mathbb{E}_\mu f_i(Y,X_i) = \mathbb{E}_{Y\sim\nu_Y}\mathbb{E}_{X_i\sim\nu_i(\cdot|Y_i)} f_i(Y,X_i), \qquad \text{for all continuous, bounded } f_i : \mathcal{Y}\times\mathcal{X}\to\mathbb{R}.$$

Hence, the marginal distribution of (Y,X_i) under μ must be equal to that of $(Y\sim\mathsf{P}_Y, X_i\sim\mathsf{Q}_i(\cdot\mid Y))$ for $i\in\{1,2\}$, which means $\mu\in\mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2)$. □

By compactness, there exists (passing to a subsequence if necessary) a $\mu_\star \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2)$ such that $\mu_{\star,n} \overset{\text{weak}}{\to} \mu_\star$ in the weak topology. Then, as ϕ_m is continuous and bounded, it follows that for all m,

$$\begin{aligned}
\mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi_m(X_1,X_2)] &= \limsup_{n\to\infty}\mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,n}}[\phi_m(X_1,X_2)] && (\mu_{\star,n}\overset{\text{weak}}{\to}\mu_\star)\\
&\leq \limsup_{n\to\infty}\mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,n}}[\phi_n(X_1,X_2)] && (\phi_m\leq\phi_n \text{ for } n\geq m)\\
&= \limsup_{n\to\infty}\mathbb{E}_Y\psi_n(Y) && (\text{(C.3)})\\
&= \mathbb{E}_Y\lim_{n\to\infty}\psi_n(Y) && (\text{Monotone Convergence})\\
&= \mathbb{E}_Y\psi(Y). && (\text{Claim C.9})
\end{aligned}$$

Thus, by the monotone convergence theorem,

$$\begin{aligned}
\mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi(X_1,X_2)] &= \mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}\left[\lim_{m\to\infty}\phi_m(X_1,X_2)\right]\\
&= \lim_{m\to\infty}\mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi_m(X_1,X_2)]\\
&\leq \lim_{m\to\infty}\mathbb{E}_Y\psi(Y) = \mathbb{E}_Y\psi(Y).
\end{aligned}$$

Similarly, repeating some of the above steps,

$$\begin{aligned}
\mathbb{E}_Y \psi(Y) &= \limsup_{n\to\infty} \mathbb{E}_Y \psi_n(Y) \\
&= \limsup_{n\to\infty} \mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,n}}[\phi_n(X_1,X_2)] \\
&\le \limsup_{n\to\infty} \mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\phi_n(X_1,X_2)] && (\mu_{\star,n} \text{ is the optimal coupling for } \phi_n) \\
&\le \mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\lim_{n\to\infty} \phi_n(X_1,X_2)] && \text{(monotone convergence)} \\
&\le \mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\phi(X_1,X_2)].
\end{aligned}$$

Hence, $\mathbb{E}_Y \psi(Y) \le \liminf_{m\ge1} \mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\phi(X_1,X_2)]$. By assumption, $\phi(X_1,X_2)$ is lower semicontinuous and bounded from below. Thus, the Portmanteau theorem (Durrett, 2019) implies that, as $\mu_{\star,m} \overset{\text{weak}}{\to} \mu_\star$, $\liminf_{m\ge1} \mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\phi(X_1,X_2)] = \mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi(X_1,X_2)]$. Hence, $\mathbb{E}_Y \psi(Y) \le \mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi(X_1,X_2)]$, proving the reverse inequality.

Proof of the last statement. To prove the last statement, we observe that if $\mu_\star \in \mathscr{C}_{\mathsf{P}_Y}(\mathsf{Q}_1,\mathsf{Q}_2)$ then there exists a version of $(\mu_\star)_{X,X'|Y}$ that is a regular conditional probability and such that for almost every y it holds that $(\mu_\star)_{X,X'|y} \in \mathscr{C}(\mathsf{Q}_1(y),\mathsf{Q}_2(y))$. Indeed, the existence of a version that is a regular conditional probability is immediate by Theorem 3. To see that this version is a valid coupling of $\mathsf{Q}_1(y)$ and $\mathsf{Q}_2(y)$, observe that under $\mu_\star$, the joint law of $(X,Y) \sim \mathsf{Q}_1$ and thus the conditional distribution under $\mu_\star$ of $X|Y$ is determined up to sets of Q_1-measure 0. In particular, again by Theorem 3, there exists a regular conditional probablity that is a version of $(\mu_\star)_{X|y}$ and this must agree almost everywhere with $(\mathsf{Q}_1)_{X|y} = \mathsf{Q}_1(y)$. The same argument holds for X' and thus $(\mu_\star)_{X,X'|y} \in \mathscr{C}(\mathsf{Q}_1(y),\mathsf{Q}_2(y))$ for almost every y. Thus, by definition of ψ as an infimum, it holds for almost every y that

$$\psi(y) \le \mathbb{E}_{(X,X')\sim(\mu_\star)_{|Y}}[\phi(X,X')].$$

By the second claim of the proposition, we also have that

$$\mathbb{E}_{\mu_\star}[\phi(X_1,X_2)] = \mathbb{E}_{\mu_\star}[\psi(Y)].$$

Because the expectations are equal and one function is pointwise almost everywhere dominated by the other function, the two functions must be equal almost everywhere, concluding the proof. □

C.5. A simple union-bound recursion.

Finally, we also use the following version of the union bound extensively in our recursion proofs.

Lemma C.11. *For any event $\mathcal{E}$ and events $\mathcal{B}_1,\mathcal{B}_2,\dots,\mathcal{B}_H$, it holds that*

$$\mathbb{P}[(\mathcal{Q}\cap\bigcap_{h=1}^{H}\mathcal{B}_h)^c] \le \mathbb{P}[\mathcal{Q}^c] + \mathbb{P}\left[\exists h\in[H] \text{ s.t. } \left(\mathcal{Q}\cap\bigcap_{j=1}^{h-1}\mathcal{B}_j\cap\mathcal{B}_h^c\right) \text{ holds}\right]$$

Proof. Note that

$$\left(\mathcal{Q}\cap\bigcap_{h=1}^{H}\mathcal{B}_h\right)^c = \mathcal{Q}^c \cup \left(\mathcal{Q}\cap\left(\bigcap_{h=1}^{H}\mathcal{B}_h\right)^c\right) = \mathcal{Q}^c \cup \bigcup_{h=1}^{H} \mathcal{Q}\cap\mathcal{B}_h\cap\bigcap_{j=1}^{h-1}\mathcal{B}_j.$$

The result follows by a union bound. □

D. Warmup: Analysis Without Augmentation

In this section, we give a simplified analysis that replaces the smoothing kernels W_σ with the assumption that the learner policy $\hat{\pi}$ is already total variation continuous. The removal of the coupling kernel makes the coupling construction considerably simpler while still communicating some intuition for the full proof in Appendix E.

Throughout this section, we make the following assumptions on the state and action spaces, along with their associated metrics:

Assumption D.1. We assume that $\mathcal{S}$ and $\mathcal{A}$ are Polish spaces. This means they are metrizable, but we do not annotate their metrics because, e.g. the metric on $\mathcal{S}$ may be other than $\mathsf{d}_{\mathcal{S}}$. We further assume that

- $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}$ are pseudometrics and Borel measurable function from $\mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$
- For any $\varepsilon \geq 0$, the set $\{(\mathsf{a}, \mathsf{a}') \in \mathcal{A} \times \mathcal{A} : \mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') > \varepsilon\}$ is an open subset of $\mathcal{A} \times \mathcal{A}$; i.e. $\mathsf{d}_{\mathcal{A}}(\cdot, \cdot)$ is lower semicontinuous. In particular, this means $\mathsf{d}_{\mathcal{A}}$ is a Borel measurable function.

Recall the definitions of total variation continuity (TVC) and input-stability in Section 4. The main result of this section is as follows.

Proposition D.1. *Let $\pi^\star$ be input-stable w.r.t. $(\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathcal{A}})$ and let $\hat{\pi}$ be γ-TVC. Then, for all $\varepsilon > 0$, $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star) \leq H\gamma(\varepsilon) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star) \parallel \pi^\star(\mathsf{s}_h^\star))$.*

Proof. The key to the proof is to construct an appropriate "interpolating sequence" of actions $\hat{\mathsf{a}}_{1:H}^{\mathrm{inter}}$ to which we couple both $(\mathsf{s}_{1:H+1}^\star, \mathsf{a}_{1:H}^\star)$ and $(\hat{\mathsf{s}}_{1:H+1}, \hat{\mathsf{a}}_{1:H})$. This technique will be used in a significantly more sophisticated manner in the sequel to prove the analogous result with smoothing.

Let $\mathcal{F}_h$ denote the σ-algebra generated by $(\mathsf{s}_{1:h}^\star, \mathsf{a}_{1:h}^\star)$, $(\hat{\mathsf{s}}_{1:h}, \hat{\mathsf{a}}_{1:h})$, and $\hat{\mathsf{a}}_{1:h}^{\mathrm{inter}}$, and let $\mathcal{F}_0$ denote the σ-algebra generated by $\mathsf{s}_1^\star, \hat{\mathsf{s}}_1$. We construct couplings of the following form:

- The initial states are generated as $\mathsf{s}_1^\star = \hat{\mathsf{s}}_1 \sim \mathsf{P}_{\mathrm{init}}$.
- The dynamics are determined by F_h:
$$\mathsf{s}_{h+1}^\star = F_h(\mathsf{s}_h^\star, \mathsf{a}_h^\star), \quad \hat{\mathsf{s}}_{h+1} = F_h(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h) \tag{D.1}$$
In particular, $\mathsf{s}_{h+1}^\star, \hat{\mathsf{s}}_{1:h+1}$ are $\mathcal{F}_h$ measurable.
- The conditional distributions of the primitive controllers satisfy the following
$$\mathsf{a}_h^\star \mid \mathcal{F}_{h-1} \sim \pi_h^\star(\mathsf{s}_h^\star), \quad \hat{\mathsf{a}}_{h-1} \mid \mathcal{F}_{h-1} \sim \hat{\pi}_h(\hat{\mathsf{s}}_h), \quad \hat{\mathsf{a}}_h^{\mathrm{inter}} \mid \mathcal{F}_h \sim \hat{\pi}_h(\mathsf{s}_h^\star). \tag{D.2}$$

Note that if μ satisfies the above construction, then $(\mathsf{s}_{1:H+1}^\star, \mathsf{s}_{1:H}^\star) \sim \mathsf{D}_{\pi^\star}$ and $(\hat{\mathsf{s}}_{1:H+1}, \hat{\mathsf{a}}_{1:H}) \sim \mathsf{D}_{\hat{\pi}}$.

Specifying the rest of the coupling. It remains to specify the coupling of the terms in (D.2). We establish our coupling sequentially. Let $\mu^{(0)}$ denote the coupling of $\hat{\mathsf{s}}_1 = \mathsf{s}_1^\star \sim \mathsf{P}_{\mathrm{init}}$.

Assume we have constructed the coupling up to state $h-1$.For ease, let Y_{h-1} denote the random variable corresponding to $(\mathsf{s}_{1:h}^\star, \hat{\mathsf{s}}_{1:h}, \mathsf{a}_{1:h-1}^\star, \hat{\mathsf{a}}_{1:h-1}, \hat{\mathsf{a}}_{1:h-1}^{\mathrm{inter}})$; note that Y_{h-1} is $\mathcal{F}_{h-1}$-measurable (as $\hat{\mathsf{s}}_h, \mathsf{s}_h^\star$ are determined by the dynamics (D.1)). Observe that, by the assumption of $\hat{\pi}_h$ being TVC, it holds that
$$\mathsf{TV}(\mathbb{P}_{\hat{\mathsf{a}}_h|Y_{h-1}}, \mathbb{P}_{\hat{\mathsf{a}}_h^{\mathrm{inter}}|Y_{h-1}}) \leq \gamma(\mathsf{d}_{\mathrm{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star)).$$

Thus by Lemma C.1, there exists a coupling $\mu_1^{(h)}$ between $Y_{h-1}, \hat{\mathsf{a}}_h, \hat{\mathsf{a}}_h^{\mathrm{inter}}$, with $Y_{h-1} \sim \mu^{(h-1)}$ such that it holds that
$$\mathbb{P}[\hat{\mathsf{a}}_h \neq \hat{\mathsf{a}}_h^{\mathrm{inter}}] \leq \mathbb{E}_{\mu^{(h-1)}}[\gamma(\mathsf{d}_{\mathrm{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star))].$$

Similarly by Proposition C.3, there is a coupling $\mu_2^{(h)}$ of $Y_{h-1}, \hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star$ such that
$$\mathbb{P}_{\mu_2^{(h)}}[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) \geq \varepsilon] \leq \mathbb{E}_{\mathsf{s}_h^\star \sim \mu^{(h-1)}}[\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))].$$

By the gluing lemma Lemma C.2 and a union bound, we may construct a coupling $\mu^{(h)}$ of $Y_h, \hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star, \hat{\mathsf{a}}_h$ such that (almost surely),
$$\begin{aligned}
&\mathbb{P}_{\mu^{(h)}}[\{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) \geq \varepsilon\} \cup \{\hat{\mathsf{a}}_h \neq \hat{\mathsf{a}}_h^{\mathrm{inter}}\} \mid \mathcal{F}_{h-1}] \\
&= \mathbb{P}_{\mu^{(h)}}[\{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) \geq \varepsilon\} \cup \{\hat{\mathsf{a}}_h \neq \hat{\mathsf{a}}_h^{\mathrm{inter}}\} \mid Y_{h-1}] \\
&\quad \leq \gamma(\mathsf{d}_{\mathrm{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star))] + \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))
\end{aligned} \tag{D.3}$$

Thus inductively, we may continue this construction for $h \leq H$ and let $\mu = \mu^{(H)}$.

Concluding the proof. Define the event $\mathcal{B}_h := \{\mathsf{d}_{\mathcal{A}}(\mathsf{a}_h, \hat{\mathsf{a}}_h^{\text{inter}}) \le \varepsilon\}$ and $\mathcal{C}_h = \{\hat{\mathsf{a}}_h^{\text{inter}} = \hat{\mathsf{a}}_h\}$. Then, by Lemma C.11

$$\mathbb{P}_\mu\left[(\bigcap_{h=1}^{H} \mathcal{B}_h \cap \mathcal{C}_h)^c\right] \le \sum_{h=1}^{H} \mathbb{P}_\mu\left[(\bigcap_{j=1}^{h-1} \mathcal{B}_j \cap \mathcal{C}_j) \cap (\mathcal{B}_h^c \cup \mathcal{C}_h^c)\right]. \tag{D.4}$$

Note first that $(\bigcap_{j=1}^{h-1} \mathcal{B}_j \cap \mathcal{C}_j)$ is $\mathcal{F}_{h-1}$ measurable. On this event, input stability at $\hat{\mathsf{a}}_j^{\text{inter}} = \hat{\mathsf{a}}_j$, $1 \le j \le h-1$, implies that

$$\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h^\star, \hat{\mathsf{s}}_h) \le \varepsilon.$$

Thus, (D.3) implies that

$$\begin{aligned}
\mathbb{P}_\mu\left[(\bigcap_{j=1}^{h-1} \mathcal{B}_j \cap \mathcal{C}_j) \cap (\mathcal{B}_h^c \cup \mathcal{C}_h^c)\right] &\le \mathbb{E}_\mu[\gamma(\mathsf{d}_{\text{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star))\mathbf{I}\{\mathsf{d}_{\text{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star) \le \varepsilon\} + \mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star)) \mid \mathcal{F}_{h-1}] \\
&\le \gamma(\varepsilon) + \mathbb{E}_\mu\left[\mathbb{E}_\mu[\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star)) \mid \mathcal{F}_{h-1}]\right] \\
&= \gamma(\varepsilon) + \mathbb{E}_\mu[\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))] \\
&= \gamma(\varepsilon) + \mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star}\mathbb{E}_\mu[\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))],
\end{aligned}$$

where the first equality follows from the tower rule for conditional expectations and the second follows because $\mathsf{s}_h^\star \sim \mathsf{P}_h^\star$ under μ. Summing and applying (D.4) implies that

$$\mathbb{P}_\mu\left[(\bigcap_{h=1}^{H} \mathcal{B}_h \cap \mathcal{C}_h)^c\right] \le H\gamma(\varepsilon) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star}[\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))].$$

Again, invoking input stability and the definitions $\mathcal{B}_h := \{\mathsf{d}_{\mathcal{A}}(\mathsf{a}_h, \hat{\mathsf{a}}_h^{\text{inter}}) \le \varepsilon\}$ and $\mathcal{C}_h = \{\hat{\mathsf{a}}_h^{\text{inter}} = \hat{\mathsf{a}}_h\}$, $(\bigcap_{h=1}^{H} \mathcal{B}_h \cap \mathcal{C}_h)^c$ implies that

$$\max_{1 \le h \le H} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}_{h+1}^\star, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}_h^\star, \hat{\mathsf{a}}_h)\} \le \varepsilon.$$

This concludes the proof.

□

E. Imitation in the Composite MDP

In this section, we prove our imitation guarantees in the composite MDP under the full generality of data augmentation. The majority of this section is devoted to proving a more general version of Theorem 2 that applies to vectorized notions of distance and helps tighten our bounds when instantiated in the control setting. In Appendix E.1, we introduce some notation and state our most general result, Theorem 4. We then proceed to show that Theorem 2 follows from Theorem 4 and in Appendix E.2, we provide a detailed and rigorous proof of the main result. In Appendix E.3, we show that the more general Theorem 4 impiles Theorem 2 from the text.

Throughout, we also assume $\mathcal{S}$ admits a direct decomposition. This is useful to capture the fact that we only apply smoothing on the $\boldsymbol{\rho}_{\mathrm{m},h}$ coordinates (memory chunk), not the full trajectory chunk $\boldsymbol{\rho}_{\mathrm{c},h}$.

Definition E.1 (Direct Decomposition)**.** Let $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ is a direct decomposition. We let $\phi_{\mathcal{Z}}$ and $\phi_{\mathcal{V}}$ denote projections onto the $\mathcal{Z}$ and $\mathcal{V}$ components, respectively. We say that the $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ is *compatible* with the dynamics if $F_h((\mathsf{z}, \mathsf{v}), \mathsf{a}) = F_h((\mathsf{z}, \mathsf{v}'), \mathsf{a})$ for all $\mathsf{v}, \mathsf{v}' \in \mathcal{V}$ and $\mathsf{z} \in \mathcal{Z}$, and *compatible* with policy π if $\pi_h((\mathsf{z}, \mathsf{v}), \mathsf{a}) = \pi_h((\mathsf{z}, \mathsf{v}'), \mathsf{a})$.; we define compatibility of a kernel W and of a pseudometric $\mathsf{d}(\cdot, \cdot) : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\ge 0}$ with $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ similarly.

We emphasize that compatibility of dynamics with a direct decomposition does not make v irrelevant because $\mathsf{d}_{\mathcal{S}}$ still depends on v. For the purposes of the instantiation for control in the following appendix, we wish to control the imitation gaps on distances that do depend on v_h, even though v_h does not figure directly into the dynamics. Note that as defined, v_h does depend on the dynamics up until time $h-1$ and thus it is necessary to deal with this component in order to provide guarantees in $\mathsf{d}_{\mathcal{S}}$.

E.1. A generalization of Theorem 2

We now state a generalization of Theorem 2, which replaces a single distance by a vector of distances of dimension K; this will be useful for our instantiation of the composite MDP as a chunked control system in our final application (in particular, for deriving a bound on $\mathcal{L}_{\mathrm{fin},\varepsilon}$). It also showcases the most general structure accomodated by our proof technique.

We begin by defining some notation:

- Let $K \in \mathbb{N}$ denote a dimension
- Let $\vec{\varepsilon} \in \mathbb{R}^K_{\geq 0}$ denote a vector of tolerances
- Let $\vec{\mathsf{d}}_{\mathcal{S}}(\cdot,\cdot)$ denote a vector of pseudometrics $\mathsf{d}_{\mathcal{S},i}$ on $\mathcal{S}$
- Let $\vec{\mathsf{d}}_{\mathcal{A}}$ denote a vector of non-negative functions $\mathsf{d}_{\mathcal{A},i} : \mathcal{A}^2 \to \mathbb{R}_{\geq 0}$, not necessarily pseudometrics.
- Let $\preceq$ denote vector wise inequality, and let the symbols $\wedge$ and $\vee$ be generalized to denote entrywise minima and maxima. Similarly, addition of vectors is coordinate wise with scalars assumed to be broadcast appropriately.
- We let $\mathsf{d}_{\mathcal{S},1} = \mathsf{d}_{\mathrm{TVC}}$ denote the metric we consider for evaluating total variation distance.

We generalize We assume the following measure-theoretic regularity conditions, generalizing Assumption D.1 as follows.

Assumption E.1. We assume that $\mathcal{S}$ and $\mathcal{A}$ are Polish spaces. This means they are metrizable, but we do not annotate their metrics because, e.g. the metric on $\mathcal{S}$ may be other than $\mathsf{d}_{\mathcal{S}}$. We further assume that

- $\mathsf{d}_{\mathcal{S},i}$ is a pseudometric and Borel measurable function from $\mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$.
- For any $\varepsilon \geq 0$, the set $\{(\mathsf{a},\mathsf{a}') \in \mathcal{A} \times \mathcal{A} : \mathsf{d}_{\mathcal{A},i}(\mathsf{a},\mathsf{a}') > \varepsilon\}$ is an open subset of $\mathcal{A} \times \mathcal{A}$; i.e. $\mathsf{d}_{\mathcal{A},i}(\cdot,\cdot)$ is lower semicontinuous. In particular, this means $\mathsf{d}_{\mathcal{A},i}$ is a Borel measurable function. Note that this implies that the

$$\{(\mathsf{a},\mathsf{a}') \in \mathcal{A} \times \mathcal{A} : \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a},\mathsf{a}') \not\preceq \vec{\varepsilon}\}.$$

is closed and thus measurable.

Note that the above assumption is the natural vectorized generalization of Assumption D.1. Next, we define vector versions of our imitation errors.

Definition E.2 (Imitation Errors, vector version)**.** Given error parameter $\vec{\varepsilon} \in \mathbb{R}^K_{\geq 0}$, define

- The ***vector joint-error***

$$\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi} \parallel \pi^\star) := \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\exists h \in [H] : \vec{\mathsf{d}}_{\mathcal{S}}(\hat{\mathsf{s}}_{h+1},\mathsf{s}^\star_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^\star_h,\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}\right],$$

where the infimum is over trajectory couplings $((\hat{\mathsf{s}}_{1:H+1},\hat{\mathsf{a}}_{1:H}),(\mathsf{s}^\star_{1:H+1},\mathsf{a}^\star_{1:H})) \sim \mu_1 \in \mathscr{C}(\mathsf{D}_{\hat{\pi}},\mathsf{D}_{\pi^\star})$ satisfying $\mathbb{P}_{\mu_1}[\hat{\mathsf{s}}_1 = \mathsf{s}^\star_1] = 1$.

- The ***vector marginal error***

$$\vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}}(\hat{\pi} \parallel \pi^\star) := \max_{h\in[H]} \max\left\{\inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\vec{\mathsf{d}}_{\mathcal{S}}(\hat{\mathsf{s}}_{h+1},\mathsf{s}^\star_{h+1}) \not\preceq \vec{\varepsilon}\right], \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^\star_h,\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}\right]\right\}$$

the same as the to joint-gap, with the "$\max$" outside the probability and infimum over couplings.

- The ***vector-wise one-step error***

$$\vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_h(\mathsf{s}) \parallel \pi^\star_h(\mathsf{s})) := \inf_{\mu_2} \mathbb{P}_{\mu_2}\left[\vec{\mathsf{d}}_{\mathcal{A}}(\hat{\mathsf{a}}_h,\mathsf{a}^\star_h) \not\preceq \vec{\varepsilon}\right],$$

where the infimum is over $(\mathsf{a}^\star_h,\hat{\mathsf{a}}_h) \sim \mu_2 \in \mathscr{C}(\hat{\boldsymbol{\pi}}_h(\mathsf{s}),\boldsymbol{\pi}^\star_h(\mathsf{s}))$.

We now describe input stability.

Definition E.3 (Input-Stability, vector version)**.** A trajectory $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ is ***input-stable*** w.r.t. $(\vec{\mathsf{d}}_{\mathcal{S}}, \vec{\mathsf{d}}_{\mathcal{A}})$ if all sequences $\mathsf{s}'_1 = \mathsf{s}_1$ and $\mathsf{s}'_{h+1} = F_h(\mathsf{s}'_h, \mathsf{a}'_h)$ satisfy

$$\mathsf{d}_{\mathcal{S},i}(\mathsf{s}'_{h+1}, \mathsf{s}_{h+1}) \le \max_{1\le j\le h} \mathsf{d}_{\mathcal{A},i}\left(\mathsf{a}'_j, \mathsf{a}_j\right), \quad \forall h \in [H], i \in [K]$$

Finally, define input process stability. A slight technicality is that, in our instantiation, $\pi^\star$ is taken to be a suitable regular condition probability of the joint distribution $\mathcal{D}_{\mathrm{exp}}$ of expert trajectories. This means that $\pi^\star$ can only really satisfy desired regularity conditions on states visited with positive probabiliy by $\mathcal{D}_{\mathrm{exp}}$. We address this subtlety by considering the following definition generalizing Definition 4.5 in the body. We also restrict the kernels under consideration to those which produce distributions *absolutely continuous* (Definition C.4) with respect to $\mathsf{P}^\star_h$, and denoted with the $\ll$ comparator. More specifically, we only care about absolute continuity under the projections onto the $\mathcal{Z}$ component of $\mathcal{S}$.

Definition E.4 (Input & Process Stability, vector version)**.** Let $p_{\textsc{ips}} \in (0,1)$, $\vec{\gamma}_{\mathrm{IPS}} = (\gamma_{\mathrm{IPS},i})_{1\le i\le K}$ be a collection non-decreasing maps $\gamma_{\mathrm{IPS},i} : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$, let $\mathsf{d}_{\textsc{ips}} : \mathcal{S}\times\mathcal{S} \to \mathbb{R}$ be a pseudometric (possibly other than any of the $\mathsf{d}_{\mathcal{S},i}$), and $r_{\textsc{ips}} > 0$. We say a policy $\pi^\star$ is $(\vec{\gamma}_{\mathrm{IPS}}, \mathsf{d}_{\textsc{ips}}, r_{\textsc{ips}}, p_{\textsc{ips}})$*-(vectorwise-input-&-process stable (vIPS)* if the following holds for any $r \in [0, r_{\textsc{ips}}]$:

Consider any sequence of kernels $\mathsf{W}_h : \mathcal{S} \to \Delta(\mathcal{S})$, $1 \le h \le H$, satisfying

$$\forall h, \mathsf{s} \in \mathcal{S}: \quad \mathbb{P}_{\tilde{\mathsf{s}}\sim \mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\textsc{ips}}(\tilde{\mathsf{s}}, \mathsf{s}) \le r] = 1, \quad \phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathcal{Z}} \circ \mathsf{P}^\star_h. \tag{E.1}$$

Define a process $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h)$, $\mathsf{a}_h \sim \pi_h(\tilde{\mathsf{s}}_h)$, and $\mathsf{s}_{h+1} := F_h(\mathsf{s}_h, \mathsf{a}_h)$. Then, with probability at least $1 - p_{\textsc{ips}}$,

(a) the sequence $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ is input-stable w.r.t $(\vec{\mathsf{d}}_{\mathcal{S}}, \vec{\mathsf{d}}_{\mathcal{A}})$ (as defined by Definition E.3).

(b) $\max_{h\in[H]} \mathsf{d}_{\mathcal{S},i}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \le \gamma_{\mathrm{IPS},i}(r)$.

We can now state our desired generalization.

Theorem 4. *Suppose that there*

(a) $\pi^\star$ *is* $(\vec{\gamma}_{\mathrm{IPS}}, \mathsf{d}_{\textsc{ips}}, r_{\textsc{ips}}, p_{\textsc{ips}})$*-vector IPS in the sense of Definition E.4.*

(b) There is a direct decomposition of $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$*, which associated projection maps* $\phi_{\mathcal{Z}}$ *and* $\phi_{\mathcal{V}}$*, and which is compatible with the dynamics, and policies* $\pi^\star$*,* $\hat{\pi}$*, and smoothing kernel* W_σ*, and* $\mathsf{d}_{\textsc{ips}}$*.*

(c) $\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma$ *is* γ_σ*-TVC with respect to the pseudometric* $\mathsf{d}_{\textsc{tvc}} = \mathsf{d}_{\mathcal{S},1}$*.*

Let $\hat{\pi}_\sigma$ *be any policy which is* $\hat{\gamma}$*-TVC, also w.r.t.* $\mathsf{d}_{\textsc{tvc}} = \mathsf{d}_{\mathcal{S},1}$*. Finally, let* $\vec{\varepsilon} \in \mathbb{R}^K_{\ge 0}$*,* $r \in (0, \frac{1}{2} r_{\textsc{ips}}]$*, and define*

$$p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}'\sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\textsc{ips}}(\mathsf{s}', \mathsf{s}) > r], \quad \vec{\varepsilon}_{\mathrm{marg}} := \vec{\varepsilon} + \vec{\gamma}_{\mathrm{IPS}}(2r).$$

Then,

- *For any policy* $\hat{\pi}$*, both* $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft})$ *and* $\vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star)$ *are upper bounded by*

$$p_{\textsc{ips}} + H(2p_r + \hat{\gamma}(\vec{\varepsilon}_1) + (\hat{\gamma} + \gamma_\sigma) \circ \gamma_{\textsc{ips},1}(2r)) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h)) \tag{E.2}$$

- *In the special case where* $\hat{\pi}_\sigma = \hat{\pi} \circ \mathsf{W}_\sigma$*, we can take* $\hat{\gamma} = \gamma_\sigma$*, and obtain that* $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft})$ *and* $\vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star)$ *are upper bounded by*

$$p_{\textsc{ips}} + H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\textsc{ips},1}(2r)\}\right) + \textstyle\sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h)). \tag{E.3}$$

We note that Theorem 2 is a special case of Theorem 4 and prove the former assuming the latter here at the end of the section.

E.2. Proof of Theorem 4

In this section, we prove Theorem 4. We begin with an intuitive overview of the proof and partially construct the relevant intermediate trajectories used to define our coupling in Appendix E.2.1. In Appendix E.2.2, we prove several prerequisite properties of the construction given in Appendix E.2.1. Finally, in Appendix E.2.3 we formally construct the coupling and rigorously prove Theorem 4.

E.2.1. Proof Overview and Coupling Construction

The proof proceeds by constucting a sophisticated coupling between the law of a trajectory evolving according to $\hat{\pi}$ and a trajectory evolving according to $\pi^\star_{\circlearrowleft}$ by introducing several intermediate sequences of composite states and composite actions.

We partially specify this coupling below and formally construct it in Appendix E.2.3. Our construction is recursive and relies on the input and process stability as well as total variation continuity to show that if the trajectories generated by $\pi^\star_{\circlearrowleft}$ and $\hat{\pi}$ are close in $\vec{\mathsf{d}}_{\mathrm{os},\tilde{\varepsilon}}$ evaluated on states at step h, then they will remain close at step $h+1$. There are a number of technical subtelties involved, especially those of a measure-theoretic nature, but much of the inuition can be gleaned from the following partial specification of the coupling μ over composite-state $(\hat{\mathsf{s}}_{1:H}, \mathsf{s}^{\circlearrowleft}_{1:H}, \mathsf{s}^{\mathrm{tel}}_{1:H}, \tilde{\mathsf{s}}^{\mathrm{tel}}_{1:H}) \subset \mathcal{S}$, composite-actions $(\mathsf{a}^{\circlearrowleft}_{1:H}, \hat{\mathsf{a}}_{1:h}, \mathsf{a}^{\mathrm{tel}}_{1:H}) \subset \mathcal{K}$ and interpolating composite-actions, $(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_{1:H}, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_{1:H}) \subset \mathcal{A}$.

To define the construction, we define the probability kernels corresponding to the replica and deconvolution policies. Note that these are slightly different from the definitions in the body due to the use of the direct decomposition; the intuition is the same, however.

Definition E.5 (Replica and Deconvolution Kernels)**.** Let $\mathsf{P}^{\mathrm{proj}}_{\mathrm{aug},h}$ denote the joint distribution over $(\mathsf{z}^\star_h, \mathsf{s}^\star_h, \tilde{\mathsf{z}}^\star_h, \mathsf{a}^\star_h)$ under the generative process

$$\mathsf{s}^\star_h \sim \mathsf{P}^\star_h, \quad \mathsf{a}^\star_h \sim \pi^\star_h(\mathsf{s}^\star_h), \quad \mathsf{z}^\star_h = \phi_{\mathcal{Z}}(\mathsf{s}^\star_h), \quad \tilde{\mathsf{z}}^\star_h \sim \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}^\star_h)$$

For $\mathsf{z} \in \mathcal{Z}$, let $\mathsf{W}^\star_{\mathrm{dec},\mathcal{Z},h}(\mathsf{z})$ denote the distribution of $\mathsf{z}^\star_h$ conditioned on $\tilde{\mathsf{z}}^\star_h = \mathsf{z}$, under $\mathsf{P}^{\mathrm{proj}}_{\mathrm{aug},h}$. Given $\mathsf{s} = (\mathsf{z}, \mathsf{v})$, define

$$\begin{aligned}
\mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}) &= \mathsf{W}^\star_{\mathrm{dec},\mathcal{Z},h}(\phi_{\mathcal{Z}}(\mathsf{s})) \otimes \delta_{\phi_{\mathcal{V}}(\mathsf{s})}, \\
\mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}) &= \mathsf{W}^\star_{\mathrm{dec},h} \circ (\mathsf{W}_\sigma(\phi_{\mathcal{Z}}(\mathsf{s})) \otimes \delta_{\phi_{\mathcal{V}}(\mathsf{s})}) = (\mathsf{W}^\star_{\mathrm{dec},\mathcal{Z},h} \circ \mathsf{W}_\sigma(\phi_{\mathcal{Z}}(\mathsf{s}))) \otimes \delta_{\phi_{\mathcal{V}}(\mathsf{s})}.
\end{aligned}$$

where we recall the dirac-delta δ. Equivalently, $\mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s})$ denotes the conditional sequence of $(\tilde{\mathsf{z}}, \mathsf{v})$, where $\mathsf{v} = \phi_{\mathcal{V}}(\mathsf{s})$, and $\tilde{\mathsf{z}} \sim \mathsf{W}^\star_{\mathrm{dec},\mathcal{Z},h}(\mathsf{s})$; $\mathsf{W}^\star_{\circlearrowleft,h}$ can be expressed similarly.

We remark that $\mathsf{W}^\star_{\mathrm{dec},h}$ and $\mathsf{W}^\star_{\circlearrowleft,h}$ are both kernels and by Theorem 3, we may assume that the joint distribution over $(\mathsf{s}^\star_h, \tilde{\mathsf{s}}^{\mathrm{tel}}_h)$ admits a regular conditional probability and thus these constructions are well-defined.

Remark E.1. Note that the kernels $\mathsf{W}^\star_{\mathrm{dec},h}$ and $\mathsf{W}^\star_{\circlearrowleft,h}$ are compatible with the decomposition $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ by construction. Moreover, note that if $\mathsf{s} = (\mathsf{z}, \mathsf{v})$, $\phi_{\mathcal{V}} \circ \mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}) = \phi_{\mathcal{V}} \circ \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s})$ is the dirac-delta distribution supported on v.

Lemma E.1. *Under our the assumption that $\pi^\star$ and W_σ are compatible with the direct decomposition,*

$$\pi^\star_{\mathrm{dec},h}(\mathsf{s}) = \pi^\star \circ \mathsf{W}^\star_{\mathrm{dec},h}, \quad \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}) = \pi^\star \circ \mathsf{W}^\star_{\circlearrowleft,h}$$

Proof. This follows imediately because $\pi^\star$ and W_σ are compatile with the direct decomposition, and by the definition of Definition 4.4. □

A template for the coupling. Our couplings are partially specified by the following generative process, and what remains unspecified are couplings between random variables at each each step h. In what follows, let $\mathcal{F}_0$ denote the σ-algebra generated by $\hat{\mathsf{s}}_1 = \mathsf{s}^{\circlearrowleft}_1 = \mathsf{s}^{\mathrm{tel}}_1$. Let $\mathcal{F}_h$ denote the sigma-algebra generated by $(\hat{\mathsf{s}}_{1:h}, \mathsf{s}^{\circlearrowleft}_{1:h}, \mathsf{s}^{\mathrm{tel}}_{1:h})$, $(\mathsf{a}^{\circlearrowleft}_{1:h}, \tilde{\mathsf{s}}^{\circlearrowleft}_{1:h}, \tilde{\mathsf{s}}^{\mathrm{tel}}_{1:h}, \mathsf{a}^{\mathrm{tel}}_{1:h}, \hat{\mathsf{a}}_{1:h})$, and $(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_{1:h}, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_{1:h})$.

- The initial states are drawn as

$$\hat{\mathsf{s}}_1 = \mathsf{s}^{\circlearrowleft}_1 = \mathsf{s}^{\mathrm{tel}}_1 \sim \mathsf{P}_{\mathrm{init}}.$$

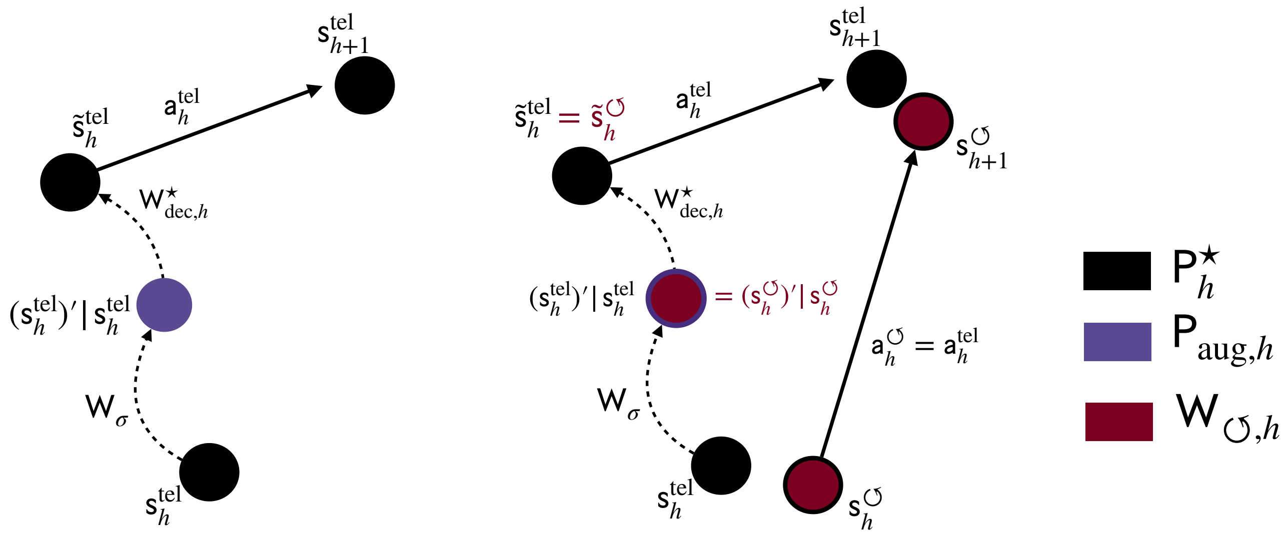


Figure 5. Graphical illustration of the coupling, in the special case where $\mathcal{Z} = \mathcal{S}$ for simplicity. **On the left** is the teleporting sequence, with $\tilde{\mathsf{s}}^{\mathrm{tel}} \sim \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}^{\mathrm{tel}}_h) = \mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{W}_\sigma(\mathsf{s}^{\mathrm{tel}}_h)$. We represent the teleporting explicitly by noising $\mathsf{s}^{\mathrm{tel}}_h$ to become $(\mathsf{s}^{\mathrm{tel}}_h)'$ by applying W_σ and then applying $\mathsf{W}^\star_{\mathrm{dec},h}$ to complete the "teleporting" to $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$. We then apply $\mathsf{a}^{\mathrm{tel}}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)$, and continue onto $\mathsf{s}^{\mathrm{tel}}_{h+1}$ *from the teleported state* $\tilde{\mathsf{s}}^{\mathrm{tel}}_{h+1}$. **On the right**, we illustrate the replica sequence next to the teleporting sequence. We start with $\mathsf{s}^{\circlearrowleft}_h$, which is close to $\mathsf{s}^{\mathrm{tel}}_h$ (a consequence of our proof). We then apply the replica kernel to achieve $\tilde{\mathsf{s}}^{\circlearrowleft}_h$. Our argument uses that $\mathsf{W}^\star_{\circlearrowleft,h} = \mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{W}_\sigma$ is TVC (a consequence of TVC of W_σ as shown in Lemma E.2). We depict this property pictorially: since W_σ is TVC and $\mathsf{s}^{\mathrm{tel}}_h$ and $\mathsf{s}^{\circlearrowleft}_h$ are close, we can couple things in such a way that, with good probability, $(\mathsf{s}^{\mathrm{tel}}_h)' \sim \mathsf{W}_\sigma(\mathsf{s}^{\mathrm{tel}}_h)$ and $(\mathsf{s}^{\circlearrowleft}_h)' \sim \mathsf{W}_\sigma(\mathsf{s}^{\circlearrowleft}_h)$ are equal. We then extend the coupling to that $\tilde{\mathsf{s}}^{\circlearrowleft}_h = \tilde{\mathsf{s}}^{\mathrm{tel}}_h$ on the event $\{(\mathsf{s}^{\mathrm{tel}}_h)' = (\mathsf{s}^{\circlearrowleft}_h)'\}$, both being drawn by applying $\mathsf{W}^\star_{\mathrm{dec},h}$ to both of $(\mathsf{s}^{\mathrm{tel}}_h)' = (\mathsf{s}^{\circlearrowleft}_h)'$. We extend the coupling once more so that $\mathsf{a}^{\mathrm{tel}}_h \sim \pi^\star(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)$ and $\mathsf{a}^{\circlearrowleft}_h \sim \pi^\star(\tilde{\mathsf{s}}^{\circlearrowleft}_h)$ are equal on this good probability event. Using our notion of stability, IPS, and the fact that $\mathsf{s}^{\circlearrowleft}_h$ and $\mathsf{s}^{\mathrm{tel}}_h$ are close, the good probability event on which $\mathsf{a}^{\mathrm{tel}}_h$ and $\mathsf{a}^{\circlearrowleft}_h$ are equal implies that $\mathsf{s}^{\circlearrowleft}_{h+1}$ remains close to $\mathsf{s}^{\mathrm{tel}}_{h+1}$. We remark that our actual analysis never explicitly computes the $(\cdot)'$-terms drawn from W_σ; rather, these terms appear implicitly in our definitions of $\mathsf{W}^\star_{\circlearrowleft,h}$ and the verification of its TVC property.

- The dynamics satisfy

$$\hat{\mathsf{s}}_{h+1} = F_h(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h), \quad \mathsf{s}^{\circlearrowleft}_{h+1} = F_h(\mathsf{s}^{\circlearrowleft}_h, \mathsf{a}^{\circlearrowleft}_h), \quad \mathsf{s}^{\mathrm{tel}}_{h+1} = F_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h, \mathsf{a}^{\mathrm{tel}}_h)$$

  Note that determinism of the dynamics implies that $\mathsf{s}^{\mathrm{tel}}_{h+1}$, $\mathsf{s}^{\circlearrowleft}_{h+1}$ and $\hat{\mathsf{s}}_{h+1}$ are $\mathcal{F}_h$-measurable.

- We generate

$$\begin{aligned}
&\tilde{\mathsf{s}}^{\circlearrowleft}_h \mid \mathcal{F}_{h-1} \sim \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}^{\circlearrowleft}_h), \quad \mathsf{a}^{\circlearrowleft}_h \mid \mathcal{F}_{h-1}, \tilde{\mathsf{s}}^{\circlearrowleft}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\circlearrowleft}_h),\\
&\tilde{\mathsf{s}}^{\mathrm{tel}}_h \mid \mathcal{F}_{h-1} \sim \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}^{\mathrm{tel}}_h), \quad \mathsf{a}^{\mathrm{tel}}_h \mid \mathcal{F}_{h-1}, \tilde{\mathsf{s}}^{\mathrm{tel}}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h).\\
&\hat{\mathsf{a}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_\sigma(\hat{\mathsf{s}}_h)
\end{aligned}$$

  Importantly, we note that, marginalizing over $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$ and $\tilde{\mathsf{s}}^{\circlearrowleft}_h$, respectively, $\mathsf{a}^{\mathrm{tel}}_h \mid \mathcal{F}_{h-1} \sim \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^{\mathrm{tel}})$ and $\mathsf{a}^{\circlearrowleft}_h \mid \mathcal{F}_{h-1} \sim \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^{\circlearrowleft}_h)$.

- Lastly, we select interpolating actions via

$$\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\circlearrowleft}_h), \qquad \hat{\mathsf{a}}^{\mathrm{tel},\mathrm{inter}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h)$$

We will say μ is "respects the construction" as shorthand to mean that μ obeys the above equations. The coupling is illustrated graphically in Figure 5. We now establish several key properties of the above constructions, separated into a subsection for the sake of clarity.

E.2.2. Properties of smoothing, deconvolution, and replicas.

In this section, we establish several useful properties of smoothed and replica policies. We begin by showing that smoothed policies are TVC.

Lemma E.2. *The following hold*

- *For any h, $\phi_{\mathcal{Z}} \circ \mathsf{W}^\star_{\circlearrowleft,h}$ and $\pi^\star_{\circlearrowleft\sigma,h}$ are γ_σ TVC.*
- *If π is any policy compatible with the direct decomposition $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ (in the sense of Definition E.1), then $\pi \circ \mathsf{W}_\sigma$ is γ_σ-TVC.*

Proof. We observe that $\phi_{\mathcal{Z}} \circ \mathsf{W}^\star_{\circlearrowleft,h} = \phi_{\mathcal{Z}} \circ \mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{W}_\sigma(\mathsf{s})$. Moreover, we observe $\mathsf{W}^\star_{\mathrm{dec},h}$ satisfies $\phi_{\mathcal{Z}} \circ \mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}) = \mathsf{W}^\star_{\mathrm{dec},\mathcal{Z},h} \circ \phi_{\mathcal{Z}}$, so that $\phi_{\mathcal{Z}} \circ \mathsf{W}^\star_{\circlearrowleft,h} = \mathsf{W}^\star_{\mathrm{dec},\mathcal{Z},h} \circ \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s})$. As $\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma$ is TVC, the first claim is a consequence of the data-processing inequality Lemma C.4. The second uses the fact that all listed objects involve composition of kernels with W_σ. $\square$

Next, we show that the replica construction preserves marginals.

Lemma E.3 (Marginal-Preservation)**.** *There exists a coupling $\mathbb{P}$ of $\mathsf{z}_h \sim \phi_{\mathcal{Z}} \circ \mathsf{P}^\star_h$, $\mathsf{z}'_h \sim \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{z}_h, \cdot)$ (where $(\cdot)$ denotes an irrelevant argument due to compatibility of W_σ with the direct decomposition), and $\tilde{\mathsf{z}}_h \sim \phi_{\mathcal{Z}} \circ \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{z}_h, \cdot)$ (again, $(\cdot)$ denotes an irrelevant argument) such that*

$$(\mathsf{z}_h, \mathsf{z}'_h) \overset{\mathrm{d}}{=} (\tilde{\mathsf{z}}_h, \mathsf{z}'_h).$$

In particular, for $\mathsf{s}^{\mathrm{tel}}_h$ and $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$ as in our construction, the marginal distributions of $\phi_{\mathcal{Z}}(\mathsf{s}^{\mathrm{tel}}_h)$ and $\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)$ are the same, where $\mathsf{s}^{\mathrm{tel}}_h \sim \mathsf{P}^\star_h$ and $\tilde{\mathsf{s}}^{\mathrm{tel}}_h \mid \mathsf{s}^{\mathrm{tel}}_h \sim \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}^{\mathrm{tel}}_h)$.

Proof. By Assumption D.1 and Theorem 3, we may assume that all joint distributions' conditional probabilities are regular conditional probabilities and thus almost surely equal to a kernel. Moreover, since all kernels are compatible with the direct decomposition, it suffices to prove the special case of the trivial direct-decomposition where $\mathcal{Z} = \mathcal{S}$. Fix a common measure $\mathbb{P}$ over which $\mathsf{s}_h^{\mathrm{tel}}, \tilde{\mathsf{s}}_h^{\mathrm{tel}}$, and s'_h are defined such that $\mathsf{s}_h^{\mathrm{tel}} \sim \mathsf{P}_h^\star$, $\mathsf{s}'_h \sim \mathsf{W}_\sigma(\mathsf{s}_h^{\mathrm{tel}})$, and $\tilde{\mathsf{s}}_h^{\mathrm{tel}} \sim \mathsf{W}_{\mathrm{dec},h}(\mathsf{s}'_h)$. Then for any measurable sets A, B, we have

$$\begin{aligned}\mathbb{P}(\mathsf{s}_h^{\mathrm{tel}} \in A,\, \mathsf{s}'_h \in B) &= \mathbb{P}(\mathsf{s}'_h \in B) \cdot \mathbb{E}_{\mathsf{s}'_h}\left[\mathbf{I}[\mathsf{s}'_h \in B] \cdot \mathbb{P}(\mathsf{s}_h^{\mathrm{tel}} \in A | \mathsf{s}'_h)\right] \\ &= \mathbb{P}(\mathsf{s}'_h \in B) \cdot \mathbb{E}_{\mathsf{s}'_h}\left[\mathbf{I}[\mathsf{s}'_h \in B] \cdot \mathbb{P}(\tilde{\mathsf{s}}_h^{\mathrm{tel}} \in A | \mathsf{s}'_h)\right] \\ &= \mathbb{P}\left(\tilde{\mathsf{s}}_h^{\mathrm{tel}} \in A,\, \mathsf{s}'_h \in B\right),\end{aligned}$$

where the first equality holds by the fact that we are working with regular conditional probabilities and Bayes' rule, the second equality holds by the definition of the deconvolution kernel above, and the last equality holds again by Bayes' rule and the tower rule for conditional expectations.

To prove the second statement, we apply induction, again assuming that $\mathcal{Z} = \mathcal{S}$ as in the proof of the first statement. Note that $\mathsf{s}_1^{\mathrm{tel}} \sim \mathsf{P}_1^\star = \mathsf{P}_{\mathrm{init}}$, and $\tilde{\mathsf{s}}_1^{\mathrm{tel}} \sim \mathsf{W}_{\circlearrowleft,1}^\star \circ \mathsf{P}_1^\star$. Thus, from the first part of the lemma, $\phi_{\mathcal{Z}}(\mathsf{s}_1^{\mathrm{tel}}) \sim \phi_{\mathcal{Z}} \circ \mathsf{P}_1^\star$. Now, suppose the induction holds up to step h. Then, $\tilde{\mathsf{s}}_h^{\mathrm{tel}} \sim \mathsf{P}_h^\star$, as $\mathsf{a}_h^{\mathrm{tel}} \sim \pi_h^\star(\mathsf{a}_h^{\mathrm{tel}})$, then $\mathsf{s}_{h+1}^{\mathrm{tel}} = F_h(\tilde{\mathsf{s}}_h^{\mathrm{tel}}, \mathsf{a}_h^{\mathrm{tel}}) \sim \mathsf{P}_{h+1}^\star$. Again $\tilde{\mathsf{s}}_{h+1}^{\mathrm{tel}} \sim \mathsf{W}_{\circlearrowleft,h+1}^\star(\mathsf{s}_{h+1}^{\mathrm{tel}})$, so that $\tilde{\mathsf{s}}_{h+1}^{\mathrm{tel}}$ has marginal $\mathsf{W}_{\circlearrowleft,h+1}^\star \circ \mathsf{P}_{h+1}^\star = \mathsf{P}_{h+1}^\star$, as needed. □

We further show that $\mathsf{W}_{\circlearrowleft,h}$ can be defined to be absolutely continuous with respect to $\mathsf{P}_h^\star$.

Lemma E.4. *The kernel $\mathsf{W}_{\circlearrowleft,h}$ satisfies that $\phi_{\mathcal{Z}} \circ \mathsf{W}_{\circlearrowleft,h} \ll \phi_{\mathcal{Z}} \circ \mathsf{P}_h^\star$ as laws, validating the second condition in (E.1). It further holds that $\phi_{\mathcal{Z}} \circ \mathsf{W}_{\mathrm{dec},h} \ll \phi_{\mathcal{Z}} \circ \mathsf{P}_h^\star$.*

Proof. The first statement follows immediately from Lemma E.3 because these distributions are the same. The second statement follows immediately from the tower law of conditional expectation and the definition of $\mathsf{W}_{\mathrm{dec},h}$. □

Lastly, we establish that the replica kernel inherits all concentration properties from the smoothing kernel.

Lemma E.5 (Replica Concentration)**.** *Recall that*

$$p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}', \mathsf{s}) > r].$$

We then have

$$\mathbb{P}_{\mathsf{s}_h \sim \mathsf{P}_h^\star, \tilde{\mathsf{s}}_h \sim \mathsf{W}_{\circlearrowleft,h}^\star(\mathsf{s}_h)}[\mathsf{d}_{\mathrm{IPS}}(\tilde{\mathsf{s}}_h, \mathsf{s}_h) > 2r] \le 2p_r$$

Proof. Again, all terms – $\mathsf{W}_\sigma, \mathsf{W}_{\circlearrowleft,h}^\star, \mathsf{W}_{\mathrm{dec},h}^\star$ and $\mathsf{d}_{\mathrm{IPS}}$ – are compatible with the direct decomposition, it suffices to consider the case of the trivial direct decomposition under whcih $\mathcal{Z} = \mathcal{S}$.

Let $\mathbb{P}$ denote a distribution over $\mathsf{s}_h \sim \mathsf{P}_h^\star$, $\mathsf{s}'_h \sim \mathsf{W}_\sigma(\mathsf{s}_h)$, and $\tilde{\mathsf{s}}_h \sim \mathsf{W}_{\mathrm{dec},h}^\star(\mathsf{s}'_h)$. In this special case, we see that $\tilde{\mathsf{s}}_h \mid \mathsf{s}_h \sim \mathsf{W}_{\circlearrowleft,h}^\star(\mathsf{s}_h)$[4]. By a union bound,

$$\begin{aligned}\mathbb{P}_{\mathsf{s}_h \sim \mathsf{P}_h^\star, \tilde{\mathsf{s}}_h \sim \mathsf{W}_{\circlearrowleft,h}^\star(\mathsf{s}_h)}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \tilde{\mathsf{s}}_h) > 2r] &\le \mathbb{P}[\mathsf{d}_{\mathrm{IPS}}(\tilde{\mathsf{s}}_h, \mathsf{s}'_h) > r] + \mathbb{P}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) > r] \\ &= 2\,\mathbb{P}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) > r] \le 2p_r,\end{aligned}$$

where the equality follows from the first statment of Lemma E.3. □

Remark E.2. Note that, in the previous lemma, it suffices that the following weaker condition holds: $\mathbb{P}_{\mathsf{s} \sim \mathsf{P}_h^\star, \mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}', \mathsf{s}) > r] \le p_r$, i.e. for concentration to hold only in distribution over $\mathsf{s} \sim \mathsf{P}_h^\star$, instead of *uniformly* over states.

We now proceed to formally prove Theorem 4

[4] Notice that, for general $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$, this condition would become $\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}_h) \mid \phi_{\mathcal{Z}}(\mathsf{s}_h) \sim \phi_{\mathcal{Z}} \circ \mathsf{W}_{\circlearrowleft,h}^\star(\phi_{\mathcal{Z}}(\mathsf{s}_h), \cdot)$, where the $\cdot$ argument is irrelevant.

E.2.3. Formal proof of Theorem 4

Key Events. For the random variables defined above, we define three groups of events.

- The *coupling events*, denoted by $\mathcal{B}$, which are controlled by carefully selecting a coupling.
- The *inductive events*, denoted by $\mathcal{C}$, which we condition on when bounding the probability of the coupling events.
- The *stability events*, denoted by $\mathcal{Q}$, which take advantage of the stability properties of the imitation policy.

Definition E.6 (Coupling Events)**.** Define the events

$$\begin{aligned}
\mathcal{B}_{\mathrm{tel},h} &= \left\{\mathsf{a}_h^{\circlearrowleft} = \mathsf{a}_h^{\mathrm{tel}},\ \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}_h^{\circlearrowleft}) = \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}_h^{\mathrm{tel}})\right\}\\
\mathcal{B}_{\mathrm{est},h} &= \left\{\vec{\mathsf{d}}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{tel,inter}}, \mathsf{a}_h^{\mathrm{tel}}) \not\preceq \vec{\varepsilon}\right\}\\
\mathcal{B}_{\mathrm{inter},h} &= \left\{\hat{\mathsf{a}}_h^{\mathrm{tel,inter}} = \hat{\mathsf{a}}_h^{\circlearrowleft,\mathrm{inter}}\right\}\\
\mathcal{B}_{\hat{\mathsf{a}},h} &= \left\{\hat{\mathsf{a}}_h^{\circlearrowleft,\mathrm{inter}} = \hat{\mathsf{a}}_h\right\}\\
\mathcal{B}_{\mathrm{all},h} &= \mathcal{B}_{\mathrm{inter},h} \cap \mathcal{B}_{\mathrm{tel},h} \cap \mathcal{B}_{\mathrm{est},h} \cap \mathcal{B}_{\hat{\mathsf{a}},h}\\
\bar{\mathcal{B}}_{\mathrm{all},h} &= \bigcap_{j=1}^{h} \mathcal{B}_{\mathrm{all},h}
\end{aligned}$$

Notice that each of the events above are $\mathcal{F}_h$-measurable. Moreover, note that on $\bar{\mathcal{B}}_{\mathrm{all},h}$, $\max_{1\le j\le h} \phi_{\mathrm{IS}}(\hat{\mathsf{a}}_j, \mathsf{a}_j^{\circlearrowleft}) \le \varepsilon$.

Definition E.7 (Inductive Event)**.** Define the events

$$\begin{aligned}
\mathcal{C}_{\hat{\mathsf{s}},h} &= \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_h^{\circlearrowleft}, \hat{\mathsf{s}}_h) \preceq \vec{\varepsilon}\right\},\\
\mathcal{C}_{\mathrm{tel},h} &= \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_h^{\circlearrowleft}, \mathsf{s}_h^{\mathrm{tel}}) \preceq \vec{\gamma}_{\mathrm{IPS}}(2r)\right\}\\
\mathcal{C}_{\mathrm{all},h} &:= \mathcal{C}_{\hat{\mathsf{s}},h} \cap \mathcal{C}_{\mathrm{tel},h}\\
\bar{\mathcal{C}}_{\mathrm{all},h} &= \bigcap_{j=1}^{h} \mathcal{C}_{\mathrm{all},j}
\end{aligned}$$

Notice that all the above events are $\mathcal{F}_{h-1}$-measurable, due to determinism of the dynamics. Note that also $\mathbb{P}_\mu[\bar{\mathcal{C}}_{\mathrm{all},1}] = 1$ for any μ that respects the construction (as $\mathsf{s}_1^{\circlearrowleft} = \mathsf{s}_1^{\mathrm{tel}} = \hat{\mathsf{s}}_1$).

Definition E.8 (Stability Events)**.** Define the events

$$\begin{aligned}
\mathcal{Q}_{\mathrm{close}} &:= \left\{\forall h \in [H] : \mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h^{\circlearrowleft}, \tilde{\mathsf{s}}_h^{\circlearrowleft}) \le 2r\right\}\\
\mathcal{Q}_{\mathrm{IS}} &:= \left\{(\mathsf{s}_{1:H+1}^{\circlearrowleft}, \mathsf{a}_{1:H}^{\circlearrowleft}) \text{ is input-stable w.r.t. } (\vec{\mathsf{d}}_{\mathcal{S}}, \vec{\mathsf{d}}_{\mathcal{A}})\right\}\\
\mathcal{Q}_{\mathrm{IPS}} &:= \left\{\vec{\mathsf{d}}_{\mathcal{S}}(F_h(\tilde{\mathsf{s}}_h^{\circlearrowleft}, \mathsf{a}_h^{\circlearrowleft}), \mathsf{s}_{h+1}^{\circlearrowleft}) \le \vec{\gamma}_{\mathrm{IPS}} \circ \mathsf{d}_{\mathrm{IPS}}\left(\tilde{\mathsf{s}}_h^{\circlearrowleft}, \mathsf{s}_h^{\circlearrowleft}\right), \quad 1 \le j \le H\right\}\\
\mathcal{Q}_{\mathrm{all}} &:= \mathcal{Q}_{\mathrm{IPS}} \cap \mathcal{Q}_{\mathrm{close}}.
\end{aligned}$$

In words, $\mathcal{Q}_{\mathrm{close}}$ the event on which $\mathsf{s}_h^{\circlearrowleft}$ and $\tilde{\mathsf{s}}_h^{\circlearrowleft} \sim \mathsf{W}^{\star}_{\circlearrowleft,h}(\mathsf{s}_h^{\mathrm{tel}})$ are close, and $\mathcal{Q}_{\mathrm{IS}}$and $\mathcal{Q}_{\mathrm{IPS}}$ ensure consequencs of (vector) input-stability and (vector) input process stability holds.

Steps of the proof. First, we use stability to reduce the event $\bar{\mathcal{C}}_{\mathrm{all},h+1}$ to $\bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h}$:

Claim E.6 (Stability Claim)**.** *By construction,*

$$\bar{\mathcal{C}}_{\mathrm{all},h+1} \subset \mathcal{Q}_{\mathrm{all}} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h}.$$

Proof. It suffices to show that on $\mathcal{Q}_{\text{all}} \cap \bar{\mathcal{C}}_{\text{all},h} \cap \bar{\mathcal{B}}_{\text{all},h}$, $\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}) \preceq \vec{\varepsilon}$ and $\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \mathsf{s}^{\text{tel}}_{h+1}) \preceq \vec{\gamma}_{\text{IPS}}(2r)$. By applying the event $\mathcal{Q}_{\text{IS}}$ to the sequence $\mathsf{a}'_h = \hat{\mathsf{a}}_h$ and $\mathsf{s}'_h = \hat{\mathsf{s}}_h$, we have that on $\mathcal{Q}_{\text{all}} \subset \mathcal{Q}_{\text{IS}}$ that

$$\forall h \in [H], i \in [K], \quad \mathsf{d}_{\mathcal{S},i}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}) \le \max_{1 \le j \le h} \mathsf{d}_{\mathcal{A},i}\left(\mathsf{a}^{\circlearrowleft}_j, \hat{\mathsf{a}}_j\right)$$

For the next point, note that the compatibility of the dynamics with the direct decomposition $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ implies that there exists a dynamics map $F^{\mathcal{Z}}_h$ for which

$$F_h(\mathsf{s}, \mathsf{a}) = F^{\mathcal{Z}}_h(\phi_{\mathcal{Z}}(\mathsf{s}), \mathsf{a}).$$

Similarly, by applying $\mathcal{Q}_{\text{IPS}}$ and $\mathcal{Q}_{\text{close}}$ and the event $\{\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h) = \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\text{tel}}_h), \mathsf{a}^{\text{tel}}_h = \mathsf{a}^{\circlearrowleft}_h\}$ on $\mathcal{B}_{\text{tel},h}$, it holds that on $\mathcal{Q}_{\text{all}} \cap \bar{\mathcal{C}}_{\text{all},h} \cap \bar{\mathcal{B}}_{\text{all},h}$ that, for all $h \in [H]$,

$$\begin{aligned}
\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, F_h(\tilde{\mathsf{s}}^{\circlearrowleft}_h, \mathsf{a}^{\circlearrowleft}_h)) &= \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, F^{\mathcal{Z}}_h(\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h), \mathsf{a}^{\circlearrowleft}_h)) \\
&= \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, F^{\mathcal{Z}}_h(\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\text{tel}}_h), \mathsf{a}^{\text{tel}}_h)) && (\mathcal{B}_{\text{tel},h}) \\
&= \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, F_h(\tilde{\mathsf{s}}^{\text{tel}}_h, \mathsf{a}^{\text{tel}}_h)) \\
&= \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \mathsf{s}^{\text{tel}}_{h+1}) \\
&\le \vec{\gamma}_{\text{IPS}} \circ \mathsf{d}_{\text{IPS}}\left(\mathsf{s}^{\text{tel}}_j, \tilde{\mathsf{s}}^{\text{tel}}_j\right) && (\mathcal{Q}_{\text{IPS}}) \\
&\le \vec{\gamma}_{\text{IPS}} \circ \mathsf{d}_{\text{IPS}}\left(2r\right). && (\mathcal{Q}_{\text{close}})
\end{aligned}$$

□

From Claim E.6, we decompose our error probability as follows:

Lemma E.7 (Key Error Decomposition)**.** *Let μ respect the construction (in the sense of Appendix E.2.1). Then*

$$\begin{aligned}
&\mathbb{P}_\mu[\exists h \in [H] : \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}), \phi_{\text{IS}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h)\} > \varepsilon] \\
&\le \mathbb{P}_\mu[\mathcal{Q}^c_{\text{all}}] + \sum_{h=1}^{H} \mathbb{P}_\mu[\mathcal{B}^c_{\text{all},h} \cap \bar{\mathcal{C}}_{\text{all},h} \cap \bar{\mathcal{B}}_{\text{all},h-1}]
\end{aligned}$$

Hence, letting $\inf_\mu$ denote the infinum over couplings μ which respect the construction,

$$\begin{aligned}
&\vec{\Gamma}_{\text{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) \vee \vec{\Gamma}_{\text{marg},\vec{\varepsilon}_{\text{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star) \\
&\le \inf_\mu \left\{ \mathbb{P}_\mu[\mathcal{Q}^c_{\text{all}}] + \sum_{h=1}^{H} \mathbb{P}_\mu[\bar{\mathcal{B}}^c_{\text{all},h} \cap \bar{\mathcal{C}}_{\text{all},h} \cap \bar{\mathcal{B}}_{\text{all},h-1}] \right\}
\end{aligned} \tag{E.4}$$

Proof. Define the events $\mathcal{E}_h := \bar{\mathcal{C}}_{\text{all},h+1} \cap \bar{\mathcal{B}}_{\text{all},h}$. Observe that the events are nested: $\mathcal{E}_h \supset \mathcal{E}_{h+1}$, and that on $\mathcal{E}_H$, we have that for all $h \in [H]$

$$\begin{aligned}
\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h) &\preceq \vec{\varepsilon} \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h) && (\mathcal{C}_{\hat{\mathsf{s}},h+1} \supset \bar{\mathcal{C}}_{\text{all},h+1} \supset \mathcal{E}_h) \\
&\preceq \vec{\varepsilon}. && (\bar{\mathcal{B}}_{\text{all},h} \supset \mathcal{E}_h)
\end{aligned}$$

Thus,

$$\mathbb{P}_\mu[\exists h \in [H] : \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h)\} \not\preceq \vec{\varepsilon}] \le \mathbb{P}_\mu[\mathcal{E}^c_H] \le \mathbb{P}_\mu[(\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H)^c] \tag{E.5}$$

As $(\mathsf{s}^{\circlearrowleft}_{1:H+1}, \mathsf{a}^{\circlearrowleft}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\circlearrowleft}}$, this shows $\vec{\Gamma}_{\text{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) \le \mathbb{P}_\mu[(\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H)^c]$. Moreover, on $\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H$, we have that

$$\max_h \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_h, \mathsf{s}^{\text{tel}}_h) \le \vec{\gamma}_{\text{IPS}}(2r),$$

so that, by the inequality preceeding (E.5), the following holds for all $h \in [H]$ on $\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H$.

$$\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h) \le \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h) \le \vec{\varepsilon}. \tag{E.6}$$

By construction, for each h, $\mathsf{a}_h^{\mathrm{tel}} \mid \mathcal{F}_h \sim \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^{\mathrm{tel}})$. Moreover, Lemma E.3 implies that $\mathsf{s}_h^{\mathrm{tel}}$ has the marginal distribution of $\mathsf{s}_h^\star \sim \mathsf{P}_h^\star$. Thus, for each h, $\mathsf{s}_{h+1}^{\mathrm{tel}}$ and $\mathsf{a}_h^{\mathrm{tel}}$ have the same *marginals* as the marginals under $\mathsf{D}_{\pi^\star}$. Consequently, (E.6) implies that,

$$\begin{aligned}\vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star) &:= \max_{h\in[H]} \max\left\{\inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\vec{\mathsf{d}}_{\mathcal{S}}(\hat{\mathsf{s}}_{h+1}, \mathsf{s}_{h+1}^\star) \not\preceq \vec{\varepsilon}\right], \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^\star, \hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}\right]\right\} \\ &\le \mathbb{P}_\mu[(\mathcal{Q}_{\mathrm{all}} \cap \mathcal{E}_H)^c].\end{aligned}$$

where above we take inf over $\mu_1 \in \mathscr{C}(\mathsf{D}_{\hat{\pi}_\sigma}, \mathsf{D}_{\pi^\star})$. Summarizing our findings thus far,

$$\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) \vee \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star) \le \mathbb{P}_\mu[(\mathcal{Q}_{\mathrm{all}} \cap \mathcal{E}_H)^c].$$

Let us conclude by bounding $\mathbb{P}_\mu[(\mathcal{Q}_{\mathrm{all}} \cap \mathcal{E}_H)^c]$. Using the nesting structure $\mathcal{E}_h = \bigcap_{j=1}^h \mathcal{E}_j$, the peeling lemma, Lemma C.11, and a union bound, it holds that

$$\begin{aligned}\mathbb{P}_\mu\left[(\mathcal{Q}_{\mathrm{all}} \cap \mathcal{E}_H)^c\right] &\le \mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c] + \mathbb{P}\left[\exists h \in [H] \text{ s.t. } (\mathcal{Q}_{\mathrm{all}} \cap \mathcal{E}_{h-1} \cap \mathcal{E}_h^c) \text{ holds}\right] \\ &\le \mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c] + \sum_{h=1}^H \mathbb{P}_\mu\left[\mathcal{Q}_{\mathrm{all}} \cap \mathcal{E}_{h-1} \cap \mathcal{E}_h^c \text{ holds}\right] \\ &= \mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c] + \sum_{h=1}^H \mathbb{P}_\mu\left[\mathcal{Q}_{\mathrm{all}} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap (\bar{\mathcal{B}}_{\mathrm{all},h} \cap \bar{\mathcal{C}}_{\mathrm{all},h+1})^c \text{ holds}\right] \\ &= \mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c] + \sum_{h=1}^H \mathbb{P}_\mu\left[\mathcal{Q}_{\mathrm{all}} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h}^c\right] \\ &= \mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c] + \sum_{h=1}^H \mathbb{P}_\mu\left[\mathcal{Q}_{\mathrm{all}} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \mathcal{B}_{\mathrm{all},h}^c\right],\end{aligned}$$

where the last step invokes Claim E.6. □

Next, we bound the contribution of $\mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c]$ in (E.4), uniformly over all couplings.

Lemma E.8. *For all μ which respect the construction,*

$$\mathbb{P}_\mu[\mathcal{Q}_{\mathrm{all}}^c] \le p_{\mathrm{IPS}} + 2Hp_r.$$

Proof. $\mathbb{P}_\mu[\mathcal{Q}_{\mathrm{close}}^c] = \mathbb{P}_\mu[\exists h : \mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h^{\mathrm{tel}}, \tilde{\mathsf{s}}_h^{\mathrm{tel}}) > 2r] \le 2Hp_r$ by Lemma E.5 and a union bound.

Let us now bound $\mathbb{P}_\mu[\mathcal{Q}_{\mathrm{close}} \cap \mathcal{Q}_{\mathrm{IPS}}^c] \le \mathbb{P}_\mu[\mathcal{Q}_{\mathrm{IPS}}^c \mid \mathcal{Q}_{\mathrm{close}}]$. Define the kernels $\mathsf{W}_h(\mathsf{s})$ to be equal to the kernel $\mathsf{W}_{\circlearrowleft,h}(\mathsf{s})$ conditioned on the event $\mathsf{s}' \sim \mathsf{W}_{\circlearrowleft,h}(\mathsf{s})$ satisfies $\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}', \mathsf{s}) \le 2r$. Then, conditional on $\mathcal{Q}_{\mathrm{close}}$, we see that the sequence $(\mathsf{s}_{1:H+1}^{\circlearrowleft}, \tilde{\mathsf{s}}_{1:H}^{\circlearrowleft}, \mathsf{a}_{1:H}^{\circlearrowleft})$ obeys the generative process

$$\tilde{\mathsf{s}}_h^{\circlearrowleft} \mid \tilde{\mathsf{s}}_{1:h-1}^{\circlearrowleft}, \mathsf{s}_{1:h}^{\circlearrowleft}, \mathsf{a}_{1:h-1}^{\circlearrowleft} \sim \mathsf{W}_h(\mathsf{s}), \quad \mathsf{a}_h^{\circlearrowleft} \mid \tilde{\mathsf{s}}_{1:h}^{\circlearrowleft}, \mathsf{s}_{1:h}^{\circlearrowleft}, \mathsf{a}_{1:h-1}^{\circlearrowleft} \sim \pi_h^\star(\tilde{\mathsf{s}}_h^{\circlearrowleft}), \quad \mathsf{s}_{h+1}^{\circlearrowleft} = F_h(\mathsf{s}_h^{\circlearrowleft}, \mathsf{a}_h^{\circlearrowleft}).$$

By construction, for each h, $\mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_{\circlearrowleft,h}(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}', \mathsf{s}) > 2r] = 0$. Thus, the definition of (vector) input process stability (Definition E.4) and assumption $r \le \frac{1}{2}r_{\mathrm{IPS}}$ implies that $\mathbb{P}_\mu[\mathcal{Q}_{\mathrm{IPS}}^c \mid \mathcal{Q}_{\mathrm{close}}] \le p_{\mathrm{IPS}}$. □

The remaining step of the proof is therefore to bound the second term in (E.4).

Lemma E.9. *There exists a coupling μ which respects the construction and satisfies the following for any $h \in [H]$*

$$\begin{aligned}&\mathbb{P}_\mu[\mathcal{B}_{\mathrm{all},h}^c \mid \mathcal{F}_{h-1}] \\ &\le \hat{\gamma} \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\circlearrowleft}, \hat{\mathsf{s}}_h) + (\hat{\gamma} + \gamma_\sigma) \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\circlearrowleft}, \mathsf{s}_h^{\mathrm{tel}}) + \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^{\mathrm{tel}})), \ \mu\text{-almost surely}\end{aligned}$$

Consequently, for all $h \in [H]$,

$$\begin{aligned}&\mathbb{P}_\mu[\mathcal{B}_{\mathrm{all},h}^c \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1}] \\ &\le \hat{\gamma}(\vec{\varepsilon}_1) + (\hat{\gamma} + \gamma_\sigma) \circ \gamma_{\mathrm{IPS},1}(2r) + \mathbb{E}_\mu[\vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}))]\end{aligned}$$

Moreover, $\mathsf{s} \mapsto \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}))$ is measurable.

Proof Sketch. We begin by giving a high level overview of the construction, which is done recursively. The key technical tool is Lemma C.2 above, which allows us to transform any coupling μ between random variables (X, Y) into a probability kernel $\mu(\cdot|X)$ mapping instances of X to probability distributions on Y such that $(X, Y) \sim \mu$ has the same law as $(X, Y \sim \mu(\cdot|X))$. For each h, we then show that, assuming the coupling has kept the states and controls close together until time $h-1$, this will imply the following chain:

$$\underbrace{(\mathsf{a}^{\circlearrowleft} \leftrightarrow \mathsf{a}^{\mathrm{tel}})}_{\gamma_{\mathrm{TVC}} \text{ and induction}} \to \underbrace{(\mathsf{a}^{\mathrm{tel}} \leftrightarrow \hat{\mathsf{a}}^{\mathrm{tel,inter}})}_{\text{learning and sampling}} \to \underbrace{(\hat{\mathsf{a}}^{\mathrm{tel,inter}} \leftrightarrow \hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}})}_{\gamma_{\mathrm{TVC}} \text{ and induction}} \to \underbrace{(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}} \leftrightarrow \hat{\mathsf{a}})}_{\gamma_{\mathrm{TVC}} \text{ and induction}},$$

where the bidirectional arrows indicate individual couplings between the laws of the random variables that are constructed by the method outlined in text below and the single directional arrows denote the probability kernels described above. The full proof of the lemma is given in Appendix E.2.4. □

Concluding the proof. Here, we finish the proof of Theorem 4. Recall that we wish to bound $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) \vee \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star)$. We begin by bounding $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) \vee \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft})$. In light of Lemma E.7, it suffices to bound

$$\mathbb{P}_\mu[\mathcal{Q}^c_{\mathrm{all}}] + \sum_{h=1}^{H} \mathbb{P}_\mu[\bar{\mathcal{B}}^c_{\mathrm{all},h} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1}],$$

where μ is the coupling in Lemma E.9. Applying Lemma E.8 and Lemma E.9,

$$\begin{aligned}
&\mathbb{P}_\mu[\mathcal{Q}^c_{\mathrm{all}}] + \sum_{h=1}^{H} \mathbb{P}_\mu[\bar{\mathcal{B}}^c_{\mathrm{all},h} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1}] \\
&\le p_{\mathrm{IPS}} + 2Hp_r + \sum_{h=1}^{H} \mathbb{P}_\mu[\bar{\mathcal{B}}^c_{\mathrm{all},h} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1}] \\
&\le p_{\mathrm{IPS}} + H(2p_r + \hat{\gamma}(\vec{\varepsilon}_1) + (\hat{\gamma} + \gamma_\sigma) \circ \gamma_{\mathrm{IPS},1}(2r)) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^{\mathrm{tel}}_h \sim \mu} \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h))
\end{aligned}$$

To conclude, we note that for any μ which respects the construction, Lemma E.3 ensures that $\mathsf{s}^{\mathrm{tel}}_h$ as the marginal distribution of $\mathsf{s}^\star_h \sim \pi^\star_h$. Thus, the above is at most

$$p_{\mathrm{IPS}} + H(2p_r + \hat{\gamma}(\vec{\varepsilon}_1) + (\hat{\gamma} + \gamma_\sigma) \circ \gamma_{\mathrm{IPS},1}(2r)) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}^\star_h) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^\star_h)) \tag{E.7}$$

which concludes the proof of (E.2) for $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi} \parallel \pi^\star_{\circlearrowleft})$.

To prove (E.3) for $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi} \parallel \pi^\star_{\circlearrowleft})$, we consider the special case that $\hat{\pi}_\sigma = \hat{\pi} \circ \mathsf{W}_\sigma$. By definition, $\hat{\pi}_{\sigma,h} = \hat{\pi} \circ \mathsf{W}_\sigma$. Thus, the data-processing inequality for optimal transport (Lemma C.5)

$$\vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}^\star_h) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^\star_h)) \le \mathbb{E}_{\mathsf{s}'_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}(\mathsf{s}'_h) \parallel \pi^\star_{\mathrm{dec},h}(\mathsf{s}'_h)),$$

for all $\mathsf{s}^\star_h$. Substituting this into (E.7), and setting $\hat{\gamma} = \gamma_\sigma$ (in view of Lemma E.2), finishes the argument.

E.2.4. Proof of Lemma E.9

Recall that Assumption E.1 ensures all of the general measure-theoretic guarantees of Appendix C hold true in our setting. Notably we need the gluing lemma (Lemma C.2) and the commuting of optimal transport metrics and conditional probabilities (Proposition C.3).

Proof strategy. Our proof follows along similar lines as that of Proposition D.1, although with the added complication of including the smoothing. We will inductively construct μ. A useful schematic for the construction at each step is the following diagram:

$$\underbrace{(\tilde{\mathsf{s}}^{\circlearrowleft} \leftrightarrow \tilde{\mathsf{s}}^{\mathrm{tel}}), (\mathsf{a}^{\circlearrowleft} \leftrightarrow \mathsf{a}^{\mathrm{tel}})}_{\mathcal{B}_{\mathrm{tel},h}} \to \underbrace{(\mathsf{a}^{\mathrm{tel}} \leftrightarrow \hat{\mathsf{a}}^{\mathrm{tel,inter}})}_{\mathcal{B}_{\mathrm{est},h}} \to \underbrace{(\hat{\mathsf{a}}^{\mathrm{tel,inter}} \leftrightarrow \hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}})}_{\mathcal{B}_{\mathrm{inter},h}} \to \underbrace{(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}} \leftrightarrow \hat{\mathsf{a}})}_{\mathcal{B}_{\hat{\mathsf{a}},h}},$$

where the events under each bidirectional arrow refer to the event such ensuring that there exists a coupling such that the objects are close. We then will apply Lemma C.2 to glue the individual couplings together. We will then use Lemma C.11 and a union bound to control the probability under our constructed coupling that any of the relevant events fail to hold, concluding the proof.

Recursive construction of μ. Let $h \geq 1$, and suppose that we have constructed the coupling $\mu^{(1:h-1)}$ for steps $1, \ldots, h-1$ which respects the construction. Recall that $\mathcal{F}_h$ denotes the sigma-algebra generated by $(\hat{\mathsf{s}}_{1:h}, \mathsf{s}^{\circlearrowleft}_{1:h}, \mathsf{s}^{\mathrm{tel}}_{1:h})$, $(\mathsf{a}^{\circlearrowleft}_{1:h}, \tilde{\mathsf{s}}^{\circlearrowleft}_{1:h}, \tilde{\mathsf{s}}^{\mathrm{tel}}_{1:h}, \mathsf{a}^{\mathrm{tel}}_{1:h}, \hat{\mathsf{a}}_{1:h})$, and $(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_{1:h}, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_{1:h})$. Notice that $\mathsf{s}^{\mathrm{tel}}_{h+1}, \mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}$ are determined by $\mathcal{F}_h$ as well. Similarly, it can be seen from Definition E.5 that $\phi_{\mathcal{V}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_{h+1})$ and $\phi_{\mathcal{V}}(\tilde{\mathsf{s}}^{\circlearrowleft}_{h+1})$ are also determined by $\mathcal{F}_h$ (since the replica kernel preserves the $\mathcal{V}$-components). We summarize all these aforementioned variables in a random variable Y_h. Let $\mathcal{F}_0$ denote the filtration generated by $\mathsf{s}^{\circlearrowleft}_1 = \mathsf{s}^{\mathrm{tel}}_1 = \hat{\mathsf{s}}_1$. We let $Y_0 = (\mathsf{s}^{\circlearrowleft}_1, \mathsf{s}^{\mathrm{tel}}_1, \hat{\mathsf{s}}_1)$.

Correspondingly, let Z_h denote the random variables $(\mathsf{a}^{\circlearrowleft}_h, \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h), \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_h), \mathsf{a}^{\mathrm{tel}}_h, \hat{\mathsf{a}}_h)$, and $(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_h)$ such that the joint law of these random variables respects the construction. Our goal is then to specify, for each $h \in [H]$, a joint distribution of (Y_{h-1}, Z_h). Note that Z_h, Y_{h-1} determines Y_h, and we call this induced law $\mu^{(h)}$.

We begin by specifying joint distributions conditional on Y_{h-1} and subsets of Z_h, then glue them together by the gluing lemma. Below, we use use information-theoretic notation.

- By total variation continuity of $\phi_{\mathcal{Z}} \circ \mathsf{W}^{\star}_{\circlearrowleft,h}$ (Lemma E.2),

$$\mathsf{TV}(\mathbb{P}_{\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h)|Y_{h-1}}, \mathbb{P}_{\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)|Y_{h-1}}) \leq \gamma_\sigma \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}^{\circlearrowleft}_h, \mathsf{s}^{\mathrm{tel}}_h).$$

  Because $\mathsf{a}^{\circlearrowleft}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\circlearrowleft}_{h+1})$ and $\mathsf{a}^{\mathrm{tel}}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)$, and $\pi^\star$ is compatible with the decomposition $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ (i.e. $\pi^\star_h(\mathsf{s})$ is a function of $\phi_{\mathcal{Z}}(\mathsf{s})$) Lemma C.4 implies that (almost surely)

$$\mathsf{TV}(\mathbb{P}_{(\mathsf{a}^{\circlearrowleft}_h, \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h))|Y_{h-1}}, \mathbb{P}_{(\mathsf{a}^{\mathrm{tel}}_h, \phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_h))|Y_{h-1}}) \leq \gamma_\sigma \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}^{\circlearrowleft}_h, \mathsf{s}^{\mathrm{tel}}_h).$$

  Hence, Corollary C.1 implies that there exists a coupling $\mu^{(h)}_{\mathrm{tel}}$ over $Y_{h-1}, (\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h), \mathsf{a}^{\circlearrowleft}_h), (\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_h), \mathsf{a}^{\mathrm{tel}}_h)$ respecting the construction such that $Y_h \sim \mu^{(h-1)}$ and such that (almost surely)

$$\mathbb{E}_{\mu^{(h)}_{\mathrm{tel}}}[\mathcal{B}_{\mathrm{tel},h} \mid Y_{h-1}] = \mathbb{P}_{\mu^{(h)}_{\mathrm{tel}}}[(\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\circlearrowleft}_h), \mathsf{a}^{\circlearrowleft}_h) \neq (\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_h), \mathsf{a}^{\mathrm{tel}}_h) \mid Y_{h-1}] \leq \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}^{\circlearrowleft}_h, \mathsf{s}^{\mathrm{tel}}_h)].$$

- In our construction, $\mathsf{a}^{\mathrm{tel}}_h \mid Y_{h-1} \sim \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h)$, and $\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h \mid Y_{h-1} \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h)$. Thus, by definition of $\vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}$, and the assumption $\mathbf{I}\{\vec{\mathsf{d}}_{\mathcal{A}}(\cdot,\cdot) \not\preceq \vec{\varepsilon}\}$ is lower semicontinuous, Proposition C.3 implies that we may find a coupling $\mu^{(h)}_{\mathrm{est}}$ of $(\mathsf{a}^{\mathrm{tel}}_h, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_h, Y_{h-1})$ respecting the construction such that, almost surely,

$$\begin{aligned}\mathbb{P}_{\mu^{(h)}_{\mathrm{est}}}\left[\mathcal{B}^c_{\mathrm{est},h} \mid Y_{h-1}\right] &= \mathbb{P}_{\mu^{(h)}_{\mathrm{est}}}\left[\vec{\mathsf{d}}_{\mathcal{A}}(\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h, \mathsf{a}^{\mathrm{tel}}_h) \not\preceq \vec{\varepsilon} \mid Y_{h-1}\right] \\ &= \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h))].\end{aligned}$$

  Moreover, that same proposition ensures measurability of $\mathsf{s} \to \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}))$.

- Since $\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h \mid \mathcal{F}_h \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h)$ and $\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_{h+1} \mid \mathcal{F}_h \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\circlearrowleft}_h)$, and since $\hat{\pi}_{\sigma,h}(\cdot)$ is $\hat{\gamma}$-TVC by assumption,

$$\mathsf{TV}(\mathbb{P}_{\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h|Y_{h-1}}, \mathbb{P}_{\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h|Y_{h-1}}) \leq \hat{\gamma} \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}^{\circlearrowleft}_h, \mathsf{s}^{\mathrm{tel}}_h).$$

  Corollary C.1 implies that there is a coupling $\mu^{(h)}_{\mathrm{inter}}$ between $(\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h, \hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h, Y_{h-1})$ such that

$$\mathbb{P}_{\mu^{(h)}_{\mathrm{inter}}}[\mathcal{B}^c_{\mathrm{inter},h} \mid Y_{h-1}] = \mathbb{P}_{\mu^{(h)}_{\mathrm{inter}}}\left[\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h \neq \hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h \mid Y_{h-1}\right] \leq \hat{\gamma} \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}^{\mathrm{tel}}_h, \mathsf{s}^{\circlearrowleft}_h)$$

- Similarly, since $\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_h(\mathsf{s}^{\circlearrowleft}_h)$ and $\hat{\mathsf{a}}_{h+1} \mid \mathcal{F}_{h-1} \sim \hat{\pi}_h(\hat{\mathsf{s}}_h)$, $\hat{\pi}_h(\cdot)$ is $\hat{\gamma}$-TVC, Corollary C.1 implies that there is a coupling $\mu^{(h)}_{\hat{\mathsf{a}}}$ between $(\hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h, \hat{\mathsf{a}}_h, Y_{h-1})$ such that

$$\mathbb{P}_{\mu^{(h)}_{\hat{\mathsf{a}}}}[\mathcal{B}^c_{\hat{\mathsf{a}},h} \mid Y_{h-1}] = \mathbb{P}_{\mu^{(h)}_{\hat{\mathsf{a}}}}\left[\hat{\mathsf{a}}_h \neq \hat{\mathsf{a}}^{\circlearrowleft,\mathrm{inter}}_h \mid Y_{h-1}\right] \leq \hat{\gamma} \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}^{\circlearrowleft}_h, \hat{\mathsf{s}}_h)$$

We can then apply the gluing lemma (Lemma C.2) to

$$\begin{aligned}
X_{h,1} &= (\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}), \mathsf{a}_h^{\mathrm{tel}}, Y_{h-1}) \\
X_{h,2} &= (\phi_{\mathcal{Z}}(\tilde{\mathsf{s}}_h^{\circlearrowleft}), \mathsf{a}_h^{\circlearrowleft}, Y_{h-1}) \\
X_{h,3} &= (\mathsf{a}_h^{\mathrm{tel}}, \hat{\mathsf{a}}_h^{\mathrm{tel,inter}}, Y_{h-1}) \\
X_{h,4} &= (\hat{\mathsf{a}}_h^{\mathrm{tel,inter}}, \hat{\mathsf{a}}_h^{\circlearrowleft,\mathrm{inter}}, Y_{h-1}) \\
X_{h,5} &= (\hat{\mathsf{a}}_h^{\circlearrowleft,\mathrm{inter}}, \hat{\mathsf{a}}_h, Y_{h-1})
\end{aligned}$$

with

$$(X_{h,1}, X_{h,2}) \sim \mu_{\mathrm{tel}}^{(h)}, \quad (X_{h,2}, X_{h,3}) \sim \mu_{\mathrm{est}}^{(h)}, \quad (X_{h,3}, X_{h,4}) \sim \mu_{\mathrm{inter}}^{(h)}, \quad (X_{h,4}, X_{h,5}) \sim \mu_{\hat{\mathsf{a}}}^{(h)}.$$

Lemma C.2 guarantees the existence of a coupling $\mu^{(h)}$ consident with all sub-couplings $\mu_{\mathrm{tel}}^{(h)}, \mu_{\mathrm{est}}^{(h)}, \mu_{\mathrm{intp}}^{(h)}, \mu_{\hat{\mathsf{a}}}^{(h)}$. Then, $\mu^{(h)}$-almost surely (and using that $\mathcal{F}_{h-1}$ is precisely the σ-algebra generated by Y_{h-1})

$$\begin{aligned}
&\mathbb{P}_{\mu^{(h)}}[\mathcal{B}^c_{\mathrm{all},h} \mid \mathcal{F}_{h-1}] \\
&\le \mathbb{P}_{\mu^{(h)}}[\mathcal{B}^c_{\mathrm{tel},h} \mid \mathcal{F}_{h-1}] + \mathbb{P}_{\mu^{(h)}}[\mathcal{B}^c_{\mathrm{est},h}\mathcal{F}_{h-1}] + \mathbb{P}_{\mu^{(h)}}[\mathcal{B}^c_{\mathrm{inter},h}\mathcal{F}_{h-1}] + \mathbb{P}_{\mu^{(h)}}[\mathcal{B}^c_{\hat{\mathsf{a}},h}\mathcal{F}_{h-1}] \\
&\le \hat{\gamma} \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\circlearrowleft}, \hat{\mathsf{s}}_h) + (\hat{\gamma} + \gamma_\sigma) \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\circlearrowleft}, \mathsf{s}_h^{\mathrm{tel}}) + \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^{\mathrm{tel}})) \\
&= \hat{\gamma} \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\circlearrowleft}, \hat{\mathsf{s}}_h) + (\hat{\gamma} + \gamma_\sigma) \circ \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\circlearrowleft}, \mathsf{s}_h^{\mathrm{tel}}) + \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}))
\end{aligned}$$

This concludes the inductive construction.

For the second statement, notice that the events $\bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1}$ are $\mathcal{F}_h$ measurable (thus determined by $\mu^{(h-1)}$) and, when they hold, $\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_h^{\circlearrowleft}, \mathsf{s}_h^{\mathrm{tel}}) \preceq \vec{\gamma}_{\mathrm{IPS}}(2r)$ and $\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h^{\circlearrowleft}, \hat{\mathsf{s}}_h) \preceq \vec{\varepsilon}$. For our purposes, we use $\mathsf{d}_{\mathrm{TVC}} = \mathsf{d}_{\mathcal{S},1}(\mathsf{s}_h^{\circlearrowleft}, \mathsf{s}_h^{\mathrm{tel}}) \preceq \gamma_{\mathrm{IPS},1}(2r)$ and $\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h^{\circlearrowleft}, \hat{\mathsf{s}}_h) \preceq \vec{\varepsilon}_1$. Hence,

$$\begin{aligned}
\max_{h \in [H]} \mathbb{P}_\mu[\mathcal{B}^c_{\mathrm{all},h} \cap \bar{\mathcal{C}}_{\mathrm{all},h} \cap \bar{\mathcal{B}}_{\mathrm{all},h-1}] \le \hat{\gamma}(\vec{\varepsilon}_1) + (\hat{\gamma} + \gamma_\sigma) \circ \gamma_{\mathrm{IPS},1}(2r) \\
+ \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^{\mathrm{tel}}) \parallel \pi^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^{\mathrm{tel}})).
\end{aligned}$$

The result follows.

E.3. Proof of Theorem 2, and generalization to direct decompositions

In this subsection, we consider the special case dealt with in Theorem 2. Note that there always exists a trivial direct decomposition that is compatible with all policies and dynamics simply by letting $\mathcal{V} = \emptyset$ and $\mathcal{S} = \mathcal{Z}$. We prove here the version of the result that involves a possibly nontrivial direct decomposition, as we will instantiate this in our control setting by letting $\mathcal{Z} = \{\rho_{\mathrm{m},h}\}$ and $\mathcal{S} = \{\rho_{\mathrm{c},h}\}$, i.e., projecting $\rho_{\mathrm{c},h}$ onto the last τ_{m} coordinates gives z_h. We further consider a restriction of IPS to consider kernels absolutely continuous with respect to $\mathsf{P}_h^\star$ in their $\mathcal{Z}$ component.

Definition E.9 (Restricted IPS)**.** For a non-decreasing maps $\gamma_{\mathrm{IPS},1}, \gamma_{\mathrm{IPS},2} : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ a pseudometric $\mathsf{d}_{\mathrm{IPS}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ (possibly other than $\mathsf{d}_{\mathcal{S}}$ or $\mathsf{d}_{\mathrm{TVC}}$), and $r_{\mathrm{IPS}} > 0$, we say a policy π is $(\gamma_{\mathrm{IPS},1}, \gamma_{\mathrm{IPS},2}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-*restricted IPS* if the following holds for any $r \in [0, r_{\mathrm{IPS}}]$. Consider any sequence of kernels $\mathsf{W}_1, \dots, \mathsf{W}_H : \mathcal{S} \to \Delta(\mathcal{S})$ satisfying

$$\max_{h, \mathsf{s} \in \mathcal{S}} \mathbb{P}_{\tilde{\mathsf{s}} \sim \mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\tilde{\mathsf{s}}, \mathsf{s}) \le r] = 1, \quad \forall s, \quad \phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s}_h) \ll \phi_{\mathcal{Z}} \circ \mathsf{P}_h^\star.$$

and define a process $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h), \mathsf{a}_h \sim \pi_h(\tilde{\mathsf{s}}_h)$, and $\mathsf{s}_{h+1} := F_h(\mathsf{s}_h, \mathsf{a}_h)$. Then, almost surely, (a) the sequence $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ is input-stable w.r.t $(\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathcal{A}})$ (b) $\max_{h \in [H]} \mathsf{d}_{\mathrm{TVC}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \le \gamma_{\mathrm{IPS},1}(r)$ and (c) $\max_{h \in [H]} \mathsf{d}_{\mathcal{S}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \le \gamma_{\mathrm{IPS},2}(r)$.

Note that the above is a slightly weaker condition than the one in Definition 4.5 in the main text and consequently, the following theorem which uses it as an assumption implies Theorem 2 in the body.

Theorem 5. *Suppose $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$ as in Definition E.1 and projections $\phi_{\mathcal{Z}}, \phi_{\mathcal{V}}$, which is compatible with the dynamics and with given policies $\hat{\pi}, \pi^\star$, smoothing kernel W_σ, and pseudometric $\mathsf{d}_{\mathrm{IPS}}$. Suppose $\pi^\star$ satisfies $(\gamma_{\mathrm{IPS},1}, \gamma_{\mathrm{IPS},2}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-restricted IPS (Definition E.9) and $\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma$ is γ_σ-TVC. Let $\varepsilon > 0$, $r \in (0, \frac{1}{2} r_{\mathrm{IPS}}]$; define $p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}', \mathsf{s}) > r]$*

and $\varepsilon' := \varepsilon + \gamma_{\text{IPS},2}(2r)$*. Then, for any policy* $\hat{\pi}$*, both* $\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\circlearrowleft})$ *and* $\Gamma_{\text{marg},\varepsilon'}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ *are upper bounded by*

$$H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\text{IPS},1}(2r)\})\right) + \sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\text{dec}}(\tilde{\mathsf{s}}^\star_h)).$$

Consider the special case $K = 2$ with $\mathsf{d}_{\mathcal{S},1} = \mathsf{d}_{\text{TVC}}$, $\mathsf{d}_{\mathcal{S},2} = \mathsf{d}_{\mathcal{S}}$, $\mathsf{d}_{\mathcal{A},1} = \mathsf{d}_{\mathcal{A},2} = \mathsf{d}_{\mathcal{A}}$ and $\vec{\varepsilon} = (\varepsilon, \varepsilon)$. In this case, applying (E.3), we see that

$$\begin{aligned}
&\vec{\Gamma}_{\text{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft}) \vee \vec{\Gamma}_{\text{marg},\vec{\varepsilon}_{\text{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft}) \\
&\quad \le p_{\text{IPS}} + H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\text{IPS},1}(2r)\}\right) + \sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \vec{\mathsf{d}}_{\text{os},\vec{\varepsilon}}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\text{dec}}(\tilde{\mathsf{s}}^\star_h))
\end{aligned}$$

We now observe that under this convention,

$$\begin{aligned}
\Gamma_{\text{joint},\varepsilon}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft}) &= \inf_{\mu_1} \mathbb{P}_{\mu_1}[\max_{h\in[H]} \mathsf{d}_{\mathcal{S}}(\hat{\mathsf{s}}_{h+1}, \mathsf{s}^\star_{h+1}) \vee \mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) > \varepsilon] \\
&\le \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\max_{h\in[H]} \left(\mathsf{d}_{\text{TVC}}(\hat{\mathsf{s}}_{h+1}, \mathsf{s}^\star_{h+1}), \mathsf{d}_{\mathcal{S}}(\hat{\mathsf{s}}_{h+1}, \mathsf{s}^\star_{h+1})\right) \vee \left(\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h), \mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h)\right) \not\preceq \vec{\varepsilon}\right] \\
&= \vec{\Gamma}_{\text{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\circlearrowleft})
\end{aligned}$$

and similarly $\Gamma_{\text{marg},\varepsilon'}(\hat{\pi}_\sigma \parallel \pi^\star) \le \vec{\Gamma}_{\text{marg},\vec{\varepsilon}+\gamma_{\text{IPS}}(2r)}(\hat{\pi}_\sigma \parallel \pi^\star)$. From the construction of $\vec{\mathsf{d}}_{\mathcal{A}}$, however, we see that $\left\{\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') \not\preceq \vec{\varepsilon}\right\} = \{\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') > \varepsilon\}$ for all a, a' and thus for all $h \in [H]$,

$$\begin{aligned}
\vec{\mathsf{d}}_{\text{os},\vec{\varepsilon}}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_h(\tilde{\mathsf{s}}^\star_h)) &= \inf_{\mu_2} \mathbb{P}_{\mu_2}\left[\vec{\mathsf{d}}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) \not\preceq \vec{\varepsilon}\right] \\
&= \inf_{\mu_2} \mathbb{P}_{\mu_2}\left[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) \ge \varepsilon\right] \\
&= \mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_h(\tilde{\mathsf{s}}^\star_h)).
\end{aligned}$$

Plugging in to (E.3) concludes the proof.

F. Lower Bounds

In this section, we establish lower bounds against the imitation results in the composite MDP. Specifically, we show that

- In Appendix F.1 we show that Theorem 2 and Proposition D.1 are sharp in the regime where $\gamma_{\text{IPS},1} = \gamma_{\text{IPS},2} = 0$.
- In Appendix F.2, we show that the marginals of an expert policy $\pi^\star$ and replica policy $\pi^\star_{\circlearrowleft\sigma}$ can coincide, but their joint distributions can be different. By considering $\hat{\pi} = \pi^\star_{\text{dec}}$ in Theorem 2, this establishes the necessity of considering the marginal imitation gap with respect to $\pi^\star$.
- In Appendix F.3, we lower bound the distance between *marginal distributions* over states under $\pi^\star$ and $\pi^\star_{\circlearrowleft\sigma}$ in the regime where $\gamma_{\text{IPS},2} \ne 0$. This example demonstrates that the dependence of $\gamma_{\text{IPS},2}$ in Theorem 2 is essentially sharp.
- In Appendix F.4, we show that for an expert policy $\pi^\star$ and smoothing kernel W_σ, the state distributions under $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star_{\text{dec}}$ can have different marginals (and thus different joint distributions). By considering $\hat{\pi} = \pi^\star_{\text{dec}}$ in Theorem 2, this explains why it is necessary to smooth $\hat{\pi}$ to $\hat{\pi} \circ \mathsf{W}_\sigma$.

Taken together, the above counterexamples show that our distinctions between joint and marginal distributions, decision to add noise at inference time, and dependence on almost all problem quantities in Section 4 are sharp. We do not, however, establish necessity of $\gamma_{\text{IPS},1}$ in the interest of brevity; we believe this quantity is necessary. Still, the $\gamma_{\text{IPS},1}$ term contributes a factor exponentially small in τ_{c} in Theorem 1, so we deem lower bounds establishing its necessity of lesser importance.

Commonalities of construction. In all but Appendix F.3, we take the action and state spaces to be

$$\mathcal{S} = \mathcal{A} = \mathbb{R},$$

which is the archetypal Polish space (Durrett, 2019). Throughout, we use δ_x to denote the dirac-delta distribution on $x \in \mathbb{R}$. We let $\mathsf{d}_{\mathcal{S}}(\mathsf{s}', \mathsf{s}) = \mathsf{d}_{\text{TVC}}(\mathsf{s}', \mathsf{s}) = |\mathsf{s}' - \mathsf{s}|$ and $\mathsf{d}_{\mathcal{A}}(\mathsf{a}', \mathsf{a}) = |\mathsf{a}' - \mathsf{a}|$ all be the Euclidean distance.

F.1. Sharpness of Proposition D.1 and Theorem 2

Here, we demonstrate that Proposition D.1 is tight up to constant factors, and that Theorem 2 is tight up to the terms $\gamma_{\text{IPS},1}, \gamma_{\text{IPS},2}$ and concentration probability p_r. Consider the simple dynamics

$$F_h(\mathsf{s}, \mathsf{a}) = \mathsf{a}.$$

Note that, as the dynamics are state-independent, we have $\gamma_{\text{IPS},1}(\cdot) = \gamma_{\text{IPS},2}(\cdot) \equiv 0$. Furthermore, let us assume policies do not depend on time index h. Let $\pi^\star : \mathsf{s} \to \delta_0$ be deterministic, and let $\mathsf{P}_{\text{init}} = \delta_0$ be an initial state distribution concentrated on 0. Then, $\mathsf{D}_{\pi^\star}$ is the dirac distribution on the all-zero trajectory.

Fix parameters $0 < \varepsilon < \sigma$, and $p \in (0, 1)$. We consider the following smoothing-kernel

$$\mathsf{W}_{\varepsilon,\sigma} = \begin{cases} \delta_0 & \mathsf{s} \le 0 \\ (1 - \frac{s}{\sigma})\delta_0 + \frac{s}{\sigma}\delta_\sigma & \mathsf{s} \in [0, \sigma] \\ \delta_\sigma & \mathsf{s} > \sigma, \end{cases}$$

Define the candidate policy

$$\hat{\pi}_{\varepsilon,p,\sigma}(\mathsf{s}) := \begin{cases} (1 - p)\delta_\varepsilon + p\delta_\sigma & \mathsf{s} \le \frac{\varepsilon}{2} \\ \delta_\sigma & \mathsf{s} > \frac{\varepsilon}{2} \end{cases}$$

Proposition F.1. *For any $p \in (0, 1)$, $0 < \varepsilon < \sigma$, set $\bar{\pi} = \hat{\pi}_{\varepsilon,p,\sigma} \circ \mathsf{W}_{\sigma,\varepsilon}$. Then,*

(a) $\pi^\star$, $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star_{\text{dec}}$ all map $\mathsf{s} \to \delta_0$, $\mathsf{P}^\star_h = \delta_0$, and thus for any $\tilde{\pi} \in \{\pi^\star, \pi^\star_{\circlearrowleft\sigma}, \pi^\star_{\text{dec}}\}$,

$$\mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{s'_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)}[\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_{\varepsilon,p,\sigma}(s'_h) \parallel \tilde{\pi}(s'_h)] = \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}[\mathsf{d}_{\text{os},\varepsilon}(\bar{\pi}(\mathsf{s}^\star_h) \parallel \tilde{\pi}(\mathsf{s}^\star_h))] = p.$$

(b) The kernel $\mathsf{W}_{\sigma,\varepsilon}$ is γ_σ-TVC, where $\gamma_\sigma(u) = u/\sigma$.

(c) For a universal constant $c > 0$,

$$\Gamma_{\text{joint},\varepsilon}(\bar{\pi} \parallel \pi^\star) = \Gamma_{\text{marg},\varepsilon}(\bar{\pi} \parallel \pi^\star) \ge c \min\{1, H(p + \varepsilon/\sigma)\},$$

and the same holds with $\pi^\star$ replaced by $\pi^\star_{\circlearrowleft\sigma}$ or $\pi^\star_{\text{dec}}$.

In particular, the above proposition shows that

$$\Gamma_{\text{joint},\varepsilon}(\bar{\pi} \parallel \pi^\star) = \Gamma_{\text{marg},\varepsilon}(\bar{\pi} \parallel \pi^\star) \gtrsim H\gamma_\sigma(\varepsilon) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}[\mathsf{d}_{\text{os},\varepsilon}(\bar{\pi}(\mathsf{s}^\star_h) \parallel \pi^\star(\mathsf{s}^\star_h))],$$

verifying the sharpness of Proposition D.1 (note that $\bar{\pi} = \hat{\pi}_{\varepsilon,p,\sigma} \circ \mathsf{W}_\sigma$ is γ_σ TVC). Similary, our above proposition shows that,

$$\Gamma_{\text{joint},\varepsilon}(\bar{\pi} \parallel \pi^\star_{\circlearrowleft\sigma}) = \Gamma_{\text{marg},\varepsilon}(\bar{\pi} \parallel \pi^\star) \gtrsim H\gamma_\sigma(\varepsilon) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}[\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_{\varepsilon,p,\sigma}(\mathsf{s}^\star_h) \parallel \pi^\star_{\text{dec},h}(\mathsf{s}^\star_h)],$$

verying that Theorem 2 is sharp up to the additional stability terms $\gamma_{\text{IPS},1}, \gamma_{\text{IPS},2}$.

Proof. We begin with a computation. Define

$$\eta(\mathsf{s}) = 1 - (1 - p)(1 - \frac{\mathsf{s}}{\sigma}) = p + (1 - p)\frac{\mathsf{s}}{\sigma}$$

We compute

$$
\begin{aligned}
\bar{\pi} = \hat{\pi}_{\varepsilon,p,\sigma} \circ \mathsf{W}_{\sigma,\varepsilon} &= \begin{cases} (1-p)\delta_\varepsilon + p\delta_\sigma & \mathsf{s} \le \frac{\varepsilon}{2} \\ \delta_\sigma & \mathsf{s} > \frac{\varepsilon}{2} \end{cases} \circ \begin{cases} \delta_0 & \mathsf{s} \le 0 \\ (1-\frac{s}{\sigma})\delta_0 + \frac{s}{\sigma}\delta_\sigma & \mathsf{s} \in [0,\sigma] \\ \delta_\sigma & \mathsf{s} \ge \sigma. \end{cases} \\
&= \begin{cases} (1-p)\delta_\varepsilon + p\delta_\sigma & \mathsf{s} \le 0 \\ (1-\eta(s))\delta_\varepsilon + \eta(s)\delta_\sigma & 0 \le \mathsf{s} \le \sigma \\ \delta_\sigma & \mathsf{s} > \sigma. \end{cases} \qquad \text{(F.1)}
\end{aligned}
$$

In particular,

$$
\hat{\pi}(0) = \pi_{\varepsilon,p,\sigma}(0) = (1-p)\delta_\varepsilon + p\delta_\sigma
$$

Part (a). Notice that the support of the deconvolution and replica distributions are always in the support of $\mathsf{P}_h^\star$, which is always $\mathsf{s} = 0$ under $\pi^\star$. Thus, $\pi^\star = \pi^\star_{\circlearrowleft\sigma} = \pi^\star_{\mathrm{dec}}$. By the same token, for any policy π,

$$
\mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star}[\mathsf{d}_{\mathrm{os},\varepsilon}(\pi(\mathsf{s}_h^\star) \parallel \tilde{\pi}_\star(\mathsf{s}_h^\star)] = \mathbb{P}[|\pi(0)| > \varepsilon].
$$

Hence, as $\bar{\pi}(0) = \hat{\pi}_{\varepsilon,p,\sigma}(0) = (1-p)\delta_\varepsilon + p\delta_\sigma$, and as $\sigma > \varepsilon$, part (a) follows.

Part (b). Consider $\mathsf{s}, \mathsf{s}' \in \mathcal{S}$. We can assume, from the functional form of $\mathsf{W}_{\varepsilon,\sigma}(\cdot)$, that $0 \le \mathsf{s} \le \mathsf{s}' \le \sigma$. Then,

$$
\mathsf{TV}(\mathsf{W}_{\varepsilon,\sigma}(\mathsf{s}), \mathsf{W}_{\varepsilon,\sigma}(\mathsf{s}')) = \mathsf{TV}(\delta_0(1 - \frac{\mathsf{s}}{\sigma}) + (\frac{\mathsf{s}}{\sigma})\delta_\sigma, \delta_0(1 - \frac{\mathsf{s}'}{\sigma}) + (\frac{\mathsf{s}'}{\sigma})\delta_\sigma = \frac{|\mathsf{s}' - \mathsf{s}|}{\sigma},
$$

establishing total variation continuity.

Part (c) In view of part (a), it suffices to bound gaps relative to $\pi^\star$. Let $\mathbb{P}$ denote probabilities over $\mathsf{s}_{1:H+1}, \mathsf{a}_h$ under $\bar{\pi}$. Let $\mathcal{A}_{1,h}$ denote the event that at step h, $\mathsf{a}_h = \varepsilon$, and let $\mathcal{A}_{2,h}$ denote the event that $\mathsf{a}_h = \sigma$. As the state s_0 is absoring and as $F_h(\mathsf{s}, \mathsf{a}) = \mathsf{a}_h$, the following events are equal

$$
\{\exists h : |\mathsf{a}_h| \vee |\mathsf{s}_{h+1}| > \varepsilon\} = \mathcal{A}_{2,H}.
$$

Hence,

$$
\Gamma_{\mathrm{joint},\varepsilon}(\bar{\pi} \parallel \pi^\star) = \mathbb{P}[\mathcal{A}_{2,H}].
$$

Moreover, as $\mathcal{A}_{2,H}$ is measurable with respect to the marginal of a_H, we also have that

$$
\Gamma_{\mathrm{marg},\varepsilon}(\bar{\pi} \parallel \pi^\star) = \mathbb{P}[\mathcal{A}_{2,H}].
$$

It thus suffices to lower bound $\mathbb{P}[\mathcal{A}_{2,H}]$. By definition of $\bar{\pi}$, the events $\mathcal{A}_{1,h}, \mathcal{A}_{2,h}$ are exhaustive: $\mathcal{A}_{1,h}^c = \mathcal{A}_{2,h}$. Moreover, from (F.1),

$$
\mathbb{P}[\mathcal{A}_{2,h+1} \mid \mathcal{A}_{2,h}] = 1, \quad \mathbb{P}[\mathcal{A}_{2,h+1} \mid \mathcal{A}_{1,h}] = \eta(\varepsilon), \quad \mathbb{P}[\mathcal{A}_{1,1}] = 1 - \eta(0) \ge 1 - \eta(\varepsilon).
$$

Thus,

$$
\begin{aligned}
\mathbb{P}[\mathcal{A}_{2,H}] &= \mathbb{P}[\mathcal{A}_{2,H} \mid \mathcal{A}_{2,H-1}]\,\mathbb{P}[\mathcal{A}_{2,H-1}] + \mathbb{P}[\mathcal{A}_{2,H} \mid \mathcal{A}_{1,H-1}]\,\mathbb{P}[\mathcal{A}_{1,H-1}] \\
&= \mathbb{P}[\mathcal{A}_{2,H-1}] + \eta(\varepsilon)\,\mathbb{P}[\mathcal{A}_{1,H-1}] \\
&= \mathbb{P}[\mathcal{A}_{2,H-2}] + \eta(\varepsilon)\left(\mathbb{P}[\mathcal{A}_{1,H-1} + \mathbb{P}[\mathcal{A}_{1,H-2}]\right) \\
&= \eta(\varepsilon)\left(\sum_{h=1}^{H-1} \mathbb{P}[\mathcal{A}_{1,h}]\right) + \mathbb{P}[\mathcal{A}_{2,1}] \\
&\ge \eta(\varepsilon)\left(\sum_{h=1}^{H-1} \mathbb{P}[\mathcal{A}_{1,h}]\right)
\end{aligned}
$$

Moreover, as s_0 is absorbing,

$$\mathbb{P}[\mathcal{A}_{1,h}] = \mathbb{P}[\mathcal{A}_{1,h} \mid \mathcal{A}_{1,h-1}]\,\mathbb{P}[\mathcal{A}_{1,h-1}] = (1-\eta(\varepsilon))\,\mathbb{P}[\mathcal{A}_{1,h-1}].$$

Combining with $\mathbb{P}[\mathcal{A}_{1,1}] = (1-p) \geq (1-\eta(0)) \geq 1-\eta(\varepsilon)$, we have $\mathbb{P}[\mathcal{A}_{1,h}] \geq (1-\eta(\varepsilon))^h$. Hence,

$$\begin{aligned}
\mathbb{P}[\mathcal{A}_{2,H+1}] &\geq \eta(\varepsilon)\left(\sum_{h=1}^{H-1}(1-\eta(\varepsilon))^h\right)\\
&= \eta(\varepsilon)\frac{1-\eta(\varepsilon)-(1-\eta(\varepsilon))^H}{1-(1-\eta(\varepsilon))}\\
&= 1-\eta(\varepsilon)-(1-\eta(\varepsilon))^H\\
&= \Omega(\min\{1, H(\eta(\varepsilon))\})
\end{aligned}$$

as $\eta(\varepsilon) \downarrow 0$. Subsituting in $\eta(\varepsilon) = p + (1-p)\varepsilon/\sigma = \Omega(p+\varepsilon/\sigma)$ concludes. $\square$

F.2. $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star$ induce the same marginals but different joint distributions, even with memoryless dynamics

We give a simple example where $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star$ induce the same marginal distributions over trajectories, but different joints. As we show, this example demonstrates the necessity of measuring the marginal imitation error of a smoothed policy, $\Gamma_{\mathrm{marg},\varepsilon}$, over the joint error, $\Gamma_{\mathrm{joint},\varepsilon}$. A graphical (but nonrigorous) demonstration of this issue can be seen in Figure 4 in Appendix B.

Again, let $\mathcal{S} = \mathcal{A} = \mathbb{R}$, and $F_h(\mathsf{s},\mathsf{a}) = \mathsf{a}$. We let

$$\mathsf{W}_\sigma(\cdot) = \mathcal{N}(\cdot, \sigma^2)$$

denote Gaussian smoothing. Fix some $\varepsilon > 0$. Define

$$\mathsf{P}_{\mathrm{init}} = \frac{1}{2}(\updelta_{-\varepsilon} + \updelta_{+\varepsilon}), \quad \pi^\star(\mathsf{s}) = \begin{cases}\updelta_{-\varepsilon} & \mathsf{s} \leq 0\\ \updelta_{\varepsilon} & \mathsf{s} > 0\end{cases}.$$

Thus, $\mathsf{D}_{\pi^\star}$ is supported on the trajectories with $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ being either all ε or all $-\varepsilon$, and

$$\mathsf{P}^\star_h = \mathsf{P}_{\mathrm{init}} = \frac{1}{2}(\updelta_{-\varepsilon} + \updelta_{+\varepsilon}).$$

Hence, the replica and deconvolution map to distributions supported on $\{\varepsilon, -\varepsilon\}$. Let $\phi_\sigma(\cdot)$ denote the Gaussian PDF with variance σ. Then,

$$\mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}) = \frac{\updelta_\varepsilon\phi_\sigma(\mathsf{s}-\varepsilon) + \updelta_{-\varepsilon}\phi_\sigma(\mathsf{s}+\varepsilon)}{\phi_\sigma(\mathsf{s}-\varepsilon)+\phi_\sigma(\mathsf{s}+\varepsilon)}.$$

Moreover,

$$\mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}) = \mathbb{E}_{Z\sim\mathcal{N}(0,\sigma^2)}\left[\frac{\updelta_\varepsilon\phi_\sigma(\mathsf{s}-\varepsilon+Z) + \updelta_{-\varepsilon}\phi_\sigma(\mathsf{s}+\varepsilon+Z)}{\phi_\sigma(\mathsf{s}-\varepsilon+Z)+\phi_\sigma(\mathsf{s}+\varepsilon+Z)}\right]. \tag{F.2}$$

One can check that for $\varepsilon \leq \sigma$,

$$\mathsf{W}^\star_{\circlearrowleft,h}(u\varepsilon) = \Theta\left(\frac{(1+\frac{c\varepsilon}{\sigma})\updelta_{u\varepsilon} + (1-\frac{c\varepsilon}{\sigma})\updelta_{-u\varepsilon}}{2}\right), \quad u \in \{-1, 1\}$$

for $\varepsilon \ll 1$. In particular, for $\mathsf{s} \in \{-\varepsilon, \varepsilon\}$

$$\mathbb{P}_{\mathsf{a}\sim\boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s})}[\mathsf{a} = -\mathsf{s}] \geq \Omega(1). \tag{F.3}$$

In particular, if $(\mathsf{s}^{\circlearrowleft}_{1:H+1}, \mathsf{a}^{\circlearrowleft}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}}$, then

$$\begin{aligned}\mathbb{P}[\exists h : \mathsf{d}(\mathsf{s}^{\circlearrowleft}_h, \mathsf{s}^{\circlearrowleft}_{h+1}) > \varepsilon] &\le \mathbb{P}[\exists h : \mathsf{s}^{\circlearrowleft}_h = -\mathsf{s}^{\circlearrowleft}_{h+1}] \\ &\le \mathbb{P}[\exists h : \mathsf{s}^{\circlearrowleft}_h = -\mathsf{a}^{\circlearrowleft}_h] = 1 - \exp(-\Omega(H)),\end{aligned}$$

where in the last step we used (F.3) and the the fact that the $\pi^\star_{\circlearrowleft\sigma}$ uses fresh randomness at each round. Moreover, as $\pi^\star$ always commits to either an all-ε or all-$(-\varepsilon)$-trajectory, we see that for any $\mu \in \mathscr{C}(\mathsf{D}_{\pi^\star}, \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}})$ over $(\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H}) \sim \mathsf{D}_{\pi^\star}$ and $(\mathsf{s}^{\circlearrowleft}_{1:H+1}, \mathsf{a}^{\circlearrowleft}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}}$,

$$\Gamma_{\mathrm{joint},\varepsilon}(\pi^\star_{\circlearrowleft\sigma}, \pi^\star) \ge \mathbb{P}_\mu[\exists 1 \le h \le H : \mathsf{d}(\mathsf{s}^\star_{h+1}, \mathsf{s}^{\circlearrowleft}_{h+1}) > \varepsilon] \ge 1 - \exp(-\Omega(H)),$$

That is, the replica and expert policies have different joint state distribution.

Remark F.1. The above result demonstrates the necessity of measuring the marginal error between $\hat{\pi} \circ \mathsf{W}_\sigma$ and $\pi^\star$ in Theorem 2: if we apply that proposition with $\hat{\pi} = \pi^\star_{\mathrm{dec}}$, then for all ε, $\mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h)) = 0$. But then $\hat{\pi} \circ \mathsf{W}_\sigma = \pi^\star_{\circlearrowleft\sigma}$, and we know that $\Gamma_{\mathrm{joint},\varepsilon}(\pi^\star_{\circlearrowleft\sigma}, \pi^\star) \ge \mathbb{P}_\mu[\exists 1 \le h \le H : \mathsf{d}(\mathsf{s}^\star_{h+1}, \mathsf{s}^{\circlearrowleft}_{h+1}) > \varepsilon] \ge 1 - \exp(-\Omega(H))$. Thus, we cannot hope for smoothed policies to imitate expert demonstrations in joint state distributions without additional assumptions.

Remark F.2 (Importance of chunking)**.** Above we have shown that $\pi^\star_{\circlearrowleft\sigma}$ oscillates between ε and $-\varepsilon$ (for actions and subsequent states). We remark that these oscillations can have very deleterious effects on performance on real control systems. This is why it is beneficial to predict entire sequences of trajectories. Indeed, consider a modified construction such that $\mathcal{S} = \mathcal{A} = \mathbb{R}^K$, and $F_h(\mathsf{s}, \mathsf{a}) = \mathsf{a}$. Here, we interpret $\mathcal{S}$ as a sequence of K-control states in $\mathbb{R}$, and a as sequence of K-actions, denoting the i-th coordinate of s via $\mathsf{s}[i]$,

$$\pi^\star(\mathsf{s}) = \begin{cases} \delta_{-\varepsilon\mathbf{1}} & \mathsf{s}[1] \le 0 \\ \delta_{\varepsilon\mathbf{1}} & \mathsf{s}[1] > 0, \end{cases}$$

Then, we can view the oscillations in $\pi^\star_{\circlearrowleft\sigma}$ as oscillations between length K trajectories, which is essentially what happens in our analysis for $K = \tau_c$.

F.3. $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star$ can have different marginals, implying necessity of $\gamma_{\mathrm{IPS},2}$

Our construction lifts the construction in Appendix F.2 to a two-dimensional state space $\mathcal{S} = \mathbb{R}^2$, keeping one dimensional actions $\mathcal{A} = \mathbb{R}$. Let $\mathsf{s} = (\mathsf{s}[1], \mathsf{s}[2])$ denote coordinate of $\mathsf{s} \in \mathcal{S}$. For some parameter ν, the dynamics are

$$\mathsf{s}_{h+1} = F_h(\mathsf{s}_h, \mathsf{a}_h) = (\mathsf{a}_h, \nu \cdot (\mathsf{s}_h[1] - \mathsf{a}_h))$$

We let $\mathsf{d}_{\mathcal{S}} = \mathsf{d}_{\mathrm{TVC}} = \mathsf{d}_{\mathrm{IPS}}$ denote the ℓ_1 norm on $\mathcal{S} = \mathcal{R}^2$. Our initial state distribution is

$$\mathsf{P}_{\mathrm{init}} = \frac{1}{2}\left(\delta_{(\varepsilon,0)} + \delta_{(-\varepsilon,0)}\right)$$

We let

$$\pi^\star(\mathsf{s}) = \begin{cases} \delta_{(-\varepsilon,0)} & \mathsf{s} \le 0 \\ \delta_{(\varepsilon,0)} & \mathsf{s} > 0 \end{cases}.$$

Thus, $\pi^\star$ induces trajectories which either stay on $\delta_{(\varepsilon,0)}$ or $\delta_{(-\varepsilon,0)}$.

$$\mathsf{P}^\star_h = \frac{1}{2}\left(\delta_{(\varepsilon,0)} + \delta_{(-\varepsilon,0)}\right), \quad \forall h \ge 1.$$

Let

$$\mathsf{W}_\sigma(\mathsf{s}) = \mathcal{N}(\mathsf{s}', \sigma^2)$$

Proposition F.2. *In the above construction, we can take $\gamma_{\mathrm{IPS},2}(u) \le \nu \cdot u$ in Definition 4.5, and p_r satisfies the conditions in Theorem 2 for $r = 2\sigma\sqrt{\log(1/p_r)}$. Moreover, for any $\varepsilon \le \sigma$,*

$$\Gamma_{\mathrm{marg},\varepsilon'}(\pi^\star_{\circlearrowleft\sigma} \parallel \pi^\star) \ge \Omega(1), \quad \varepsilon' = \nu\varepsilon$$

Remark F.3 (Sharpness of $\gamma_{\mathrm{IPS},2}$)**.** Before proving this proposition, we note that if we take $\varepsilon = \sigma$ and $r = 2\sigma\sqrt{\log(1/p_r)}$, then $\nu\varepsilon = \tilde{\Omega}(\gamma_{\mathrm{IPS}}(2r))$, showing that our dependence on $\gamma_{\mathrm{IPS},2}$ is sharp up to logarithmic factors. Moreover, the looseness up to logarithmic factors in the above point is an artifact of using the Gaussian smoothing W_σ, and can be remover by replaced W_σ with a truncated-Gaussian kernel.

Proof of Proposition F.2. To see $\gamma_{\mathrm{IPS},2}(u) \le \nu \cdot u$, we have $\|F_h(\mathsf{s},\mathsf{a}) - F_h(\mathsf{s}',\mathsf{a})\| = \|(\mathsf{a}, \nu\cdot(\mathsf{s}[1]-\mathsf{a})) - (\mathsf{a}, \nu\cdot(\mathsf{s}'[1]-\mathsf{a}))\| = \nu|\mathsf{s}[1] - \mathsf{s}'[1]| \le \nu \mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}')$. That we can take $r = 2\sigma\sqrt{\log(1/p_r)}$ follows from Gaussian concentration.

To prove the final claim, one can directly generalize (F.2) to find that, for any $b \in \mathbb{R}$,

$$\mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s}) = \mathbb{E}_{Z\sim\mathcal{N}(0,\sigma^2)}\left[\frac{\delta_{(\varepsilon,0)}\phi_\sigma(\mathsf{s}[1] - \varepsilon + Z) + \delta_{(-\varepsilon,0)}\phi_\sigma(\mathsf{s}[1] + \varepsilon + Z)}{\phi_\sigma(\mathsf{s}[1] - \varepsilon + Z) + \phi_\sigma(\mathsf{s}[1] + \varepsilon + Z)}\right].$$

This follows form the observation that $\mathsf{W}^\star_{\circlearrowleft,h}$ and $\mathsf{P}^\star_h$ have the same support, and as $\mathsf{P}^\star_h$ always is support on vectors with second coordinate zero, that the second coordinate of s in $\mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s})$ is uninformative. For $\varepsilon \le \sigma$, we find that

$$\mathsf{W}^\star_{\circlearrowleft,h}((\varepsilon, b)) = c\delta_{(\varepsilon,0)} + (1-c)\delta_{(-\varepsilon,0)}, c = \Omega(1), b \in \mathbb{R}.$$

and $\mathsf{W}^\star_{\circlearrowleft,h}((-\varepsilon, b))$ is defined symmetrically, Hence, under $(\mathsf{s}^{\circlearrowleft}_{1:H+1}, \mathsf{a}^{\circlearrowleft}_{1:H}) \sim \pi^\star_{\circlearrowleft\sigma}$,

$$\mathbb{P}[\mathsf{s}^{\circlearrowleft}_1 \ne \mathsf{a}^{\circlearrowleft}_1] \ge \Omega(1)$$

Moroever, when $\mathsf{s}^{\circlearrowleft}_2 \ne \mathsf{a}^{\circlearrowleft}_h$, we have that $|\mathsf{s}^{\circlearrowleft}_2[2]| = \nu|\mathsf{s}^{\circlearrowleft}_1 - \mathsf{a}^{\circlearrowleft}_1|$, which as $\pi^\star$ is supported on $\{\delta_{(\varepsilon,0)}, \delta_{(-\varepsilon,0)}\}$, means, $|\mathsf{s}^{\circlearrowleft}_2(2)| \ge 2\nu\varepsilon$. Thus,

$$\mathbb{P}[|\mathsf{s}^{\circlearrowleft}_2[2]| \ge 2\nu\varepsilon] \ge \Omega(1)$$

On the other hand, $\mathsf{s}^\star_2 \sim \mathsf{P}^\star_h$ has $\mathsf{s}^\star_2[2] = 0$ with probability one. Thus, for any coupling μ between $\mathsf{D}_{\pi^\star}, \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}}$,

$$\mathbb{P}_\mu[\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_2, \mathsf{s}^\star_2)| \ge 2\nu\varepsilon] \ge \Omega(1)$$

Thus,

$$\Gamma_{\mathrm{marg},\nu\varepsilon}(\pi^\star_{\circlearrowleft\sigma} \parallel \pi^\star) \ge \Omega(1).$$

□

F.4. $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star_{\mathrm{dec}}$ have different marginals, even with memoryless dynamics

Here, we show how $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star_{\mathrm{dec}}$ have different marginals even if the dynamics are memoryless. By considering $\hat{\pi} = \pi^\star_{\mathrm{dec}}$ in Theorem 2, the discussion below demonstrates why one needs to consider $\hat{\pi} \circ \mathsf{W}_\sigma$ in order to obtain small imitation gap.

For simplicity, we use a discrete smoothing kernel W_σ, though the example extends to the Gaussian smoothing kernel in the previous counter example. Again, let $\mathcal{S} = \mathcal{A} = \mathbb{R}$, and $F_h(\mathsf{s},\mathsf{a}) = \mathsf{a}$. Take

$$\pi^\star(\mathsf{s}) = \begin{cases} \delta_{-\sigma} & \mathsf{s} \le 0 \\ \delta_\sigma & \mathsf{s} > 0 \end{cases}$$

Let us consider an asymmetric initial state distribution

$$\mathsf{P}_{\mathrm{init}} = \frac{1}{4}\delta_{-\sigma} + \frac{3}{4}\delta_{+\sigma}.$$

Note then that

$$\forall h, \quad \mathsf{P}^\star_h = \mathsf{P}_{\mathrm{init}} = \frac{1}{4}\delta_{-\sigma} + \frac{3}{4}\delta_\sigma, \tag{F.4}$$

We consider a smoothing kernel,

$$\mathsf{W}_\sigma(\mathsf{s}) = \begin{cases} (\frac{1}{2} + \frac{\mathsf{s}}{4\sigma})\delta_\sigma + (\frac{1}{2} - \frac{\mathsf{s}}{4\sigma})\delta_\sigma & -2\sigma \le \mathsf{s} \le 2\sigma \\ \delta_\sigma & \mathsf{s} \ge 2\sigma \\ \delta_{-\sigma} & \mathsf{s} \le -2\sigma \end{cases}$$

The salient part of our construction of W_σ is that

$$\mathsf{W}_\sigma(\sigma) = \frac{1}{4}\delta_{-\sigma} + \frac{3}{4}\delta_\sigma, \; \mathsf{W}_\sigma(-\sigma) = \frac{1}{4}\delta_\sigma + \frac{3}{4}\delta_{-\sigma}.$$

Denote the marginals of $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star_{\mathrm{dec}}$ with $\mathsf{P}^\star_{\circlearrowleft,h}$ and $\mathsf{P}^\star_{\mathrm{dec},h}$. One can show via the lack of memory in the dynamics and the structure of $\pi^\star$ that

$$\mathsf{P}^\star_{\circlearrowleft,h+1} = \mathsf{W}^\star_{\circlearrowleft,h} \circ \mathsf{P}^\star_{\circlearrowleft,h}, \quad \mathsf{W}^\star_{\mathrm{dec},h+1} = \mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{P}^\star_{\mathrm{dec},h}, \tag{F.5}$$

By the replica property (Lemma E.3), $\mathsf{W}^\star_{\circlearrowleft,h} \circ \mathsf{P}^\star_h = \mathsf{P}^\star_h$ for all h. Thus, for all h, (F.4) and (F.5) imply

$$\mathsf{P}^\star_{\circlearrowleft,h} = \mathsf{P}^\star_h = \frac{1}{4}\delta_{-\sigma} + \frac{3}{4}\delta_{+\sigma}. \tag{F.6}$$

The following claim computes $\mathsf{P}^\star_{\mathrm{dec},h}$.

Claim F.3. *Consider any distribution of the form* $\mathsf{P} = (1-p)\delta_\sigma + p\delta_{-\sigma}$. *Then*

$$\mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{P} = (\frac{9}{10} - \frac{p}{5})\delta_\sigma + (\frac{1}{10} + \frac{p}{5})\delta_{-\sigma}.$$

Thus,

$$\mathsf{P}^\star_{\mathrm{dec},h+1}[-\sigma] = \frac{1}{10}\left(\sum_{i=0}^{h-1} 5^{-i}\right) + \frac{1}{4}5^{1-h}.$$

Before proving the claim, let us remark on its implications. As $h \to \infty$,

$$\mathsf{P}^\star_{\mathrm{dec},h}[-\sigma] \to \frac{1}{10}\left(\frac{1}{1-1/5}\right) = \frac{1}{10} \cdot \frac{5}{4} = \frac{1}{8}.$$

Thus,

$$\lim_{h\to\infty} \mathsf{P}^\star_{\mathrm{dec},h} = \frac{7}{8}\delta_\sigma + \frac{1}{8}\delta_{-\sigma},$$

achieving a different stationary distribution that $\mathsf{P}^\star_h = \mathsf{P}^\star_{\circlearrowleft,h}$. This shows that

$$\lim_{H\to\infty} \Gamma_{\mathrm{marg},\sigma}(\pi^\star_{\circlearrowleft\sigma}, \pi^\star_{\mathrm{dec}}) \ge \mathsf{TV}(\frac{7}{8}\delta_\sigma + \frac{1}{8}\delta_{-\sigma}, \frac{3}{4}\delta_\sigma + \frac{1}{4}\delta_{-\sigma}) = \frac{1}{8},$$

which implies that the deconvolution policy $\pi^\star_{\mathrm{dec}}$ does approximate $\pi^\star_{\circlearrowleft\sigma}$. From (F.6), it also follows that $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star$ have identical marginals, so

$$\lim_{H\to\infty} \Gamma_{\mathrm{marg},\sigma}(\pi^\star, \pi^\star_{\mathrm{dec}}) \ge \mathsf{TV}(\frac{7}{8}\delta_\sigma + \frac{1}{8}\delta_{-\sigma}, \frac{3}{4}\delta_\sigma + \frac{1}{4}\delta_{-\sigma}) = \frac{1}{8}$$

as well. In particular, if we take $\hat{\pi} = \pi^\star_{\mathrm{dec}}$ in Theorem 2, we see that there is no hope to for bounding $\Gamma_{\mathrm{marg},\varepsilon}(\pi^\star, \hat{\pi})$; we must bound $\Gamma_{\mathrm{marg},\varepsilon}(\pi^\star, \hat{\pi} \circ \mathsf{W}_\sigma)$ (again noting that if $\hat{\pi} = \pi^\star_{\mathrm{dec}}$, $\hat{\pi} \circ \mathsf{W}_\sigma = \pi^\star_{\circlearrowleft\sigma}$).

Proof of Claim F.3. We have that for $\mathsf{s}' \in \{-\sigma, \sigma\}$,

$$\mathsf{W}^\star_{\mathrm{dec},\mathsf{s}'|\mathsf{s}} = \frac{\mathsf{W}_\sigma(\mathsf{s}')[\mathsf{s}] \cdot \mathsf{P}^\star_h(\mathsf{s}')}{\mathsf{W}_\sigma(\mathsf{s}')[\mathsf{s}] \cdot \mathsf{P}^\star_h(\mathsf{s}') + \mathsf{W}_\sigma(-\mathsf{s}')[\mathsf{s}] \cdot \mathsf{P}^\star_h(-\mathsf{s}')}$$

With $\mathsf{s} = \mathsf{s}' = \sigma$, the above is

$$\mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}' = \sigma \mid \mathsf{s} = \sigma) = \frac{\frac{3}{4} \cdot \frac{3}{4}}{\frac{3}{4} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{4}} = \frac{9}{10}.$$

And

$$\mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}' = \sigma \mid \mathsf{s} = -\sigma) = \frac{\frac{1}{4} \cdot \frac{3}{4}}{\frac{1}{4} \cdot \frac{3}{4} + \frac{3}{4} \cdot \frac{1}{4}} = \frac{1}{2}.$$

Hence, for any $p \in [0, 1]$,

$$\begin{aligned}
\mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}' = \sigma \mid \mathsf{s} = -\sigma)((1-p)\delta_\sigma + p\delta_{-\sigma}) &= ((1-p)\frac{9}{10} + \frac{p}{2})\delta_\sigma + (1 - ((1-p)\frac{9}{10} + \frac{p}{2})))\delta_\sigma \\
&= (\frac{9}{10} - \frac{p}{5})\delta_\sigma + (\frac{1}{10} + \frac{p}{5})\delta_{-\sigma}.
\end{aligned}$$

Consequently, by (F.5), we can unfold a recursion to compute

$$\begin{aligned}
\mathsf{P}^\star_{\mathrm{dec},h+1}[-\sigma] &= \mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}' = \sigma \mid \mathsf{s} = -\sigma)\mathsf{P}^\star_{\mathrm{dec},h} \\
&= (\frac{1}{10} + \frac{\mathsf{P}^\star_{\mathrm{dec},h}[\sigma]}{5}) \\
&= \frac{1}{10}\sum_{i=0}^{h-1} 5^{-i} + \mathsf{P}^\star_{\mathrm{dec},1}[\sigma] \cdot 5^{1-h} \\
&= \frac{1}{10}\sum_{i=0}^{h-1} 5^{-i} + \mathsf{P}^\star_{1}[\sigma] \cdot 5^{1-h} \\
&= \frac{1}{10}\left(\sum_{i=0}^{h-1} 5^{-i}\right) + \frac{1}{4}5^{1-h}.
\end{aligned}$$

□

Part II

The Control Setting

G. Stability in the Control System

This section proves our various stability conditions. One wrinkle in the exposition is that we are able to derive far sharper perturbation guarantees than are needed in our analysis. However, as the guarantees are rather technically burdensome to derive, we endeavor to present the sharpest possible results so that we may save others from having to rederive these bounds in future applications.

Importantly, this section also contains the definition of the constants $c_1, \ldots, c_5 > 0$ present in Theorem 1, Proposition 4.1, and other main results (see Definition G.7).

The section is organized as follows:

- Appendix G.1 recalls various preliminaries.

- Appendix G.2 provides the definition of numerous problem-dependent constants, all of which are polynomial in $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$ and $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$ defined in Assumptions 3.1 and 3.2.
- Appendix G.3 gives IPS guarantees in terms of the constants in the previous section. Specifically, it provides Definition G.7, which instantiates the constants $c_1, \dots, c_5 > 0$ present in Theorem 1, Proposition 4.1, and other main results. We then state Corollary G.1, from which we derive Proposition 4.1 used in the body. This corollary is derived from a sharper guarantee, Proposition G.3 (whose improvements over the corollary are detailed in Remark G.2).
- The results in Appendix G.3 are derived from two building blocks in Appendix G.4: Lemma G.4 which bounds sensitive of regular trajectories to initial state, and Proposition G.5 which addresses perturbations of control inputs and gain.
- Proposition G.3 is derived from Proposition G.5 in Appendix G.5. Lemma G.4 and Proposition G.5 are proven in Appendix G.7 in Appendix G.7, respectively.
- Appendix G.8 explains how to implement a synthesis oracle which produces Jacobian Stabilizing primitives controllers from trajectories which satisfy a natural stabilizability condition.
- Finally, Appendix G.9 gives the solutions to various scalar recursions used in the proofs of Lemma G.4 and Proposition G.5.

G.1. Recalling preliminaries and assumptions.

Recall the following definitions.

- A length-K *control trajectory* is denoted $\boldsymbol{\rho} = (x_{1:K+1}, u_{1:K}) \in \mathscr{P}_K = (\mathbb{R}^{d_x})^{K+1} \times (\mathbb{R}^{d_u})^K$.
- Its *Jacobian linearizations* are denoted $\mathbf{A}_k(\boldsymbol{\rho}) := \frac{\partial}{\partial x} f_\eta(\mathbf{x}_k, \mathbf{u}_k)$ and $\mathbf{B}_k(\boldsymbol{\rho}) := \frac{\partial}{\partial u} f_\eta(\mathbf{x}_k, \mathbf{u}_k)$ for $k \in [K]$.
- Recalling our dynamics map $f(\cdot,\cdot)$, and step size $\eta > 0$, we say $\boldsymbol{\rho}$ is *feasible* if, for all $k \in [K]$,

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k), \quad \text{where } f(\mathbf{x}, \mathbf{u}) = \mathbf{x} + \eta f_\eta(\mathbf{x}, \mathbf{u}).$$

We regular the definition of regular trajectories from Section 3.

Definition G.1. A control path $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K})$ is $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular if for all $k \in [K]$ and all $(\mathbf{x}'_k, \mathbf{u}'_k) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_u}$ such that $\|\mathbf{x}'_k - \mathbf{x}_k\| \vee \|\mathbf{u}_k - \mathbf{u}'_k\| \le R_{\mathrm{dyn}}$,[5]

$$\|\nabla f_\eta(\mathbf{x}'_k, \mathbf{u}'_k)\|_{\mathrm{op}} \le L_{\mathrm{dyn}}, \quad \|\nabla^2 f_\eta(\mathbf{x}'_k, \mathbf{u}'_k)\|_{\mathrm{op}} \le M_{\mathrm{dyn}}.$$

We also recall the definitions around Jacobian stabilization. We start with a definition of Jacobian stabilization for feedback gains, from which we then recover the definition of Jacobian stabilization for primitive controllers given in the body.

Definition G.2. Consider $R_{\mathrm{stab}}, L_{\mathrm{stab}}, B_{\mathrm{stab}} \ge 1$. Consider sequence of gains $\mathbf{K}_{1:K} \in (\mathbb{R}^{d_u \times d_u})^K$ and trajectory $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K}) \in \mathscr{P}_K$. We say that $(\boldsymbol{\rho}, \mathbf{K}_{1:K})$-is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian Stable if $\max_k \|\mathbf{K}_k\|_{\mathrm{op}} \le B_{\mathrm{stab}}$, and if the closed-loop transition operators defined by

$$\boldsymbol{\Phi}_{\mathrm{cl},k,j} := (\mathbf{I} + \eta \mathbf{A}_{\mathrm{cl},k-1}) \cdot (\mathbf{I} + \eta \mathbf{A}_{\mathrm{cl},k-2}) \cdot (\dots) \cdot (\mathbf{I} + \eta \mathbf{A}_{\mathrm{cl},j})$$

with $\mathbf{A}_{\mathrm{cl},k} = \mathbf{A}_k(\boldsymbol{\rho}) + \mathbf{B}_{k-1}(\boldsymbol{\rho})\mathbf{K}_{k-1}$ satisfies the following inequality

$$\|\boldsymbol{\Phi}_{\mathrm{cl},k,j}\|_{\mathrm{op}} \le B_{\mathrm{stab}}(1 - \frac{\eta}{L_{\mathrm{stab}}})^{k-j}.$$

The definition of Jacobian stability of primitive controllers in Section 3 may be recovered as follows.

Definition G.3. Consider $R_{\mathrm{stab}}, L_{\mathrm{stab}}, B_{\mathrm{stab}} \ge 1$. Consider a sequence of primitive controllers $\kappa_{1:K} \in \mathcal{K}^K$, each expressed as $\kappa_k(\mathbf{x}) = \bar{\mathbf{u}}_k = \bar{\mathbf{K}}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k)$ and $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K}) \in \mathscr{P}_K$. We say $(\boldsymbol{\rho}, \kappa_{1:K})$ is Jacobian Stable if $\kappa_{1:K}$ is consistent with $\boldsymbol{\rho}$, and if $(\boldsymbol{\rho}, \bar{\mathbf{K}}_{1:K})$ is $R_{\mathrm{stab}}, L_{\mathrm{stab}}, B_{\mathrm{stab}} > 0$-Jacobian stable.

Note that in Jacobian stability (both with primitive controllers and with gain-matrices), we take all parameters to be no less than one.

[5] Here, $\|\nabla^2 f_\eta(\mathbf{x}'_t, \mathbf{u}'_t)\|_{\mathrm{op}}$ denotes the operator-norm of a three-tensor.

G.1.1. Properties satisfied by $\pi^\star$

Finally, we show that actions produced by $\pi^\star$ in our control instantiation of the composite MDP satisy the assumptions in Assumptions 3.1 and 3.2.

Lemma G.1. *Suppose Assumptions 3.1 and 3.2 hold. Let $\pi^\star = (\pi^\star_h)_{1\le h\le H}$ denote the policy constructed as a regular conditional probability from the conditionals of $\mathcal{D}_{\mathrm{exp}}$. Furthermore, let $\mathcal{D}_{\mathrm{exp},\boldsymbol{\rho}_{\mathrm{m},h}}$ denote the distribution over $\boldsymbol{\rho}_{\mathrm{m},h}$ corresponding to $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$. Then, with probability one over $\boldsymbol{\rho}_{\mathrm{m},h} \sim \mathcal{D}_{\mathrm{exp},\boldsymbol{\rho}_{\mathrm{m},h}}$ and $\mathsf{a}_h \sim \pi^\star_h(\boldsymbol{\rho}_{\mathrm{m},h})$, expressed as $\boldsymbol{\rho}_{\mathrm{m},h} = (\mathbf{x}_{t_h:t_h-\tau_{\mathrm{m}}+1}, \mathbf{u}_{t_h-1:t_h-\tau_{\mathrm{m}}+1})$, and $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$. Consider the unique feasible trajectory for which*

$$\boldsymbol{\rho}_{\mathrm{c},h+1} = (\mathbf{x}'_{t_h:t_{h+1}}, \mathbf{u}'_{t_h:t_{h+1}-1}), \quad \mathbf{x}'_{t_h} = \mathbf{x}_{t_h}, \quad \mathbf{u}_t = \kappa_t(\mathbf{x}_t), \quad t_h \le t < t_{h+1}.$$

Then,

- *$\boldsymbol{\rho}'_{\mathrm{c},h+1}$ is $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular*
- *$(\boldsymbol{\rho}'_{\mathrm{c},h+1}, \kappa_{t_h:t_{h+1}-1})$ is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian stable.*

Proof. Since $\pi^\star$ in Definition I.3 is constructed as the regular conditional probabiliy of $\mathsf{a}_h \mid \boldsymbol{\rho}_{\mathrm{m},h}$ under $\mathcal{D}_{\mathrm{exp}}$, $(\mathsf{a}_h, \boldsymbol{\rho}_{\mathrm{m},h})$ is the above lemma have the same joint distribution as under $\mathcal{D}_{\mathrm{exp}}$. Thus, the lemma follows from the assumptions Assumptions 3.1 and 3.2 placed on $\mathcal{D}_{\mathrm{exp}}$. □

The following is a direct consequence of the above lemma.

Lemma G.2. *Consider the instantiation of the composite MDP for the control setting as in Section 4.1 and in Appendix I, with $\pi^\star$ as in Definition I.3, and $\phi_{\mathcal{Z}}$ as in Definition E.1. Suppose that $\mathsf{W}_1, \dots, \mathsf{W}_h : \mathcal{S} \to \Delta(\mathcal{S})$ satisfy*[6]

$$\phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathcal{Z}} \circ \mathsf{P}^\star_h, \tag{G.1}$$

Consider a sequence of actions $\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H}$ generated via

$$\mathsf{a}_h \sim \pi^\star_h(\tilde{\mathsf{s}}_h), \quad \tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h), \quad \mathsf{s}_{h+1} = F_h(\mathsf{s}_h, \mathsf{a}_h), \quad \mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}.$$

Let $\tilde{\mathsf{s}}_h = (\tilde{\mathbf{x}}_{t_{h-1}:t_h}, \tilde{\mathbf{u}}_{t_{h-1}:t_h-1})$ and $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$. Then, with probability one, for each h, the unique feasible trajectory for which

$$\boldsymbol{\rho}_{\mathrm{c},h+1} = (\mathbf{x}'_{t_h:t_{h+1}}, \mathbf{u}'_{t_h:t_{h+1}-1}), \quad \mathbf{x}'_{t_h} = \mathbf{x}_{t_h}, \quad \mathbf{u}_t = \kappa_t(\mathbf{x}_t), \quad t_h \le t < t_{h+1}.$$

satisfies

- *$\mathbf{x}'_{t_h} = \tilde{\mathbf{x}}_{t_h}$, and $\boldsymbol{\rho}'_{\mathrm{c},h+1}$ is feasible and $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular*
- *$(\boldsymbol{\rho}'_{\mathrm{c},h+1}, \kappa_{t_h:t_{h+1}-1})$ is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian Stable.*

Remark G.1 (On the absolute continuity constraint in (G.1))**.** Recall that $\phi_{\mathcal{Z}}$ as defined in Definition I.1 simply extracts the memory chunk $\boldsymbol{\rho}_{\mathrm{m},h}$ from the trajectory chunk $\boldsymbol{\rho}_{\mathrm{c},h}$. The condition $\mathrm{supp}(\phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s})) \subset \mathrm{supp}(\phi_{\mathcal{Z}} \circ \mathsf{P}^\star_h)$ just means that the distribution of the memory chunk-components from $\mathsf{W}_h(\mathsf{s})$is absolutely continuous with respect to the memory-chunks $\boldsymbol{\rho}_{\mathrm{m},h}$ under $\mathcal{D}_{\mathrm{exp}}$.

G.1.2. Norm notation.

Lastly, given our parameter $\eta > 0$, we define two types of norms. First, for sequences of vectors $\mathbf{z}_{1:K} \in (\mathbb{R}^d)^K$ and matrices $(\mathbf{X}_{1:K}) \in \mathbb{R}^{d_1 \times d_2})^K$, define

$$\|\mathbf{z}_{1:K}\|_{\ell_2} = \left(\eta \sum_{k=1}^K \|\mathbf{z}_k\|^2\right), \quad \|\mathbf{X}_{1:K}\|_{\ell_2,\mathrm{op}} = \left(\eta \sum_{k=1}^K \|\mathbf{X}_k\|^2\right),$$

where again the standard $\|\cdot\|$ notation denotes Euclidean norm for vectors and operator norm for matrices. We also define

$$\|\mathbf{z}_{1:K}\|_{\max,2} = \max_{1\le k\le K} \|\mathbf{z}_k\|, \quad \|\mathbf{X}_{1:K}\|_{\max,\mathrm{op}} = \max_{1\le k\le K} \|\mathbf{X}_k\|.$$

[6] Recall the absolute-continuity comparator $\ll$ defined in Definition C.4.

G.2. Composite Problem Constants

We begin by writing down numerous problem constants, all of which are polynomial in the quantities $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$ and $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$. First, we define the ***stability exponent***,

$$\beta_{\mathrm{stab}} := (1 - \frac{\eta}{L_{\mathrm{stab}}}) \in (0,1).$$

Definition G.4. Given the regularity parameters $R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}}$, stability parameters $R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}}$, and the step size $\eta > 0$, we define the "little-c" constants

$$c_{\mathbf{u}} = 12 B_{\mathrm{stab}} \sqrt{L_{\mathrm{stab}}} L_{\mathrm{dyn}}, \quad c_{\mathbf{K}} = 2 B_{\mathrm{stab}} + 12 B_{\mathrm{stab}} L_{\mathrm{stab}}^{1/2} L_{\mathrm{dyn}}, \quad c_{\mathbf{\Delta}} = 6 B_{\mathrm{stab}}$$

as well as "big-C" constants

$$\begin{aligned}
C_{\mathbf{u}} &:= \min\left\{ \frac{\sqrt{L_{\mathrm{stab}}} L_{\mathrm{dyn}}}{M_{\mathrm{dyn}}}, \frac{1}{256 B_{\mathrm{stab}}^2 M_{\mathrm{dyn}} L_{\mathrm{dyn}} L_{\mathrm{stab}}^{3/2}} \right\} \\
C_{\mathbf{\Delta}} &:= \frac{1}{4 \cdot 324 B_{\mathrm{stab}}^2 M_{\mathrm{dyn}} L_{\mathrm{stab}}} \\
C_{\hat{\mathbf{x}}} &:= \min\left\{ \frac{R_{\mathrm{dyn}}}{2 R_{\mathrm{stab}} B_{\mathrm{stab}}}, \frac{1}{16 L_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathrm{stab}}^2 B_{\mathrm{stab}}^2} \right\} \\
C_{\mathbf{K}} &:= \min\left\{ \frac{1}{24 \sqrt{L_{\mathrm{stab}}} B_{\mathrm{stab}} L_{\mathrm{dyn}}}, \frac{C_{\hat{\mathbf{x}}}^{-1} L_{\mathrm{dyn}}}{8 \cdot 324 B_{\mathrm{stab}}^2 M_{\mathrm{dyn}} L_{\mathrm{stab}}^{3/2}} \right\} \\
C_{\mathbf{K},\hat{\mathbf{x}}} &:= \frac{L_{\mathrm{dyn}}}{8 \cdot 324 B_{\mathrm{stab}}^2 M_{\mathrm{dyn}} L_{\mathrm{stab}}^{3/2}}.
\end{aligned}$$

The "little-c" constants enter directly into our error bounds, where as the "big-C" constants function as constraints on errors, above which we lose guarantees. We define some additional "big-C" constants which take in a radius argument R_0.

Definition G.5 (Final Stability Constants)**.** In term of the constants in Definition G.5, we define the following final stability constants, as functions of a parameter R_0:

$$\begin{aligned}
C_{\mathrm{stab},1}(R_0) &:= \min\left\{ C_{\mathbf{u}}, \frac{C_{\mathbf{\Delta}}}{4 c_{\mathbf{u}}}, \frac{R_{\mathrm{dyn}}}{R_0} \cdot \frac{1}{48 c_{\mathbf{u}} c_{\mathbf{\Delta}}} \right\} \\
C_{\mathrm{stab},2}(R_0) &:= \min\left\{ C_{\mathbf{K}}, \frac{\beta_{\mathrm{stab}}^{-\tau_{\mathrm{c}}/3} C_{\mathbf{\Delta}}}{4 c_{\mathbf{u}}}, \frac{R_{\mathrm{dyn}}}{R_0} \cdot \frac{\beta_{\mathrm{stab}}^{-\tau_{\mathrm{c}}/3}}{48 c_{\mathbf{K}} c_{\mathbf{\Delta}}} \right\} \\
C_{\mathrm{stab},3}(R_0) &:= \frac{R_{\mathrm{dyn}}}{12 R_0 c_{\mathbf{u}} \sqrt{L_{\mathrm{stab}}} + 3} \\
C_{\mathrm{stab},4}(R_0) &:= \min\left\{ C_{\hat{\mathbf{x}}}, \frac{C_{\mathbf{K},\hat{\mathbf{x}}}}{C_{\mathbf{K}}}, \frac{R_{\mathrm{dyn}}}{R_0} \cdot \frac{1}{12 c_{\mathbf{K}}} \right\}
\end{aligned}$$

G.3. IPS Guarantees & Proof of Proposition 4.1

Here we provide our main stability guarantees for the learned policy $\pi^\star$ under Assumptions 3.1 and 3.2, from which we derive Proposition 4.1. This section adopts the notation from Section 4.1.

We begin by introducing a functional form for our distances.

Definition G.6 (Distances)**.** Let τ_{c} be given, and let $0 \le \tau \le \tau_{\mathrm{c}}$. For $h > 1$ and chunk-states $\mathsf{s}_h = (\mathbf{x}_{t_{h-1}:t_h}, \mathbf{u}_{t_{h-1}:t_h-1}) \in \mathscr{P}_{\tau_{\mathrm{f}}}$ and $\mathsf{s}'_h = (\mathbf{x}'_{t_{h-1}:t_h}, \mathbf{u}'_{t_{h-1}:t_h-1})$, define

$$\begin{aligned}
\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_h, \mathsf{s}'_h) &:= \max_{t \in [t_h - \tau : t_h]} \|\mathbf{x}_t - \mathbf{x}'_t\| \\
\mathsf{d}_{\mathcal{S},\mathbf{u},\tau}(\mathsf{s}_h, \mathsf{s}'_h) &:= \max_{t \in [t_h - \tau : t_h - 1]} \|\mathbf{u}_t - \mathbf{u}'_t\| \\
\mathsf{d}_{\mathcal{S},\tau}(\mathsf{s}_h, \mathsf{s}'_h) &:= \max\left\{ \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_h, \mathsf{s}'_h), \mathsf{d}_{\mathcal{S},\mathbf{u},\tau}(\mathsf{s}_h, \mathsf{s}'_h) \right\},
\end{aligned}$$

For $h = 1$, $\mathsf{s}_1 = \bar{\mathbf{x}}_1$ and $\mathsf{s}'_1 = \bar{\mathbf{x}}'_1$, we define $\mathsf{d}_{\mathcal{S}}(\mathsf{s}_1, \mathsf{s}'_1) = \|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}'_1\|$. Note therefore that

$$\mathsf{d}_{\mathcal{S},\tau_{\mathrm{c}}}(\cdot,\cdot) = \mathsf{d}_{\mathcal{S}}(\cdot,\cdot), \quad \mathsf{d}_{\mathcal{S},\tau_{\mathrm{m}}-1}(\cdot,\cdot) = \mathsf{d}_{\text{TVC}}(\cdot,\cdot), \quad \mathsf{d}_{\mathcal{S},0}(\cdot,\cdot) = \mathsf{d}_{\text{IPS}}(\cdot,\cdot)$$

Next, we introduce five problem-dependent constants $c_1, \dots, c_5$, all of which are stated in terms of the constants in Appendix G.2; one can readily check that these are all polynomial in the constants $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$ and $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$ in Assumptions 3.1 and 3.2.

Definition G.7 (IPS Constants)**.** In terms of constants in Appendix G.2, we define the IPS constants as follows:

$$\begin{aligned} c_1 &:= (6 \max\{R_{\mathrm{stab}}(1 + 2c_{\mathbf{u}}\sqrt{L_{\mathrm{stab}}}), B_{\mathrm{stab}} + \sqrt{L_{\mathrm{stab}}}c_{\mathbf{K}}\}). \\ c_2 &:= \min\left\{\frac{C_{\mathrm{stab},1}(2R_{\mathrm{stab}})}{4R_{\mathrm{stab}}\sqrt{L_{\mathrm{stab}}}}, \frac{C_{\mathrm{stab},2}(2R_{\mathrm{stab}})}{\sqrt{L_{\mathrm{stab}}}}, \frac{C_{\mathrm{stab},3}(2R_{\mathrm{stab}})}{4R_{\mathrm{stab}}}, \frac{1}{2R_{\mathrm{stab}}}\right\}. \end{aligned} \tag{G.2}$$

We further define

$$c_3 = 3L_{\mathrm{stab}} \log(2c_{\boldsymbol{\Delta}}), \quad c_4 = \min\{1, C_{\mathrm{stab},4}(2R_{\mathrm{stab}})\}, \quad c_5 = 2(1 + R_{\mathrm{stab}})B_{\mathrm{stab}}.$$

In terms of the constants $c_1, c_2 > 0$ above, we introduce a family of distance-like functions on $\bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}, \mathsf{a}' \mid r) : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_{\geq 0} \cup \{\infty\}$, defined as follows.

Definition G.8. Consider $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{x}}_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{K}}_{1:\tau_{\mathrm{c}}})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{x}}'_{1:\tau_{\mathrm{c}}}, \bar{\mathbf{K}}'_{1:\tau_{\mathrm{c}}})$.

$$\begin{aligned} \bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}, \mathsf{a}' \mid r) := c_1 \max_{1 \leq k \leq \tau_{\mathrm{c}}} \left(\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + re^{-\frac{\eta(\tau_{\mathrm{c}}-\tau)}{3L_{\mathrm{stab}}}} \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\| \right) \\ + \mathbf{I}_{0,\infty}\left\{ \max_{1 \leq k \leq \tau_{\mathrm{c}}} \left(\max\left\{\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\|, \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\|, \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|\right\}\right) \leq c_2 \right\}, \end{aligned}$$

where $\mathbf{I}_{0,\infty}(\mathcal{E})$ is 0 if clase $\mathcal{E}$ is true and ∞ otherwise.

In words, $\bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}, \mathsf{a}' \mid r)$ measures the maximal differences between $\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k$, $\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k$, and $\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k$, subject to a constraint that each of these quantities is within some bound c_2. One the latter threshold is met, the dependence on $\|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|$ is scaled down by r, and is also exponentially small in $\tau_{\mathrm{c}} - \tau$; this latter bit is not necessary for our results, but illustrates an interesting feature of our stability guarantees: *they are far less sensitive to errors in $\bar{\mathbf{K}}$ than to errors in $\bar{\mathbf{u}}$.*

In terms of $\bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}, \mathsf{a}' \mid r)$ defined above, we can now ensure the following stability guarantee.

Corollary G.1. *Suppose that $\tau_{\mathrm{c}} \geq c_3/\eta$ and $r \leq c_4$, and consider any sequence of kernels $\{\mathsf{W}_h\}_{h=1}^h$, where $\mathsf{W}_h : \mathcal{S} \to \Delta(\mathcal{S})$, and*[7]

$$\max_{h,\mathsf{s} \in \mathcal{S}} \mathbb{P}_{\tilde{\mathsf{s}} \sim \mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\text{IPS}}(\tilde{\mathsf{s}}, \mathsf{s}) \leq r] = 1, \phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathcal{Z}} \circ \mathsf{P}^{\star}_h,$$

for $\phi_{\mathcal{Z}}$ is from the direct decomposition instantiated in Definition I.1, and where $\mathsf{P}^{\star}_h$ denotes the law of $\boldsymbol{\rho}_{\mathrm{c},h}$ under $\mathcal{D}_{\exp}$ as in Definition I.3.

Define a process $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h)$, $\mathsf{a}_h \sim \pi^{\star}_h(\tilde{\mathsf{s}}_h)$, and $\mathsf{s}_{h+1} := F_h(\mathsf{s}_h, \mathsf{a}_h)$. Then, almost surely, the following hold for all $0 \leq \tau \leq \tau_{\mathrm{c}}$:

- *For each $1 \leq h \leq H$, $\mathsf{d}_{\mathcal{S},\tau}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_h) \leq c_5 re^{-\frac{\eta(\tau_{\mathrm{c}}-\tau)}{L_{\mathrm{stab}}}}$.*
- *For any sequence $(\mathsf{a}'_{1:H})$, the dynamics $\mathsf{s}'_1 = \mathsf{s}_1$, $\mathsf{s}_{h+1} = F_h(\mathsf{s}'_h, \mathsf{a}'_h)$ satisfy*

$$\max_{1 \leq h \leq H+1} \mathsf{d}_{\mathcal{S},\tau}(\mathsf{s}_h, \mathsf{s}'_h) \leq \max_{1 \leq h \leq H} \bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}_h, \mathsf{a}'_h \mid r).$$

Corollary G.1 is derived in Appendix G.3.2 from an even more granular result stated just below. Before continuing, we explain how Proposition 4.1 follows.

Proof of Proposition 4.1. This follows directly from the above corollary notice that c_4 is define to be at most 1, so we always invoke the corollary with $r \leq 1$, and thus $\bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}_h, \mathsf{a}'_h \mid r) \leq \bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}_h, \mathsf{a}'_h \mid 1) \leq \mathsf{d}_{\mathcal{A}}$. We remark that the guarantee only applies to kernels for which □

[7] See Remark G.1 above for iterpretation of this condition below.

G.3.1. A MORE GRANULAR STABILITY STATEMENT

Here, we state an even more granular stability guarantee. The notation is rather onerous, but captures another nice feature of our bound: that our stability depends not on the maximal errors over $\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k$, $\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k$, $\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k$, but rather on ℓ_2-errors. Again, not necessary for our guarantees, but it speaks to the sharpness of our perturbation bounds. See Remark G.2 at the end of the section for more discussion.

Definition G.9 (Action Differences (inputs and gains))**.** Consider $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_c}, \bar{\mathbf{x}}_{1:\tau_c}, \bar{\mathbf{K}}_{1:\tau_c})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_c}, \bar{\mathbf{x}}'_{1:\tau_c}, \bar{\mathbf{K}}'_{1:\tau_c})$. We define

$$
\begin{aligned}
\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a},\mathsf{a}') &= \max_{1\le k\le\tau_c}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\bar{\mathbf{u}}_j - \bar{\mathbf{u}}'_j\|^2\right)^{1/2}\\
\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a},\mathsf{a}') &= \max_{1\le k\le\tau_c}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\bar{\mathbf{x}}_j - \bar{\mathbf{x}}'_j\|^2\right)^{1/2}\\
\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a},\mathsf{a}') &= \max_{1\le k\le\tau_c}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\bar{\mathbf{K}}_j - \bar{\mathbf{K}}'_j\|\right)^{1/2}.
\end{aligned}
$$

We further define

$$
\begin{aligned}
\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a},\mathsf{a}') &:= \max_{1\le k\le\tau_c}\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| = \|\bar{\mathbf{u}}_{1:\tau_c} - \bar{\mathbf{u}}'_{1:\tau_c}\|_{\max,2}\\
\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a},\mathsf{a}') &:= \max_{1\le k\le\tau_c}\|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| = \|\bar{\mathbf{x}}_{1:\tau_c} - \bar{\mathbf{x}}'_{1:\tau_c}\|_{\max,2}\\
\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a},\mathsf{a}') &:= \max_{1\le k\le\tau_c}\|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\| = \|\bar{\mathbf{K}}_{1:\tau_c} - \bar{\mathbf{K}}'_{1:\tau_c}\|_{\max,\mathrm{op}}.
\end{aligned}
$$

and

$$
\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) := \max_{1\le k\le\tau_c}\|\bar{\mathbf{K}}_k\| = \|\bar{\mathbf{K}}_{1:\tau_c}\|_{\max,\mathrm{op}}.
$$

We note further that as $\beta_{\mathrm{stab}} \in (0,1)$ and $\eta\sum_{i\ge0}\beta_{\mathrm{stab}}^i = L_{\mathrm{stab}}$, we have

$$
\begin{aligned}
\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a},\mathsf{a}') &\le \|\bar{\mathbf{u}}_{1:\tau_c} - \bar{\mathbf{u}}'_{1:\tau_c}\|_{\ell_2} \wedge \sqrt{L_{\mathrm{stab}}}\|\bar{\mathbf{u}}_{1:\tau_c} - \bar{\mathbf{u}}'_{1:\tau_c}\|_{\max,2}\\
&\le \|\bar{\mathbf{u}}_{1:\tau_c} - \bar{\mathbf{u}}'_{1:\tau_c}\|_{\ell_2} \wedge \sqrt{L_{\mathrm{stab}}}\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a},\mathsf{a}')
\end{aligned}
\tag{G.3}
$$

and analogously for $\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a},\mathsf{a}')$ and $\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a},\mathsf{a}')$.

Next, recall the constants $\{C_{\mathrm{stab},i}(R_0)\}_{i=1}^4$ in Definition G.5, and $c_{\mathbf{u}}, c_{\mathbf{K}}$ in Definition G.4, all of which are polynomial in relevant problem parameters $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$, $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$, and argument R_0. We now define a very general distance-like function between actions.

Definition G.10 (Action Divergences)**.** Define, for $R_0 \ge 1$, the following

$$
\mathsf{d}_{\mathcal{A},R_0,\tau}(\mathsf{a}_h,\mathsf{a}'_h \mid r) := 2((1+R_0)\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h,\mathsf{a}'_h \mid r) + \mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{u}}(\mathsf{a}_h,\mathsf{a}'_h \mid r)),
$$

where

$$
\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{u}}(\mathsf{a}_h,\mathsf{a}'_h \mid r) := \mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h,\mathsf{a}'_h) + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h,\mathsf{a}'_h) + 2rB_{\mathrm{stab}}\beta_{\mathrm{stab}}^{\tau_c-\tau}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h,\mathsf{a}'_h),
$$

and where

$$
\begin{aligned}
\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a},\mathsf{a}' \mid r) = {}& 2c_{\mathbf{u}}(\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a},\mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a},\mathsf{a}')) + 2c_{\mathbf{K}}r(\beta_{\mathrm{stab}})^{\frac{\tau_c-\tau}{3}}\cdot\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a},\mathsf{a}')\\
&+ \mathbf{I}_{0,\infty}\{\bigcap_{i=1}^{3}\mathcal{E}_{\mathrm{close},R_0,i}\} + \mathbf{I}_{0,\infty}\{\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \vee \mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le R_0\},
\end{aligned}
$$

where $\mathbf{I}_{0,\infty}\{\mathcal{E}\}$ denotes 0 if clause $\mathcal{E}$ is true, and ∞ otherwise, and where we define the clauses

$$\begin{aligned}
\mathcal{E}_{\mathrm{close},R_0,1}(\mathsf{a},\mathsf{a}') &= \{(\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a},\mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a},\mathsf{a}') \le C_{\mathrm{stab},1}(R_0)\} \\
\mathcal{E}_{\mathrm{close},R_0,2}(\mathsf{a},\mathsf{a}') &= \{\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a},\mathsf{a}') \le C_{\mathrm{stab},2}(R_0)\} \\
\mathcal{E}_{\mathrm{close},R_0,3}(\mathsf{a},\mathsf{a}') &= \{\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a},\mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a},\mathsf{a}') \le C_{\mathrm{stab},3}(R_0)\}.
\end{aligned}$$

Again, we see that asside from the $\mathbf{I}_{0,\infty}\{\cdot\}$ terms, our distances $\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a},\mathsf{a}' \mid r)$ depends only on ℓ_2-guarantees.

We may now state our most general stability guarantee.

Proposition G.3 (Main Stability Guarantees)**.** *Suppose that*

$$\tau_{\mathrm{c}} \ge 3L_{\mathrm{stab}}\log(2c_{\boldsymbol{\Delta}})/\eta.$$

In addition, fix an $R_0 > 0$, and $r_{\max}$ such that $r_{\max} \le C_{\mathrm{stab},4}(R_0)$. Consider any sequence of kernels $\{\mathsf{W}_h\}_{h=1}^{h}$, where $\mathsf{W}_h : \mathcal{S} \to \Delta(\mathcal{S})$ and[8]

$$\max_{h,\mathsf{s}\in\mathcal{S}} \mathbb{P}_{\tilde{\mathsf{s}}\sim\mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\tilde{\mathsf{s}},\mathsf{s}) \le r] = 1, \quad \phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathcal{Z}} \circ \mathsf{P}_h^\star, \tag{G.4}$$

and define a process $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h), \mathsf{a}_h \sim \pi_h^\star(\tilde{\mathsf{s}}_h)$, and $\mathsf{s}_{h+1} := F_h(\mathsf{s}_h,\mathsf{a}_h)$. Then, almost surely, the following hold for all $0 \le \tau \le \tau_{\mathrm{c}}$:

- *For each $1 \le h \le H$, $\mathsf{d}_{\mathcal{S},\tau}(F_h(\tilde{\mathsf{s}}_h,\mathsf{a}_h),\mathsf{s}_h) \le 2(1+R_{\mathrm{stab}})B_{\mathrm{stab}}r\beta_{\mathrm{stab}}^{(\tau_{\mathrm{c}}-\tau)}$.*
- *For any sequence $(\mathsf{a}'_{1:H})$, the dynamics $\mathsf{s}'_1 = \mathsf{s}_1$, $\mathsf{s}_{h+1} = F_h(\mathsf{s}'_h,\mathsf{a}'_h)$ satisfy*

  $$\max_{1\le h\le H+1} \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_h,\mathsf{s}'_h) \le \max_{1\le h\le H} \mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h,\mathsf{a}'_h \mid r).$$

  and

  $$\max_{1\le h\le H+1} \mathsf{d}_{\mathcal{S},\tau}(\mathsf{s}_h,\mathsf{s}'_h) \le \max_{1\le h\le H} \mathsf{d}_{\mathcal{A},R_0,\tau}(\mathsf{a}_h,\mathsf{a}'_h \mid r).$$

The above proposition is proven in Appendix G.5, wher it is derived from two key guarantees given in Appendix G.4 below.

Remark G.2 (Remark on the Scaling)**.** We now justify the extreme granularity of the above result. We demonstrate that our guarantees satisfy the following favorable properties:

- As in Corollary G.1, the dependence of $\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k$ in the non $\mathbf{I}_{0,\infty}\{\cdot\}$ scales down with r and with $r \cdot \beta_{\mathrm{stab}}^{(\tau_{\mathrm{c}}-\tau)/3)}$, so that errors in $\bar{\mathbf{K}}_k$ become less relevant as $\tau \to \tau_{\mathrm{c}}$ and as $r \to 0$.
- If we restrict our attention only to errors in states, captured by $\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}$, the non-$\mathbf{I}_{0,\infty}\{\cdot\}$ terms depend only on ℓ_2-errors rather than maximal ∞-norm ones.
- In the special case where $R_{\mathrm{dyn}} = \infty$, i.e., the regularity properties in Assumption 3.2 hold globally, then all terms $C_{\mathrm{stab},i}(R_0)$ defined in Definition G.5 no longer need depend on R_0, as the terms in which R_0 appears have an $R_{\mathrm{dyn}} = \infty$ in the numerator, and each $C_{\mathrm{stab},i}(R_0)$ serves as an upper bound on a certain quantity of interest. Hence, we can drop the dependence on R_0 in all of these terms. $C_{\mathrm{stab},i}(R_0)$ (
- In particular, the term $C_{\mathrm{stab},3}(R_0)$ if equal to ∞ when $R_{\mathrm{dyn}} = \infty$. Thus, for $R_{\mathrm{dyn}} = \infty$, we can drop the indictor of $\mathcal{E}_{\mathrm{close},R_0,3}(\mathsf{a},\mathsf{a}') := \{\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a},\mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a},\mathsf{a}') \le C_{\mathrm{stab},3}(R_0)\}.$, and hence each $\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}$ has depends only on ℓ_2-type errors.

The proof of Proposition G.3 is given in Appendix G.5, derived from the results in the subsection directly below. Before we do this, we first demonstrate how Corollary G.1 follows from Proposition G.3.

[8] Again, we refer to Remark G.1 for explanation of the second condition in the display (G.4)

G.3.2. Deriving Corollary G.1 from Proposition G.3

The proof is mostly notational bookkeeping.

By assumption $\phi_{\mathcal{Z}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathcal{Z}} \circ \mathsf{P}_h^\star$ and Lemma G.2, and the R_{stab}-term in $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian stability, the action a_h with $\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \le R_{\mathrm{stab}}$. Further, notice that the parameter β_{stab} used throughout this section can be bounded by at most

$$\beta_{\mathrm{stab}} := (1 - \frac{\eta}{L_{\mathrm{stab}}}) \le \exp(-\eta/L_{\mathrm{stab}}).$$

Hence, Corollary G.1 from Proposition G.3 as soon as we show that

$$\forall \mathsf{a} \ \text{ s.t. } \mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \le R_{\mathrm{stab}}, \quad \mathsf{d}_{\mathcal{A},R_0,\tau}(\mathsf{a}_h, \mathsf{a}'_h \mid r) \le \bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a}, \mathsf{a}' \mid r).$$

Consider the action divergences in Definition G.10. Take $R_0 = 2R_{\mathrm{stab}}$, where $R_{\mathrm{stab}} \ge 1$ by assumption. and upper bound $\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\cdot,\cdot) \le \sqrt{L_{\mathrm{stab}}}\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\cdot)$ (as in (G.3)), and similarly for $\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\cdot,\cdot)$ and $\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\cdot,\cdot)$. .Then,

$$\begin{aligned}
\mathsf{d}_{\mathcal{A},R_0,\tau}(\mathsf{a}_h, \mathsf{a}'_h \mid r) &:= 2((1+R_0)\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h, \mathsf{a}'_h \mid r) + \mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{u}}(\mathsf{a}_h, \mathsf{a}'_h \mid r)) \\
&= \mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + R_0 \mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + 2rB_{\mathrm{stab}}\beta_{\mathrm{stab}}^{\tau_c - \tau}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h, \mathsf{a}'_h) \\
&\quad + 2c_{\mathbf{u}}(\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a}, \mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a}, \mathsf{a}')) + 2c_{\mathbf{K}}r(\beta_{\mathrm{stab}})^{\frac{\tau_c - \tau}{3}} \cdot \mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a}, \mathsf{a}') \\
&\quad\quad + \mathbf{I}_{0,\infty}\{\bigcap_{i=1}^{3} \mathcal{E}_{\mathrm{close},R_0,i}\} + \mathbf{I}_{0,\infty}\{\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \vee \mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le R_0\} \\
&\le 2R_{\mathrm{stab}}(1 + 2c_{\mathbf{u}}\sqrt{L_{\mathrm{stab}}})(\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + \mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h, \mathsf{a}'_h) \\
&\quad + (2B_{\mathrm{stab}} + 2\sqrt{L_{\mathrm{stab}}}c_{\mathbf{K}})r\exp^{-\frac{\eta(\tau_c - \tau)}{3L_{\mathrm{stab}}}}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h, \mathsf{a}'_h) \\
&\quad\quad + \mathbf{I}_{0,\infty}\{\bigcap_{i=1}^{3} \mathcal{E}_{\mathrm{close},R_0,i}\} + \mathbf{I}_{0,\infty}\{\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \vee \mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le 2R_{\mathrm{stab}}\} \\
&\le \frac{c_1}{3}(\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + \mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + r\exp^{-\frac{\eta(\tau_c - \tau)}{3L_{\mathrm{stab}}}}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h, \mathsf{a}'_h)) \\
&\quad\quad + \mathbf{I}_{0,\infty}\{\bigcap_{i=1}^{3} \mathcal{E}_{\mathrm{close},R_0,i}\} + \mathbf{I}_{0,\infty}\{\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \vee \mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le 2R_{\mathrm{stab}}\},
\end{aligned}$$

where we recall from (G.2)

$$c_1 := 6\max\{R_{\mathrm{stab}}(1 + 2c_{\mathbf{u}}\sqrt{L_{\mathrm{stab}}}), B_{\mathrm{stab}} + \sqrt{L_{\mathrm{stab}}}c_{\mathbf{K}}\}.$$

Let's now simplify the indictators. Restricting our attention to a with $\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \le R_{\mathrm{stab}}$, $\mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le R_{\mathrm{stab}} + \mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h, \mathsf{a}'_h)$ by the triangle inequality. Thus, we can replace $\mathbf{I}_{0,\infty}\{\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \vee \mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le 2R_{\mathrm{stab}}\}$ with $\mathbf{I}_{0,\infty}\{\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h, \mathsf{a}'_h) \le R_{\mathrm{stab}}\}$. We now recall the definitions

$$\begin{aligned}
\mathcal{E}_{\mathrm{close},R_0,1}(\mathsf{a}, \mathsf{a}') &= \{(\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a}, \mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a}, \mathsf{a}') \le C_{\mathrm{stab},1}(R_0)\} \\
\mathcal{E}_{\mathrm{close},R_0,2}(\mathsf{a}, \mathsf{a}') &= \{\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a}, \mathsf{a}') \le C_{\mathrm{stab},2}(R_0)\} \\
\mathcal{E}_{\mathrm{close},R_0,3}(\mathsf{a}, \mathsf{a}') &= \{\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}, \mathsf{a}') + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}, \mathsf{a}') \le C_{\mathrm{stab},3}(R_0)\}.
\end{aligned}$$

Again, recall that we take $R_0 = 2R_{\mathrm{stab}}$. Again, invoke the upper bounds of the form $\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\cdot,\cdot) \le \sqrt{L_{\mathrm{stab}}}\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\cdot)$ (as in (G.3)).Thus, $\bigcap_{i=1}^{3} \mathcal{E}_{\mathrm{close},R_0,i} \cap \{\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \vee \mathsf{rad}_{\mathbf{K}}(\mathsf{a}') \le 2R_{\mathrm{stab}}\}$ holds as soon as

$$\max\{\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}, \mathsf{a}'), \mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}, \mathsf{a}')\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}, \mathsf{a}')\} \le c_2,$$

where we recall from

$$c_2 := \min\left\{\frac{C_{\mathrm{stab},1}(2R_{\mathrm{stab}})}{4R_{\mathrm{stab}}\sqrt{L_{\mathrm{stab}}}}, \frac{C_{\mathrm{stab},2}(2R_{\mathrm{stab}})}{\sqrt{L_{\mathrm{stab}}}}, \frac{C_{\mathrm{stab},3}(2R_{\mathrm{stab}})}{4R_{\mathrm{stab}}}, \frac{1}{2R_{\mathrm{stab}}}\right\}.$$

In sum, for any a for which $\mathsf{rad}_{\mathbf{K}}(\mathsf{a}) \le R_{\mathrm{stab}}$, we have

$$\mathsf{d}_{\mathcal{A},R_0,\tau}(\mathsf{a}_h,\mathsf{a}_h' \mid r) \le \frac{c_1}{3}(\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h,\mathsf{a}_h') + \mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h,\mathsf{a}_h') + r\exp^{-\frac{\eta(\tau_c-\tau)}{3L_{\mathrm{stab}}}}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h,\mathsf{a}_h')) \\ + \mathbf{I}_{0,\infty}\{\max\{\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a},\mathsf{a}'),\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a},\mathsf{a}')\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a},\mathsf{a}')\} \le c_2\}.$$

To conclude, we observe that, for any nonnegative coefficients $a_1, a_2, a_3 > 0$ and sequences $v_{1,1:n}, v_{2,1:n}, v_{3,n} \ge 0$ in $\mathbb{R}^n$,

$$\sum_{i=1}^{3} a_i(\max_{j\in[n]} v_{i,j}) \le 3\max_{j\in[n]}\sum_{i=1} a_i v_{i,j}.$$

Thus, if we express $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_c}, \bar{\mathbf{x}}_{1:\tau_c}, \bar{\mathbf{K}}_{1:\tau_c})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_c}, \bar{\mathbf{x}}'_{1:\tau_c}, \bar{\mathbf{K}}'_{1:\tau_c})$, we can bound

$$\begin{aligned}\mathsf{d}_{\mathcal{A},R_0,\tau}(\mathsf{a}_h,\mathsf{a}_h' \mid r) &\le \bar{\mathsf{d}}_{\mathcal{A},\tau}(\mathsf{a},\mathsf{a}' \mid r) \\ &:= c_1 \max_{1\le k\le\tau_c}\left(\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + re^{-\frac{\eta(\tau_c-\tau)}{3L_{\mathrm{stab}}}}\|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|\right) \\ &\quad + \mathbf{I}_{0,\infty}\left\{\max_{1\le k\le\tau_c}\left(\max\left\{\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\|, \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\|, \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|\right\}\right) \le c_2\right\}.\end{aligned}$$

□

G.4. Stability guarantees for single control (sub-)trajectories.

At the heart of the IPS guarantees in Appendix G.3 above are two building blocks: one controller the perturbation of initial state around a regular (in the sense of Assumption 3.1) trajectory, and the second extending this guarantee to perturbations of control inputs and gains.

Lemma G.4 (Stability to State Perturbation). *Let $\bar{\boldsymbol{\rho}} = (\bar{\mathbf{x}}_{1:K+1}, \bar{\mathbf{u}}_{1:K}) \in \mathscr{P}_K$ be an $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular and feasible path, and let $\mathbf{K}_{1:K}$ be gains such that $(\bar{\boldsymbol{\rho}}, \mathbf{K}_{1:K})$ is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable. Assume, that $R_{\mathrm{stab}} \ge 1$, $L_{\mathrm{stab}} \ge 2\eta$. Fix another $\mathbf{x}_1$ and define another trajectory $\boldsymbol{\rho}$ via*

$$\mathbf{u}_k = \bar{\mathbf{u}}_k + \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k), \quad \mathbf{x}_{k+1} = \bar{\mathbf{x}}_k + \eta f_\eta(\mathbf{x}_k, \mathbf{u}_k)$$

Then, if $\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le \min\{(16L_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathrm{stab}}^2B_{\mathrm{stab}}^2)^{-1}, \frac{R_{\mathrm{dyn}}}{2R_{\mathrm{stab}}B_{\mathrm{stab}}}\}$, then

- $\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| \le 2B_{\mathrm{stab}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\|\beta_{\mathrm{stab}}^k$.
- $(\boldsymbol{\rho}, \mathbf{K}_{1:K})$ *is* $(R_{\mathrm{stab}}, 2B_{\mathrm{stab}}, L_{\mathrm{stab}})$*-stable.*
- $\|\mathbf{B}_k(\boldsymbol{\rho})\| \le L_{\mathrm{dyn}}$.

This lemma is proven in Appendix G.6, and the following proposition in Appendix G.7.

Proposition G.5 (Single Trajectory Stability Guarantee). *Let $\bar{\boldsymbol{\rho}} = (\bar{\mathbf{x}}_{1:K+1}, \bar{\mathbf{u}}_{1:K}) \in \mathscr{P}_K$ be $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular and feasible, and let $\mathbf{K}_{1:K}$ be such that $(\bar{\boldsymbol{\rho}}, \mathbf{K}_{1:K})$ is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable. Assume $R_{\mathrm{stab}} \ge 1$, $L_{\mathrm{stab}} \ge 2\eta$, and given another $\mathbf{x}_1, \mathbf{x}'_1 \in \mathcal{X}$, $\bar{\mathbf{u}}'_{1:K}$ and $\mathbf{K}'_{1:K}$, define trajectoris $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K})$ and $\boldsymbol{\rho}' = (\mathbf{x}'_{1:K+1}, \mathbf{u}'_{1:K})$*

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \eta f_\eta(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{u}_k = \bar{\mathbf{u}}_k + \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k)$$
$$\mathbf{x}'_{k+1} = \mathbf{x}'_k + \eta f_\eta(\mathbf{x}'_k, \mathbf{u}'_k), \quad \mathbf{u}'_k = \bar{\mathbf{u}}'_k + \mathbf{K}'_k(\mathbf{x}'_k - \bar{\mathbf{x}}_k)$$

Let all constants be as defined in Definition G.4, and define (recalling the stability exponent $\beta_{\mathrm{stab}} := (1 - \frac{\eta}{L_{\mathrm{stab}}})$) the terms

$$\mathrm{Err}_{\mathbf{u}} := \max_{k\in[K]}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\bar{\mathbf{u}}_j - \bar{\mathbf{u}}'_j\|^2\right)^{1/2}, \quad \mathrm{Err}_{\mathbf{K}} := \max_{k\in[K]}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\mathbf{K}_j - \mathbf{K}'_j\|^2\right)^{1/2}$$

Then, the conclusions of Lemma G.4 applies to the trajectory $\boldsymbol{\rho}$, and moreover, for all $1 \le k \le K$,

$$\|\mathbf{x}_{k+1} - \mathbf{x}'_{k+1}\| \le c_{\mathbf{u}}\mathrm{Err}_{\mathbf{u}} + (c_{\mathbf{K}}\mathrm{Err}_{\mathbf{K}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| + c_{\boldsymbol{\Delta}}\|\mathbf{x}_1 - \mathbf{x}'_1\|)\,\beta_{\mathrm{stab}}^{k/3},$$

provided that the following two conditions hold:

- *The above error terms satisfy*

$$\mathrm{Err}_{\mathbf{u}} \le C_{\mathbf{u}}, \quad \mathrm{Err}_{\mathbf{K}} \le C_{\mathbf{K}}, \quad \|\mathbf{x}_1 - \mathbf{x}_1'\| \le C_{\boldsymbol{\Delta}}, \quad \|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le C_{\hat{\mathbf{x}}}, \quad \mathrm{Err}_{\mathbf{K}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le C_{\mathbf{K},\hat{\mathbf{x}}}$$

- *In addition, if* $R_{\mathrm{dyn}} < \infty$, $\tilde{R}_{\mathrm{stab}} := \max\{1, R_{\mathrm{stab}}, \max_{1\le j\le K} \|\mathbf{K}_k'\|\}$ *and* $\Delta_{\mathbf{u},\infty} := \max_j \|\bar{\mathbf{u}}_j - \bar{\mathbf{u}}_j'\|$ *satisfy*

$$R_{\mathrm{dyn}} \ge (4\tilde{R}_{\mathrm{stab}} c_{\mathbf{u}} \sqrt{L_{\mathrm{stab}}} + 1)\Delta_{\mathbf{u},\infty} + 4\tilde{R}_{\mathrm{stab}} c_{\mathbf{K}} \|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| + 4\tilde{R}_{\mathrm{stab}} c_{\boldsymbol{\Delta}} \|\mathbf{x}_1 - \mathbf{x}_1'\|.$$

The proofs of both this proposition and the lemma before it consist of translating the differences in trajectories into recursions satisfying certain functional forms. Taking norms, we obtain scalar recursions whose solutions are upper bounded in a series of technical lemmas detailed in Appendix G.9. We believe these Proposition G.5 and Lemma G.4 are useful more broadly in the study of perturbation of non-linear control systems.

Notice that, for convenience, both the $\mathbf{x}$ and $\mathbf{x}'$ trajectories are stabilizing around the same $\bar{\mathbf{x}}$. This is for convenience, and simplifies the analysis. Indeed, difference generalizing to accomodate $\mathbf{x}'$ stabilizing around $\bar{\mathbf{x}}'$ can be accomplished by a change of variables in the $\bar{\mathbf{u}}'$, which is precisely what is done in deriving Proposition G.3 in the section that follows.

G.5. Deriving Proposition G.3 from Proposition G.5

The majority of this proof is (also) notational bookkeeping, whereby we convert two trajectories (in the abstract states/actions notation) into separate trajectories for each a sequence of $h = 1, 2, \ldots, H = T/\tau_{\mathrm{c}}$, to each of which we apply Proposition G.5.

Constructing the (perturbed) expert trajectory We begin by unfolding the generative process for abstract-states $\mathsf{s}_1, \ldots, \mathsf{s}_H$ in our proposition. Recall further that $\mathsf{s}_h = \boldsymbol{\rho}_{\mathrm{c},h}$ corresponds to the trajectory-chunk.

We let the (control) states and inputs for the corresponding sequence be denote as $(\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T})$ be generated as follows. Start with

$$\mathbf{x}_1 \leftarrow \mathsf{s}_1$$

drawn from the inital state distribution. Assume that we have constructed the states $\mathsf{s}_1, \ldots, \mathsf{s}_{h-1}$; this meangs in particular that we have constructed $\mathbf{x}_{1:t_h}, \mathbf{u}_{1:t_h-1}$, as well as the memory-chunks $\boldsymbol{\rho}_{\mathrm{m},1}, \ldots, \boldsymbol{\rho}_{\mathrm{m},j-1}$. We extend the construction to step $h+1$ as follows:

- Define $\mathbf{x}_{h,1} = \mathbf{x}_{t_h}$
- Select a perturbation of the state $\tilde{\mathsf{s}}_h = \tilde{\boldsymbol{\rho}}_{\mathrm{c},h} = (\tilde{\mathbf{x}}_{t_{h-1}:t_h}, \tilde{\mathbf{u}}_{t_{h-1}:t_h-1})$, with corresponding memory-chunk $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} = (\tilde{\mathbf{x}}_{t_{h-\tau_{\mathrm{m}}+1}:t_h}, \tilde{\mathbf{u}}_{t_h-\tau_{\mathrm{m}}+1:t_h-1})$. As per the proposition, $\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \tilde{\mathsf{s}}_h) \le r$. This means that $\|\mathbf{x}_{t_h} - \tilde{\mathbf{x}}_{t_h}\| \le r$.
- Draw $\mathsf{a}_h = \kappa_{t_h:t_h:t_h+\tau_{\mathrm{m}}-1} \sim \pi_h^\star(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h})$. We express

$$\kappa_t(\mathbf{x}) = \bar{\mathbf{u}}_t + \bar{\mathbf{K}}_t(\mathbf{x} - \bar{\mathbf{x}}_t), \quad t_h \le t \le t_{h+1} - 1,$$

  and reindexed trajectory

$$\kappa_{h,k} = \kappa_{t_h+k-1}.$$

  Denote

$$\bar{\mathbf{x}}_{h,k} = \bar{\mathbf{x}}_{t_h+k-1}, \quad \bar{\mathbf{u}}_{h,k} = \bar{\mathbf{u}}_{t_h+k-1}, \quad \bar{\mathbf{K}}_{h,k} = \bar{\mathbf{K}}_{t_h+k-1}$$

  and

$$\bar{\boldsymbol{\rho}}_{[h+1]} = (\bar{\mathbf{x}}_{h,1:\tau_{\mathrm{c}}+1}, \bar{\mathbf{u}}_{h,1:\tau_{\mathrm{c}}}).$$

- Moreover, because we assumg $\phi_{\mathcal{Z}} \circ \mathsf{W}_h \ll \phi_{\mathcal{Z}} \circ \mathsf{P}_h^\star$, we inherbit the conclusions of Lemma G.2. Hence, $\bar{\boldsymbol{\rho}}_{h+1}$ such be feasible, $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular, and $(\bar{\boldsymbol{\rho}}_h, \kappa_{h,1:\tau_c})$ is be $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable. In addition, Lemma G.2 ensures $\bar{\mathbf{x}}_{h,1} = \tilde{\mathbf{x}}_{t_h}$. Consequently, we have that the composite action map F_h satifies

$$F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h) = \bar{\boldsymbol{\rho}}_{[h+1]} = (\bar{\mathbf{x}}_{h,1:\tau_{\mathrm{c}}+1}, \bar{\mathbf{u}}_{h,1:\tau_c}). \tag{G.5}$$

- We execute a_h for τ_c steps from our *actual* state $\mathbf{x}_t$ (not $\tilde{\mathbf{x}}_t$), giving states and actions

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t), \quad \mathbf{u}_t = \kappa_t(\mathbf{x}_t), \qquad 1 \le t \le \tau_c.$$

  And define

$$\mathbf{x}_{h,k+1} = \mathbf{x}_{t_h+k}, \quad \mathbf{u}_{h,k} = \mathbf{u}_{t_h+k-1}, \quad 1 \le k \le \tau_c.$$

- Finally, define the chunks the trajectories $\boldsymbol{\rho}_{[h+1]} = (\mathbf{x}_{h,1:\tau_c+1}, \mathbf{x}_{h,1:\tau_c})$, which is equal to the next abstract-state

$$\mathsf{s}_{h+1} = (\mathbf{x}_{h,1:\tau_c+1}, \mathbf{u}_{h,1:\tau_c}) = (\mathbf{x}_{t_h:t_{h+1}}, \mathbf{u}_{t_h:t_{h+1}-1}) \tag{G.6}$$

Construction of the imitation trajectory. We now construct the imitation trajectory by setting $\mathbf{x}'_1 = \mathbf{x}_1$, and

- For each h, select $\mathsf{a}'_h = (\kappa'_{t_h:t_h+\tau_c-1}) \in \mathcal{K}^{\tau_c}$. Define the re-indexed primitive controllers

$$\kappa'_{h,k} = \kappa'_{t_h+k-1},$$

  and express

$$\kappa'_{h,k}(\mathbf{x}) = \bar{\mathbf{K}}'_{h,k}(\bar{\mathbf{x}} - \bar{\mathbf{x}}'_{k,h}) + \bar{\mathbf{u}}'_{h,k}.$$

- Execute a'_h for τ_c steps, giving states and actions

$$\mathbf{x}'_{t+1} = f(\mathbf{x}'_t, \mathbf{u}'_t), \quad \mathbf{x}_t = \kappa'_t(\mathbf{x}_t), \qquad 1 \le t \le \tau_c.$$

  And define

$$\mathbf{x}'_{h,k+1} = \mathbf{x}'_{t_h+k}, \quad \mathbf{u}'_{h,k} = \mathbf{u}'_{t_h+k-1}, \quad 1 \le k \le \tau_c.$$

- Finally, define the chunks

$$\mathsf{s}'_h = (\mathbf{x}'_{t_h:t_{h+1}}, \mathbf{u}'_{t_h:t_{h+1}-1}) = (\mathbf{x}'_{h,1:\tau_c+1}, \mathbf{u}'_{h,1:\tau_c}).$$

Further Notation. Let's define the following errors analgous to Proposition G.5.

$$\begin{aligned}
\mathrm{Err}^2_{\bar{\mathbf{u}},h} &= \max_{1 \le k \le \tau_c} \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \|\bar{\mathbf{u}}_{h,j} - \kappa'_{h,j}(\bar{\mathbf{x}}_{h,j})\|^2 \\
\mathrm{Err}^2_{\bar{\mathbf{K}},h} &= \max_{1 \le k \le \tau_c} \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \|\bar{\mathbf{K}}_{h,j} - \bar{\mathbf{K}}'_{h,j}\|^2 \\
\Delta_{\bar{\mathbf{u}},\infty,h} &:= \max_k \|\bar{\mathbf{u}}_{h,k} - \kappa'_{h,j}(\bar{\mathbf{x}}_{h,j})\|.
\end{aligned}$$

Importantly, in Proposition G.5, it is assumeded that other the primed and unprimed sequence stabilize to the same $\mathbf{x}_{h,k}$, whereas here, the primed sequence stabilized to $\mathbf{x}_{h,k'}$. This is addressed by replacing the role of $\mathbf{u}'_{h,k}$ with $\kappa'_{h,k}(\bar{\mathbf{x}}_{h,k})$.

G.5.1. Interpreting the error terms.

First, we unpack the above error terms.

Lemma G.6. *Suppose* $\max_h \mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h, \mathsf{a}'_h \mid r)$ *is finite. Then,*

$$\begin{aligned}
\mathrm{Err}_{\bar{\mathbf{K}},h} &= \mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h) \\
\mathrm{Err}_{\bar{\mathbf{u}},h} &\le \mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h) + R_0 \mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h) \\
\Delta_{\bar{\mathbf{u}},\infty,h} &= \mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + R_0 \mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h, \mathsf{a}'_h)
\end{aligned}$$

Proof of Lemma G.6. The equality of $\mathrm{Err}_{\bar{\mathbf{K}},h}$ follows from the reinxing $\bar{\mathbf{K}}_{h,k} = \bar{\mathbf{K}}_{t_h+k-1}$ and the definition of Definition G.9.Next, unpacking our notation of $\bar{\mathbf{u}}, \bar{\mathbf{u}}'$, we compute

$$\begin{aligned}\bar{\mathbf{u}}_{h,j} - \kappa_{h,j}(\bar{\mathbf{x}}_{h,j}) &= \bar{\mathbf{u}}_{h,j} - \bar{\mathbf{K}}'_{h,j}(\bar{\mathbf{x}}'_{h,j} - \bar{\mathbf{x}}_{h,j}) \\ &= \mathbf{u}'_{t_h+k-1} - \mathbf{u}'_{t_h+k-1} - \bar{\mathbf{K}}'_{t_h+k-1}(\mathbf{x}_{t_h+k-1} - \mathbf{x}'_{t_h+k-1})\end{aligned}$$

So that as long as $\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h, \mathsf{a}'_h \mid r)$ for all h, then $\|\bar{\mathbf{K}}_{h,j}\| \le R_0$. Thus

$$\|\bar{\mathbf{u}}_{h,j} - \bar{\mathbf{u}}'_{h,j}\| \le \|\mathbf{u}'_{t_h+k-1} - \mathbf{u}'_{t_h+k-1}\| + R_0\|\mathbf{x}_{t_h+k-1} - \mathbf{x}'_{t_h+k-1}\|$$

and thus by the triangle and moving the max outside the sum,

$$\begin{aligned}\mathrm{Err}_{\bar{\mathbf{u}},h} &= \max_{1\le k\le \tau_c}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\bar{\mathbf{u}}_{h,j} - \bar{\mathbf{u}}'_{h,j}\|^2\right)^{1/2} \\ &\le \max_{1\le k\le \tau_c}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\mathbf{u}'_{t_h+k-1} - \mathbf{u}'_{t_h+k-1}\|^2\right)^{1/2} \\ &\quad + R_0\max_{1\le k\le \tau_c}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\mathbf{x}'_{t_h+k-1} - \mathbf{x}'_{t_h+k-1}\|^2\right)^{1/2} \\ &\le \mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h) + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h).\end{aligned}$$

The inequality $\Delta_{\bar{\mathbf{u}},\infty,h} \le \mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h, \mathsf{a}'_h) + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h, \mathsf{a}'_h)$ follows similarly. □

G.5.2. An intermediate guarantee.

Next, we establish an intermediate guarantee, from which Proposition G.3 is readily derived.

Lemma G.7. *Suppose* $\max_h \mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h, \mathsf{a}'_h \mid r)$ *is finite, and further that* $\tau_c \ge 3L_{\mathrm{stab}}\log(2c_{\boldsymbol{\Delta}})/\eta$. *Then,*

- *For all* $k \in \{0, \dots, \tau_c\}$ *and* $h \in [H]$,

$$\|\mathbf{x}_{h,k+1} - \mathbf{x}'_{h+1,k+1}\| \le \max_{h'}\left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + 2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{k/3}\right)$$

,

- *For all* $h \in [H]$ *and* $1 \le k \le \tau_c$,

$$\|\mathbf{x}_{h,k} - \bar{\mathbf{x}}_{h,k}\| \le 2B_{\mathrm{stab}}r\beta_{\mathrm{stab}}^{k-1}.$$

Proof of Lemma G.7. First, an algebraic computation. Observe that $\log(1/\beta_{\mathrm{stab}}) = \log(1/(1-\frac{\eta}{L_{\mathrm{stab}}})) = -\log(1-\frac{\eta}{L_{\mathrm{stab}}}) \ge \frac{\eta}{L_{\mathrm{stab}}}$. Hence, if $\tau_c \ge 3L_{\mathrm{stab}}\log(2c_{\boldsymbol{\Delta}})/\eta$, we have $\tau_c \ge 3\log(2c_{\boldsymbol{\Delta}})/\log(1/\beta_{\mathrm{stab}})$, so that

$$c_{\boldsymbol{\Delta}}\beta_{\mathrm{stab}}^{\tau_c/3} \le 1/2. \tag{G.7}$$

We continue. Suppose $\max_h \mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h, \mathsf{a}'_h \mid r)$ is finite. Then, from the definition of $\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}$ in Definition G.10, the constants Definition G.5, and the inequalities in Lemma G.6 above, we can check that

$$\begin{aligned}&\max_h \mathrm{Err}_{\bar{\mathbf{u}},h} \le \max_h(\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h) + R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h)) \le C_{\mathrm{stab},1}(R_0) \\ &\max_h \mathrm{Err}_{\bar{\mathbf{K}},h} = \max_h \mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a}_h, \mathsf{a}'_h) \le C_{\mathrm{stab},2}(R_0) \\ &\max_h \Delta_{\bar{\mathbf{u}},\infty,h} = \max_h \mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h, \mathsf{a}'_h) \le C_{\mathrm{stab},3}(R_0) \\ &r \le C_{\mathrm{stab},4}(R_0).\end{aligned}$$

We begin with an induction on states $\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|$ for $h \ge 1$. Recall the assumption that $c_{\boldsymbol{\Delta}}\beta_{\mathrm{stab}}^{\tau_c/3} \le 1/2$. We prove inductively that

$$\forall h \ge 1, \|\mathbf{x}'_{h,1} - \mathbf{x}_{h,1}\| \le \max_{h'} \left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + 2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{\tau_c/3})\right) \tag{G.8}$$

For the base case, we have $\mathbf{x}_{1,1} = \mathbf{x}'_{1,1}$. Now, suppose the result holds up to some $h \ge 1$. Using the definitions of various constants in Definitions G.4 and G.5, and $r \ge \|\bar{\mathbf{x}}_{h,1} - \mathbf{x}_{h,1}\|$, as well as our inductive hypotheis, one can check that

$$\begin{aligned}
&\mathrm{Err}_{\bar{\mathbf{u}},h} \le C_{\mathbf{u}}, \quad \mathrm{Err}_{\bar{\mathbf{K}},h} \le C_{\mathbf{K}}\\
&\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\| \le \max_{h'} \left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + 2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{\tau_c/3}\right) \le C_{\boldsymbol{\Delta}}\\
&\|\bar{\mathbf{x}}_{h,1} - \mathbf{x}_{h,1}\| \le C_{\hat{\mathbf{x}}}, \quad \|\bar{\mathbf{x}}_{h,1} - \mathbf{x}_{h,1}\|\mathrm{Err}_{\mathbf{K}} \le C_{\mathbf{K},\hat{\mathbf{x}}}\\
&(4R_0c_{\mathbf{u}}\sqrt{L_{\mathrm{stab}}} + 1)\Delta_{\mathbf{u},\infty} + 4R_0c_{\mathbf{K}}\|\mathbf{x}_{h,1} - \bar{\mathbf{x}}_{h,1}\| + 4R_0c_{\boldsymbol{\Delta}}\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|. \le R_{\mathrm{dyn}}.
\end{aligned}$$

Then, by Proposition G.5,

$$\begin{aligned}
\|\mathbf{x}_{h+1,1} - \mathbf{x}'_{h+1,1}\| &= \|\mathbf{x}_{h,\tau_c+1} - \mathbf{x}'_{h,\tau_c+1}\|\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h+1} + \left(c_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h+1}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| + c_{\boldsymbol{\Delta}}\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|\right)\beta_{\mathrm{stab}}^{\tau_c/3}\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h+1} + \left(c_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h+1}r + c_{\boldsymbol{\Delta}}\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|\right)\beta_{\mathrm{stab}}^{\tau_c/3} \quad (c_{\boldsymbol{\Delta}}\beta_{\mathrm{stab}}^{\tau_c/3} \le \tfrac{1}{2}, \text{ as established in (G.7)})\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h} + rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h}\beta_{\mathrm{stab}}^{\tau_c/3} + \frac{1}{2}\max_{h'}\left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + 2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{\tau_c/3})\right)\\
&\hspace{20em}\text{(inductive hypothesis)}\\
&\le \max_{h'}\left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + 2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{\tau_c/3})\right)
\end{aligned}$$

This establishes (G.8). A second invocation of Proposition G.5 gives

$$\begin{aligned}
&\|\mathbf{x}_{h,k+1} - \mathbf{x}'_{h+1,k+1}\|\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h+1} + \left(c_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h+1}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| + c_{\boldsymbol{\Delta}}\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|\right)\beta_{\mathrm{stab}}^{k/3}\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h+1} + \left(c_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h+1}r + c_{\boldsymbol{\Delta}}\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|\right)\beta_{\mathrm{stab}}^{k/3}\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h+1} + rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h+1}\beta_{\mathrm{stab}}^{k/3} + \frac{1}{2}\|\mathbf{x}_{h,1} - \mathbf{x}'_{h,1}\|\\
&\le c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h+1} + rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h}\beta_{\mathrm{stab}}^{k/3} + \max_{h'}\left(c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{\tau_c/3})\right)\\
&\le \max_{h'}\left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'} + 2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{k/3})\right).
\end{aligned}$$

Moreover, as Proposition G.5 implies that the conclusions of Lemma G.4 also hold, we further find that

$$\|\mathbf{x}_{h,k} - \bar{\mathbf{x}}_{h,k}\| \le 2B_{\mathrm{stab}}\|\mathbf{x}_{h,k} - \bar{\mathbf{x}}_{h,k}\|\beta_{\mathrm{stab}}^{k-1} \le 2B_{\mathrm{stab}}r\beta_{\mathrm{stab}}^{k-1},$$

as needed. □

G.5.3. Concluding the proof of Proposition G.3.

Completing the proof of Proposition G.3. Let us start with the first item, bound $\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}$. We may assume that $\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h, \mathsf{a}'_h \mid r)$ is finite for all h.

Controlling $\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_h,\mathsf{s}'_h)$. Not that $\mathsf{s}_1=\mathsf{s}'_1$. For any $2\le h\le H+1$,

$$\begin{aligned}
\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_h,\mathsf{s}'_h) &:= \max_{t\in[t_h-\tau:t_h]}\|\mathbf{x}_t-\mathbf{x}'_t\|\\
&= \max_{\tau_c-\tau\le k\le\tau_c}\|\mathbf{x}_{h-1,1+k}-\mathbf{x}'_{h-1,k+1}\| && \text{(our indexing scheme)}\\
&\le \max_{\tau_c-\tau\le k\le\tau_c}\max_{h'}\left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'}+2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{k/3}\right) && \text{(Lemma G.7)}\\
&= \max_{h'}\left(2c_{\mathbf{u}}\mathrm{Err}_{\bar{\mathbf{u}},h'}+2rc_{\mathbf{K}}\mathrm{Err}_{\bar{\mathbf{K}},h'}\beta_{\mathrm{stab}}^{(\tau_c-\tau)/3}\right)\\
&\le \max_{h'}\left(2c_{\mathbf{u}}(\mathsf{d}_{\mathcal{A},\mathbf{u},\ell_2}(\mathsf{a}_{h'},\mathsf{a}'_{h'})+R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\ell_2}(\mathsf{a}_{h'},\mathsf{a}'_{h'}))+2rc_{\mathbf{K}}\mathsf{d}_{\mathcal{A},\mathbf{K},\ell_2}(\mathsf{a}_h,\mathsf{a}'_h)\beta_{\mathrm{stab}}^{(\tau_c-\tau)/3}\right) && \text{(Lemma G.6)}\\
&\le \max_{h'}\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_{h'},\mathsf{a}'_{h'}\mid r) && \text{(Definition G.10)}
\end{aligned}$$

That is,

$$\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_h,\mathsf{s}'_h)\le\max_h\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h,\mathsf{a}'_h\mid r) \tag{G.9}$$

Bounding $\mathsf{d}_{\mathcal{S},\tau}$. To bound $\mathsf{d}_{\mathcal{S},\tau}$, we also need to account for differents in inputs. We have

$$\begin{aligned}
&\mathbf{u}_{h,k}-\mathbf{u}'_{h,k}=\kappa_{h,k}(\mathbf{x}_{h,k})-\kappa'_{h,k}(\mathbf{x}'_{h,k})\\
&=\bar{\mathbf{u}}_{h,k}-\bar{\mathbf{u}}'_{h,k}+\bar{\mathbf{K}}_t(\mathbf{x}_{h,k}-\bar{\mathbf{x}}_{h,k})-\bar{\mathbf{K}}'_{h,k}(\mathbf{x}'_{h,k}-\bar{\mathbf{x}}'_{h,k})\\
&=\bar{\mathbf{u}}_{h,k}-\bar{\mathbf{u}}'_{h,k}+(\bar{\mathbf{K}}_{h,k}-\bar{\mathbf{K}}'_{h,k})(\mathbf{x}_{h,k}-\bar{\mathbf{x}}_{h,k})+\bar{\mathbf{K}}'_{h,k}(\mathbf{x}_t-\mathbf{x}'_{h,k}-(\bar{\mathbf{x}}_{h,k}-\bar{\mathbf{x}}'_{h,k}))
\end{aligned}$$

Where $\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h,\mathsf{a}'_h\mid r)$ is finite for all h, then $\|\bar{\mathbf{K}}'_{h,k}\|=\|\bar{\mathbf{K}}_{t_h+k-1}\|\le R_0$. Thus,

$$\begin{aligned}
&\|\mathbf{u}_{h,k}-\mathbf{u}'_{h,k}\|\\
&\le\|\bar{\mathbf{u}}_{h,k}-\bar{\mathbf{u}}'_{h,k}\|+R_0\|\bar{\mathbf{x}}_{h,k}-\bar{\mathbf{x}}'_{h,k}\|+R_0\|\bar{\mathbf{x}}_{h,k}-\bar{\mathbf{x}}'_{h,k}\|+\|\bar{\mathbf{K}}_{h,k}-\bar{\mathbf{K}}'_{h,k}\|\|\mathbf{x}_{h,k}-\bar{\mathbf{x}}_{h,k}\|\\
&\le\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+R_0\|\bar{\mathbf{x}}_{h,k}-\bar{\mathbf{x}}'_{h,k}\|+\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h,\mathsf{a}')\|\mathbf{x}_{h,k}-\bar{\mathbf{x}}_{h,k}\|\\
&\le\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+2rB_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-1}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h,\mathsf{a}')+R_0\|\bar{\mathbf{x}}_{h,k}-\bar{\mathbf{x}}'_{h,k}\|
\end{aligned}$$

Hence,

$$\begin{aligned}
&\max_h\mathsf{d}_{\mathcal{S},\tau}(\mathsf{s}_h,\mathsf{s}'_h)\\
&=\max_h\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{a}_h,\mathsf{a}'_h)\vee\max_h\max_{\tau_c-\tau\le k\le\tau_c-1}\|\mathbf{u}_{h,k+1}-\mathbf{u}'_{h,k+1}\|\\
&=\max_h\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{a}_h,\mathsf{a}'_h)+R_0\max_h\max_{\tau_c-\tau\le k\le\tau_c-1}\|\bar{\mathbf{x}}_{h,k+1}-\bar{\mathbf{x}}'_{h,k+1}\|\\
&\quad+\max_h\max_{\tau_c-\tau\le k\le\tau_c-1}\left(\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+2rB_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h,\mathsf{a}')\right)\\
&=(1+R_0)\max_h\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{a}_h,\mathsf{a}'_h)\\
&\quad+\max_h\left(\underbrace{\mathsf{d}_{\mathcal{A},\mathbf{u},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+R_0\mathsf{d}_{\mathcal{A},\mathbf{x},\infty}(\mathsf{a}_h,\mathsf{a}'_h)+2rB_{\mathrm{stab}}\beta_{\mathrm{stab}}^{\tau_c-\tau}\mathsf{d}_{\mathcal{A},\mathbf{K},\infty}(\mathsf{a}_h,\mathsf{a}'_h)}_{:=\max_h\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{u}}(\mathsf{a}_h,\mathsf{a}'_h|r)}\right)\\
&\le(1+R_0)\max_h\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h,\mathsf{a}'_h\mid r)+\max_h\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{u}}(\mathsf{a}_h,\mathsf{a}'_h\mid r) && \text{((G.9))}\\
&\le\max_h 2\left((1+R_0)\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{x}}(\mathsf{a}_h,\mathsf{a}'_h\mid r)+\mathsf{d}_{\mathcal{A},R_0,\tau,\mathbf{u}}(\mathsf{a}_h,\mathsf{a}'_h\mid r)\right).
\end{aligned}$$

Bounding $\mathsf{d}_{\mathcal{S},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h))$. Next, by (G.5) and (G.6) that

$$\begin{aligned}
\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) &= \max_{\tau_c-\tau\le k\le\tau_c} \|\mathbf{x}_{h,k+1} - \bar{\mathbf{x}}_{h,k+1}\| && \text{((G.5) and (G.6))}\\
&= \max_{\tau_c-\tau\le k\le\tau_c} 2B_{\mathrm{stab}} r \beta_{\mathrm{stab}}^{k} && \text{(Lemma G.7)}\\
&= 2B_{\mathrm{stab}} r \beta_{\mathrm{stab}}^{(\tau_c-\tau)}.
\end{aligned}$$

Thus, We have

$$\begin{aligned}
\mathsf{d}_{\mathcal{S},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) &= \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) \vee \max_{\tau_c-\tau\le k\le\tau_c-1} \|\mathbf{u}_{h,k+1} - \bar{\mathbf{u}}_{h,k+1}\|\\
&= \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) \vee \max_{\tau_c-\tau\le k\le\tau_c-1} \|\kappa_{h,k+1}(\mathbf{x}_{h,k+1}) - \bar{\mathbf{u}}_{h,k+1})\|\\
&= \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) \vee \max_{\tau_c-\tau\le k\le\tau_c-1} \|(\bar{\mathbf{K}}_{h,k+1}(\mathbf{x}_{h,k+1} - \bar{\mathbf{x}}_{h,k+1}) + \bar{\mathbf{u}}_{h,k+1}) - \bar{\mathbf{u}}_{h,k+1})\|\\
&\le \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) \vee \max_{\tau_c-\tau\le k\le\tau_c-1} \|\bar{\mathbf{K}}_{h,k+1}\|\|\mathbf{x}_{h,k+1} - \bar{\mathbf{x}}_{h,k+1}\|\\
&\overset{(i)}{\le} \mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)) \vee \max_{\tau_c-\tau\le k\le\tau_c-1} R_{\mathrm{stab}}\|\mathbf{x}_{h,k+1} - \bar{\mathbf{x}}_{h,k+1}\|\\
&\le (1+R_{\mathrm{stab}})\mathsf{d}_{\mathcal{S},\mathbf{x},\tau}(\mathsf{s}_{h+1}, F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h))\\
&\le 2(1+R_{\mathrm{stab}})B_{\mathrm{stab}} r \beta_{\mathrm{stab}}^{(\tau_c-\tau)}
\end{aligned}$$

where in (i), we used $\|\bar{\mathbf{K}}_{h,k+1}\| \le R_{\mathrm{stab}}$ because $(\bar{\boldsymbol{\uprho}}_{[h+1]}, \kappa_{h,1:\tau_c})$ is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable, so that the gains are bounded in operator norm by R_{stab}. □

G.6. Proof of Lemma G.4 (state perturbation)

Define $\bar{\Delta}_{\mathbf{x},k} = \mathbf{x}_k - \bar{\mathbf{x}}_k$. Then

$$\begin{aligned}
\bar{\Delta}_{\mathbf{x},k+1} &= \bar{\Delta}_{\mathbf{x},k} + \eta\left(f_\eta(\mathbf{x}_k, \bar{\mathbf{u}}_k + \bar{\mathbf{K}}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k) - f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\right)\\
&= \bar{\Delta}_{\mathbf{x},k} + \eta(\mathbf{A}_k + \mathbf{B}_k\mathbf{K}_k)\Delta_{\mathbf{x},k} + \mathrm{rem}_k,
\end{aligned} \tag{G.10}$$

where

$$\mathrm{rem}_k = f_\eta(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k)) - f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k) - (\mathbf{A}_k + \mathbf{B}_k\mathbf{K}_k)\bar{\Delta}_{\mathbf{x},k}.$$

Claim G.8. *Take $R_{\mathrm{stab}} \ge 1$, and suppose $\|\bar{\Delta}_{\mathbf{x},k}\| \le R_{\mathrm{dyn}}/R_{\mathrm{stab}}$. Then,*

$$\|\bar{\mathbf{x}}_k - \mathbf{x}_k\| \vee \|\bar{\mathbf{u}}_k - \mathbf{u}_k\| \le R_{\mathrm{dyn}}, \tag{G.11}$$

and $\|\mathrm{rem}_k\| \le M_{\mathrm{dyn}} R_{\mathrm{stab}}^2 \|\bar{\Delta}_{\mathbf{x},k}\|^2$.

Proof. Let $\mathbf{u}_k = \bar{\mathbf{u}}_k + \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k)$. The conditions of the claim imply $\|\mathbf{u}_k - \bar{\mathbf{u}}_k\| \vee \|\mathbf{x}_k \vee \bar{\mathbf{x}}_k\| \le R_{\mathrm{dyn}}$. From Taylor's theorem and the fact that $\bar{\boldsymbol{\uprho}}$ is $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular imply that

$$\begin{aligned}
\|f_\eta(\mathbf{x}_k, \mathbf{u}_k) - f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| &\le \frac{1}{2} M_{\mathrm{dyn}}(\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^2 + \|\mathbf{u}_k - \bar{\mathbf{u}}_k\|)\\
&\le \frac{1}{2}(1 + R_{\mathrm{stab}}^2) M_{\mathrm{dyn}} \|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^2 \le R_{\mathrm{stab}}^2 M_{\mathrm{dyn}} \|\bar{\Delta}_{\mathbf{x},k}\|^2,
\end{aligned}$$

where again use $R_{\mathrm{stab}} \ge 1$ above. □

Solving the recursion from (G.10), we have

$$\bar{\Delta}_{\mathbf{x},k+1} = \eta \sum_{j=1}^{k} \boldsymbol{\Phi}_{\mathrm{cl},k+1,j+1} \mathrm{rem}_k + \boldsymbol{\Phi}_{\mathrm{cl},k+1,1} \bar{\Delta}_{\mathbf{x},1}.$$

Set $\beta_{\mathrm{stab}} := (1 - \frac{\eta}{L_{\mathrm{stab}}})$, so that $M := \frac{\eta}{\beta_{\mathrm{stab}}^{-1}-1} = L_{\mathrm{stab}}$. Further, recall $R_0 \le R_{\mathrm{dyn}}/R_{\mathrm{stab}}$. By assumpion, $\mathbf{\Phi}_{\mathrm{cl},k,j} \le B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}$, so using Claim G.8 implies that, if $\max_{j\in[k]} \|\bar{\Delta}_{\mathbf{x},j}\| \le R_0 \le R_{\mathrm{dyn}}/R_{\mathrm{stab}}$ for all $j \in [k]$,

$$\|\bar{\Delta}_{\mathbf{x},k+1}\| \le \eta \sum_{j=1}^{k} B_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathrm{stab}}^2 \beta_{\mathrm{stab}}^{k-j} \|\bar{\Delta}_{\mathbf{x},j}\|^2 + B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\|\bar{\Delta}_{\mathbf{x},1}\|.$$

Appling Lemma G.17 with $\alpha = 0$, $C_1 = B_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathrm{stab}}^2$, and $C_2 = B_{\mathrm{stab}} \ge 1$ and $M = L_{\mathrm{stab}}$ (noting $\beta_{\mathrm{stab}} \ge 1/2$), it holds that for $\|\bar{\Delta}_{\mathbf{x},1}\| = \varepsilon_1 \le 1/4MC_1C_3 = 1/4L_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathrm{stab}}^2 B_{\mathrm{stab}}^2$,

$$\|\bar{\Delta}_{\mathbf{x},k+1}\| \le 2B_{\mathrm{stab}}\|\bar{\Delta}_{\mathbf{x},1}\|(1 - \frac{\eta}{L_{\mathrm{stab}}})^k.$$

To ensure the inductive hypothesis that $\max_{j\in[k]} \|\bar{\Delta}_{\mathbf{x},j}\| \le R_{\mathrm{dyn}} R_{\mathrm{stab}}$, it suffices to ensure that $2B_{\mathrm{stab}}\|\bar{\Delta}_{\mathbf{x},1}\| \le R_0$, which is assumed by the lemma. Thus, we have shown that, if

$$\|\bar{\Delta}_{\mathbf{x},1}\| \le \min\{1/2B_{\mathrm{stab}}R_0, 1/8L_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathrm{stab}}^2 B_{\mathrm{stab}}^2\},$$

it holds that $\|\bar{\Delta}_{\mathbf{x},k+1}\| \le 2B_{\mathrm{stab}}\|\bar{\Delta}_{\mathbf{x},1}\|(1 - \frac{\eta}{L_{\mathrm{stab}}})^k \le R_0$ for all k.

Next, we adress the stability of the gains for the perturbed trajectory $\boldsymbol{\rho}$. Using $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regularity of $\bar{\boldsymbol{\rho}}$ and (G.11),

$$\begin{aligned}
&\|\mathbf{A}_k(\boldsymbol{\rho}) + \mathbf{B}_k(\boldsymbol{\rho})\mathbf{K}_k - \mathbf{A}_k(\bar{\boldsymbol{\rho}}) + \mathbf{B}_k(\bar{\boldsymbol{\rho}})\mathbf{K}_k\| \\
&= \left\| \begin{bmatrix} \mathbf{A}_k(\boldsymbol{\rho}) - \mathbf{A}_k(\bar{\boldsymbol{\rho}}) & \hat{\mathbf{B}}_k(\boldsymbol{\rho}) - \mathbf{B}_k(\bar{\boldsymbol{\rho}}) \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_k \end{bmatrix} \right\| \\
&= \left\| (\nabla f_\eta(\hat{\mathbf{x}}_k, \mathbf{u}_k) - \nabla f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)) \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_k \end{bmatrix} \right\| \\
&\le M_{\mathrm{dyn}} \|(\mathbf{x}_k - \bar{\mathbf{x}}_k, \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k))\| \left\| \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_k \end{bmatrix} \right\| \\
&= M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\| \left\| \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_k \end{bmatrix} \right\|^2 \le M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|(1 + \|\mathbf{K}_k\|_{\mathrm{op}}^2) \\
&= M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\| \left\| \begin{bmatrix} \mathbf{I} \\ \mathbf{K}_k \end{bmatrix} \right\|^2 \le M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|(1 + \|\mathbf{K}_k\|_{\mathrm{op}}^2) \\
&\le 2R_{\mathrm{stab}}^2 M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\| \\
&\le 4B_{\mathrm{stab}} R_{\mathrm{stab}}^2 M_{\mathrm{dyn}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\|\beta_{\mathrm{stab}}^{k-1}, \quad \beta_{\mathrm{stab}} = (1 - \frac{\eta}{L_{\mathrm{stab}}}).
\end{aligned}$$

Invoking Lemma G.20 with $\beta_{\mathrm{stab}} \ge 1/2$, $\|\hat{\mathbf{\Phi}}_{\mathrm{cl},k,j}\| \le 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}$ for all j, k provided that $4B_{\mathrm{stab}} R_{\mathrm{stab}}^2 M_{\mathrm{dyn}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le 1/4B_{\mathrm{stab}}L_{\mathrm{stab}}$, which requires $\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le 1/16B_{\mathrm{stab}}^2 R_{\mathrm{stab}}^2 L_{\mathrm{stab}} M_{\mathrm{dyn}}$.

The last part of the lemma uses $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regularity of $\bar{\boldsymbol{\rho}}$ and (G.11).

G.7. Proof of Proposition G.5 (input and gain perturbation)

Recall the trajectories $\bar{\mathbf{x}}_{k+1} = \bar{\mathbf{x}}_k + \eta f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)$, and

$$\begin{aligned}
\mathbf{x}_{k+1} &= \mathbf{x}_k + \eta f_\eta(\mathbf{x}_k, \mathbf{u}_k), \quad & \mathbf{u}_k &= \bar{\mathbf{u}}_k + \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k) \\
\mathbf{x}'_{k+1} &= \mathbf{x}'_k + \eta f_\eta(\mathbf{x}'_k, \mathbf{u}'_k), \quad & \mathbf{u}'_k &= \bar{\mathbf{u}}'_k + \mathbf{K}'_k(\mathbf{x}'_k - \bar{\mathbf{x}}_k).
\end{aligned}$$

Further introduce the shorthand $\hat{\mathbf{A}}_k = \mathbf{A}_k(\hat{\boldsymbol{\rho}})$, $\hat{\mathbf{B}}_k = \mathbf{B}_k(\hat{\boldsymbol{\rho}})$, $\hat{\mathbf{A}}_{\mathrm{cl},k} = \hat{\mathbf{A}}_k + \hat{\mathbf{B}}_k + \mathbf{K}_k$, as well as

$$\begin{aligned}
&\Delta_{\mathbf{x},k} = \mathbf{x}'_k - \mathbf{x}_k, \quad \Delta_{\mathbf{u},k} = \bar{\mathbf{u}}'_k - \bar{\mathbf{u}}_k, \quad \Delta_{\mathbf{K},k} = \mathbf{K}'_k - \mathbf{K} \\
&\tilde{\Delta}_{\mathbf{x},k} = \mathbf{x}'_k - \bar{\mathbf{x}}_k, \quad \bar{\Delta}_{\mathbf{x},k} = \mathbf{x}_k - \bar{\mathbf{x}}_k,
\end{aligned}$$

Then,

$$\begin{aligned}
\Delta_{\mathbf{x},k+1} &= \Delta_{\mathbf{x},k} + \eta\left(f(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \tilde{\mathbf{K}}_k\tilde{\Delta}_{\mathbf{x},k}) - f(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k})\right)\\
&= \Delta_{\mathbf{x},k} + \eta\left(f(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k}) - f(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k})\right)\\
&\quad + \eta(\mathrm{rem}_{k,1})\\
&= \Delta_{\mathbf{x},k} + \eta\left(\underbrace{\frac{\partial}{\partial x}f(\mathbf{x}_k, \mathbf{u}_k)}_{=\hat{\mathbf{A}}_k}\Delta_{\mathbf{x},k} + \underbrace{\frac{\partial}{\partial u}f(\mathbf{x}_k, \mathbf{u}_k)}_{=\hat{\mathbf{B}}_k}(\Delta_{\mathbf{u},k} + \mathbf{K}_k\underbrace{\tilde{\Delta}_{\mathbf{x},k} - \bar{\Delta}_{\mathbf{x},k}}_{\Delta_{\mathbf{x},k}})\right)\\
&\quad + \eta(\mathrm{rem}_{k,1} + \mathrm{rem}_{k,2})\\
&= \Delta_{\mathbf{x},k} + \eta\left(\hat{\mathbf{A}}_{\mathrm{cl},k}\Delta_{\mathbf{x},k} + \hat{\mathbf{B}}_k\Delta_{\mathbf{u},k}\right) + \eta(\mathrm{rem}_{k,1} + \mathrm{rem}_{k,2}).
\end{aligned}$$

where, above

$$\begin{aligned}
\mathrm{rem}_{k,1} &= f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}'_k\tilde{\Delta}_{\mathbf{x},k}) - f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k})\\
\mathrm{rem}_{k,2} &= f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k}) - f(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k})\\
&\quad - \frac{\partial}{\partial x}f_\eta(\mathbf{x}_k, \mathbf{u}_k)\Delta_{\mathbf{x},k} + \frac{\partial}{\partial u}f_\eta(\mathbf{x}_k, \mathbf{u}_k)(\Delta_{\mathbf{u},k} + \mathbf{K}_k(\tilde{\Delta}_{\mathbf{x},k} - \bar{\Delta}_{\mathbf{x},k})).
\end{aligned}$$

Solving the recursion,

$$\Delta_{\mathbf{x},k+1} = \sum_{j=1}^{k}\hat{\mathbf{\Phi}}_{\mathrm{cl},k+1,j+1}(\hat{\mathbf{B}}_j\Delta_{\mathbf{u},j} + \eta(\mathrm{rem}_{j,1} + \mathrm{rem}_{j,2})) + \hat{\mathbf{\Phi}}_{\mathrm{cl},k+1,1}\Delta_{\mathbf{x},1}$$

Recall that Lemma G.4 implies $(\mathbf{K}_{1:K}, \boldsymbol{\rho})$ is $(R_{\mathrm{stab}}, 2B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable. Thus, recalling $\beta_{\mathrm{stab}} = (1 - \frac{\eta}{L_{\mathrm{stab}}}) \in [1/2, 1)$, we have

$$\begin{aligned}
\|\Delta_{\mathbf{x},k+1}\| &\le \eta\sum_{j=1}^{k}\|\hat{\mathbf{\Phi}}_{\mathrm{cl},k+1,j+1}\|(\|\hat{\mathbf{B}}_j\|\|\Delta_{\mathbf{u},j}\| + \|\mathrm{rem}_{j,1}\| + \|\mathrm{rem}_{j,2}\|)) + \|\hat{\mathbf{\Phi}}_{\mathrm{cl},k+1,1}\|\|\Delta_{\mathbf{x},1}\|\\
&\le \eta\sum_{j=1}^{k}2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}(L_{\mathrm{dyn}}\|\Delta_{\mathbf{u},j}\| + \|\mathrm{rem}_{j,1}\| + \|\mathrm{rem}_{j,2}\|)) + 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\|\Delta_{\mathbf{x},1}\|.
\end{aligned}$$

Let us now bound each of these remainder terms. The following claim, as well as all subsequent claims, is proven at the end of the section.

Claim G.9. *Suppose that it holds that for a given k, it holds that*

$$\|\Delta_{\mathbf{x},k}\| \le c_{\mathbf{u}}\mathrm{Err}_{\mathbf{u}} + c_{\mathbf{K}}\mathrm{Err}_{\mathbf{K}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| + c_{\boldsymbol{\Delta}}\|\mathbf{x}_1 - \mathbf{x}'_1\| \tag{G.12}$$

Then,

$$\begin{aligned}
\|\mathrm{rem}_{k,1}\| &\le L_{\mathrm{dyn}}\|\Delta_{\mathbf{K},k}\|(\|\bar{\Delta}_{\mathbf{x},k}\| + \|\Delta_{\mathbf{x},k}\|)\\
\|\mathrm{rem}_{k,2}\| &\le \frac{3}{2}M_{\mathrm{dyn}}R_{\mathrm{stab}}^2\|\Delta_{\mathbf{x},k}\|^2 + M_{\mathrm{dyn}}\|\Delta_{\mathbf{u},k}\|^2
\end{aligned}$$

We now proceed by strong induction on the condition in (G.12). Observe that if this condition holds for all $1 \le j \le k$, we have

$$
\begin{aligned}
\|\Delta_{\mathbf{x},k+1}\| &\le \eta \sum_{j=1}^{k} 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}(L_{\mathrm{dyn}}\|\Delta_{\mathbf{u},j}\| + \|\mathrm{rem}_{j,1}\| + \|\mathrm{rem}_{j,2}\|)) + 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\|\Delta_{\mathbf{x},1}\| \\
&\le \underbrace{\eta \sum_{j=1}^{k} 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}\left(L_{\mathrm{dyn}}\|\Delta_{\mathbf{u},j}\| + M_{\mathrm{dyn}}\|\Delta_{\mathbf{u},j}\|^2\right)}_{=\mathrm{Term}_{1,k}} \\
&\quad + \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j}\left(\underbrace{3B_{\mathrm{stab}}M_{\mathrm{dyn}}}_{C_1}\|\Delta_{\mathbf{x},j}\|^2 + \underbrace{2B_{\mathrm{stab}}L_{\mathrm{dyn}}}_{C_2}\|\Delta_{\mathbf{K},j}\|\|\Delta_{\mathbf{x},j}\|\right) \\
&\quad \underbrace{+ 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\|\Delta_{\mathbf{x},1}\| + \eta\sum_{j=1}^{k} 2L_{\mathrm{dyn}}B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}\|\Delta_{\mathbf{K},j}\|\|\bar{\Delta}_{\mathbf{x},j}\|}_{\mathrm{Term}_{2,k}}
\end{aligned} \tag{G.13}
$$

Define the terms

$$
\begin{aligned}
C_1 &:= 3B_{\mathrm{stab}}M_{\mathrm{dyn}}, \quad C_2 := 2B_{\mathrm{stab}}L_{\mathrm{dyn}}, \\
\alpha &:= 2B_{\mathrm{stab}}\mathrm{Err}_{\mathbf{u}}\left(M_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{u}} + \sqrt{L_{\mathrm{stab}}}L_{\mathrm{dyn}}\right) \\
\bar{\varepsilon}_1 &:= 2B_{\mathrm{stab}}\left(\|\Delta_{\mathbf{x},1}\| + 2L_{\mathrm{stab}}^{1/2}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\|\right)
\end{aligned}
$$

where above

$$
\mathrm{Err}_{\mathbf{u}} := \max_{k\in[K]}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\Delta_{\mathbf{u},j}\|^2\right)^{1/2}, \quad \mathrm{Err}_{\mathbf{K}} := \max_{k\in[K]}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\Delta_{\mathbf{K},j}\|^2\right)^{1/2}.
$$

We bound the two underlined terms in the above display.

Claim G.10. *Recall* $\mathrm{Err}_{\mathbf{u}} = \max_{k\in[K]}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\|\Delta_{\mathbf{u},j}\|^2\right)^{1/2}$. *Then, for any* k*,*

$$
\mathrm{Term}_{1,k} \le \alpha := 2B_{\mathrm{stab}}\mathrm{Err}_{\mathbf{u}}\left(M_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{u}} + \sqrt{L_{\mathrm{stab}}}L_{\mathrm{dyn}}\right)
$$

Claim G.11. *Assume* $\beta_{\mathrm{stab}} \in [1/2, 1)$ *and recall* $\mathrm{Err}_{\mathbf{K}} := \max_{k\in[K]}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{j}\|\Delta_{\mathbf{K},j}\|^2\right)^{1/2}$. *Then,*

$$
\mathrm{Term}_2 \le \bar{\varepsilon}_1\beta_{\mathrm{stab}}^{\frac{k}{2}}, \quad \bar{\varepsilon}_1 := 2B_{\mathrm{stab}}\left(\|\Delta_{\mathbf{x},1}\| + 2L_{\mathrm{stab}}^{1/2}L_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\|\right)
$$

The previous two claims and (G.13) show that as soon as (G.12) holds for all indices $1 \le j \le k$,

$$
\|\Delta_{\mathbf{x},k+1}\| \le \alpha + \bar{\varepsilon}_1\beta_{\mathrm{stab}}^{k/2} + \eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}\left(C_1\|\Delta_{\mathbf{x},j}\|^2 + C_2\|\Delta_{\mathbf{K},j}\|\|\Delta_{\mathbf{x},j}\|\right)
$$

Set $\varepsilon_j = \|\Delta_{\mathbf{x},j}\|$. Note that $\bar{\varepsilon}_1 \ge \varepsilon_1$, $\beta_{\mathrm{stab}} \in [1/2, 1)$, we can apply Lemma G.19 with $\delta_j \leftarrow \|\Delta_{\mathbf{K},j}\|$ and $M \leftarrow \frac{\eta}{1-\beta} = L_{\mathrm{stab}}$ to find that

$$
\|\Delta_{\mathbf{x},k+1}\| = \varepsilon_{k+1} \le 3(\alpha + \bar{\varepsilon}_1)\beta_{\mathrm{stab}}^{k/3}
$$

provided it holds that (we take$L_{\mathrm{stab}} \geq 1, B_{\mathrm{stab}} \geq 1$)

$$2B_{\mathrm{stab}}\mathrm{Err}_{\mathbf{u}}\left(M_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{u}} + \sqrt{L_{\mathrm{stab}}}L_{\mathrm{dyn}}\right) = \alpha \leq \frac{1}{18C_1 L_{\mathrm{stab}}} = \frac{1}{64B_{\mathrm{stab}}M_{\mathrm{dyn}}L_{\mathrm{stab}}}$$

$$2B_{\mathrm{stab}}\left(\|\Delta_{\mathbf{x},1}\| + 2L_{\mathrm{dyn}}L_{\mathrm{stab}}^{1/2}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\|\right) = \bar{\varepsilon}_1 \leq \frac{1}{108C_1 L_{\mathrm{stab}}} = \frac{1}{324B_{\mathrm{stab}}M_{\mathrm{dyn}}L_{\mathrm{stab}}}$$

$$\mathrm{Err}_{\mathbf{K}} \leq \frac{1}{12\sqrt{L_{\mathrm{stab}}}\max\{C_2, 1\}} \leq \frac{1}{24\sqrt{L_{\mathrm{stab}}}B_{\mathrm{stab}}L_{\mathrm{dyn}}}.$$

For these first two equation, it is enough that

$$\mathrm{Err}_{\mathbf{u}} \leq \min\left\{\frac{\sqrt{L_{\mathrm{stab}}}L_{\mathrm{dyn}}}{M_{\mathrm{dyn}}}, \frac{1}{256B_{\mathrm{stab}}^2 M_{\mathrm{dyn}}L_{\mathrm{dyn}}L_{\mathrm{stab}}^{3/2}}\right\}$$

$$\mathrm{Err}_{\mathbf{K}} \leq \frac{1}{24\sqrt{L_{\mathrm{stab}}}B_{\mathrm{stab}}L_{\mathrm{dyn}}}$$

$$\|\Delta_{\mathbf{x},1}\| \leq \frac{1}{4 \cdot 324B_{\mathrm{stab}}^2 M_{\mathrm{dyn}}L_{\mathrm{stab}}}$$

$$\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| \leq \frac{L_{\mathrm{dyn}}}{8 \cdot 324B_{\mathrm{stab}}^2 M_{\mathrm{dyn}}L_{\mathrm{stab}}^{3/2}}$$

for which $\mathrm{Err}_{\mathbf{u}} \leq C_{\mathbf{u}}$, $\|\Delta_{\mathbf{x},1}\| \leq C_{\boldsymbol{\Delta}}$,$\|\bar{\Delta}_{\mathbf{x},1}\| \leq C_{\hat{\mathbf{x}}}$, $\mathrm{Err}_{\mathbf{K}} \leq C_{\mathbf{K}}$, $\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| \leq C_{\mathbf{K},\hat{\mathbf{x}}}$. Moreover, under the above condition on $\mathrm{Err}_{\mathbf{u}}$, we have

$$\begin{aligned}
\|\Delta_{\mathbf{x},k+1}\| &\leq 3(\alpha + \bar{\varepsilon}_1)\beta_{\mathrm{stab}}^{k/3} \\
&\leq 12B_{\mathrm{stab}}\sqrt{L_{\mathrm{stab}}}L_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{u}} + 2B_{\mathrm{stab}}\left(\|\Delta_{\mathbf{x},1}\| + 2L_{\mathrm{stab}}^{1/2}L_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\|\right)\beta_{\mathrm{stab}}^{k/3} \\
&\leq 12B_{\mathrm{stab}}\sqrt{L_{\mathrm{stab}}}L_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{u}} + 2B_{\mathrm{stab}}\left(\|\Delta_{\mathbf{x},1}\| + 2L_{\mathrm{stab}}^{1/2}L_{\mathrm{dyn}}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\|\right)\beta_{\mathrm{stab}}^{k/3} \\
&\leq c_{\mathbf{u}}\mathrm{Err}_{\mathbf{u}} + \left(c_{\mathbf{K}}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| + c_{\boldsymbol{\Delta}}\|\Delta_{\mathbf{x},1}\|\right)\beta_{\mathrm{stab}}^{k/3}.
\end{aligned}$$

This in turn shows that the inductive hypothesis (G.12) holds, completing the induction.

G.7.1. Deferred Claims

Proof of Claim G.9. We argue in steps. Recall also $\tilde{R}_{\mathrm{stab}}$ be such that $\tilde{R}_{\mathrm{stab}} \geq \max_k\{\|\mathbf{K}_k\|, \|\mathbf{K}'_k\|, 1\}$.

Ensuring within radius of regularity. Our first step is to establish that the maximum of the following three terms is at most R_{dyn}:

$$\begin{aligned}
&\|(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}'_k\tilde{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| \\
&\vee \|(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}'_k\tilde{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| \\
&\vee \|(\mathbf{x}_k, \mathbf{u}_k) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| \leq R_{\mathrm{dyn}}
\end{aligned}$$

First, we observe

$$\begin{aligned}
&\|(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}'_k\tilde{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| \\
&\leq \|(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}'_k\tilde{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| + \|\Delta_{\mathbf{K},k}\|\|\tilde{\Delta}_{\mathbf{x},k}\| \\
&\leq \|(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| + \|\Delta_{\mathbf{K},k}\|\|\Delta_{\mathbf{x},k}\| + \|\Delta_{\mathbf{K},k}\|\|\bar{\Delta}_{\mathbf{x},k}\| \\
&\leq \|(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| + \|\Delta_{\mathbf{K},k}\|\|\Delta_{\mathbf{x},k}\| + \|\Delta_{\mathbf{K},k}\|\|\bar{\Delta}_{\mathbf{x},k}\| + \|\mathbf{K}_k\|\|\Delta_{\mathbf{x},k}\| \\
&\leq \|(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k}) - (\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| + (1 + \|\Delta_{\mathbf{K},k}\|)\|\Delta_{\mathbf{x},k}\| + \|\Delta_{\mathbf{K},k}\|\|\bar{\Delta}_{\mathbf{x},k}\| + \|\mathbf{K}_k\|\|\Delta_{\mathbf{x},k}\| + \|\Delta_{\mathbf{u},k}\| \\
&\leq \|\bar{\Delta}_{\mathbf{x},k}\|(1 + \|\mathbf{K}\|) + (1 + \|\Delta_{\mathbf{K},k}\| + \|\mathbf{K}_k\|)\|\Delta_{\mathbf{x},k}\| + \|\bar{\Delta}_{\mathbf{x},k}\| + \|\Delta_{\mathbf{u},k}\| \\
&\leq (1 + \|\mathbf{K}_k\| + \|\Delta_{\mathbf{K},k}\|)(\|\Delta_{\mathbf{x},k}\| + \|\bar{\Delta}_{\mathbf{x},k}\|) + \|\Delta_{\mathbf{u},k}\| \\
&\leq (2R_{\mathrm{stab}} + \max_j \|\mathbf{K}_k - \mathbf{K}'_j\|)(\|\Delta_{\mathbf{x},k}\| + \|\bar{\Delta}_{\mathbf{x},k}\|) + \|\Delta_{\mathbf{u},k}\|
\end{aligned}$$

Recall the notation

$$\Delta_{\mathbf{K},\infty} := \max_j \|\mathbf{K}_j - \mathbf{K}'_j\|, \quad \Delta_{\mathbf{u},\infty} := \max_j \|\bar{\mathbf{u}}_j - \bar{\mathbf{u}}'_j\|.$$

Hence, it is enough that

$$(2R_{\text{stab}} + \Delta_{\mathbf{K},\infty})(\|\Delta_{\mathbf{x},k}\| + \|\bar{\Delta}_{\mathbf{x},k}\|) + \Delta_{\mathbf{u},\infty} \leq R_{\text{dyn}}.$$

Thus, since $\|\Delta_{\mathbf{x},k}\| \leq c_{\mathbf{u}}\text{Err}_{\mathbf{u}} + (c_{\mathbf{K}}\text{Err}_{\mathbf{K}} - 2B_{\text{stab}})\|\bar{\Delta}_{\mathbf{x},1}\| + c_{\mathbf{\Delta}}\|\Delta_{\mathbf{x},1}\|$ due to (G.12) and $\|\bar{\Delta}_{\mathbf{x},k}\| \leq 2B_{\text{stab}}\|\bar{\Delta}_{\mathbf{x},1}\|$ by Lemma G.4

$$\|\Delta_{\mathbf{x},k}\| \leq c_{\mathbf{u}}\text{Err}_{\mathbf{u}} + c_{\mathbf{K}}\text{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| + c_{\mathbf{\Delta}}\|\Delta_{\mathbf{x},1}\|$$

Hence, it is enough that

$$R_{\text{dyn}} \geq (2R_{\text{stab}} + \Delta_{\mathbf{K},\infty})(c_{\mathbf{u}}\text{Err}_{\mathbf{u}} + c_{\mathbf{K}}\text{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| + c_{\mathbf{\Delta}}\|\Delta_{\mathbf{x},1}\|)) + \Delta_{\mathbf{u},\infty},$$

We can bound $2R_{\text{stab}} + \Delta_{\mathbf{K},\infty} \leq 4\tilde{R}_{\text{stab}}$, and solving the geometric series, bound $\text{Err}_{\mathbf{u}} \leq \sqrt{L_{\text{stab}}}\Delta_{\mathbf{u},\infty}$ and $\text{Err}_{\mathbf{K}} \leq \sqrt{L_{\text{stab}}}\Delta_{\mathbf{K},\infty} \leq 2\sqrt{L_{\text{stab}}}\tilde{R}_{\text{stab}}$. Thus, it is enough that

$$R_{\text{dyn}} \geq (4\tilde{R}_{\text{stab}}c_{\mathbf{u}}\sqrt{L_{\text{stab}}} + 1)\Delta_{\mathbf{u},\infty} + 4\tilde{R}_{\text{stab}}c_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| + 4\tilde{R}_{\text{stab}}c_{\mathbf{\Delta}}\|\Delta_{\mathbf{x},1}\|.$$

which is ensured by Proposition G.5.

Controlling the first remainder. Using that the relevant terms are within the radius of regularity,

$$\begin{aligned}
\|\text{rem}_{k,1}\| &= \|f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}'_k\tilde{\Delta}_{\mathbf{x},k}) - f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k})\| \\
&\leq L_{\text{dyn}}\|(\mathbf{K}'_k - \mathbf{K}_k)\tilde{\Delta}_{\mathbf{x},k}\| \\
&\leq L_{\text{dyn}}\Delta_{\mathbf{K},k}(\|\bar{\Delta}_{\mathbf{x},k}\| + \|\Delta_{\mathbf{x},k}\|).
\end{aligned}$$

Controlling the first remainder. Using the definitions of $\mathbf{x}_k = \bar{\mathbf{x}}_k + \bar{\Delta}_{\mathbf{x},k}$ $\mathbf{u}_t = \bar{\mathbf{u}}_k + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k}$, and the fact that $(\mathbf{x}_k, \mathbf{u}_k)$ is in the radius of regularity around $(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)$, a Taylor expansion implies

$$\begin{aligned}
\|\text{rem}_{k,2}\| &= \left\| \begin{array}{c} f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k}) - f(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k\bar{\Delta}_{\mathbf{x},k}) \\ -\frac{\partial}{\partial x}f_\eta(\mathbf{x}_k, \mathbf{u}_k)\Delta_{\mathbf{x},k} + \frac{\partial}{\partial u}f_\eta(\mathbf{x}_k, \mathbf{u}_k)(\Delta_{\mathbf{u},k} + \mathbf{K}_k(\tilde{\Delta}_{\mathbf{x},k} - \bar{\Delta}_{\mathbf{x},k})) \end{array} \right\| \\
&= \left\| \begin{array}{c} f_\eta(\mathbf{x}_k + \Delta_{\mathbf{x},k}, \bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k}) - f(\mathbf{x}_k, \mathbf{u}_k) \\ -\frac{\partial}{\partial x}f_\eta(\mathbf{x}_k, \mathbf{u}_k)\Delta_{\mathbf{x},k} + \frac{\partial}{\partial u}f_\eta(\mathbf{x}_k, \mathbf{u}_k)(\bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k} - \mathbf{u}_k) \end{array} \right\| \\
&\leq \frac{M_{\text{dyn}}}{2}\left(\|\Delta_{\mathbf{x},k}\|^2 + \|\bar{\mathbf{u}}_k + \Delta_{\mathbf{u},k} + \mathbf{K}_k\tilde{\Delta}_{\mathbf{x},k} - \mathbf{u}_k\|\right)^2 \\
&= \frac{M_{\text{dyn}}}{2}\left(\|\Delta_{\mathbf{x},k}\|^2 + \|\Delta_{\mathbf{u},k} + \mathbf{K}_k(\tilde{\Delta}_{\mathbf{x},k} - \bar{\Delta}_{\mathbf{x},k})\|\right)^2 \\
&= \frac{M_{\text{dyn}}}{2}\left(\|\Delta_{\mathbf{x},k}\|^2 + \|\Delta_{\mathbf{u},k} + \mathbf{K}_k\Delta_{\mathbf{x},k}\|\right)^2 \\
&= \frac{M_{\text{dyn}}}{2}\left((1 + 2\|\mathbf{K}_k\|^2)\|\Delta_{\mathbf{x},k}\|^2 + 2\|\Delta_{\mathbf{u},k}\|^2\right)^2 \\
&= \frac{3}{2}M_{\text{dyn}}R_{\text{stab}}^2\|\Delta_{\mathbf{x},k}\|^2 + M_{\text{dyn}}\|\Delta_{\mathbf{u},k}\|^2
\end{aligned}$$

$\square$

Proof of Claim G.10. Recall $\mathrm{Err}_{\mathbf{u}} = \max_{k\in[K]} \left(\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \|\Delta_{\mathbf{u},j}\|^2\right)^{1/2}$ and $\beta_{\mathrm{stab}} = 1 - \frac{\eta}{L_{\mathrm{stab}}}$. Then,

$$
\begin{aligned}
\mathrm{Term}_{1,k} &= \eta \sum_{j=1}^{k} 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{k-j} \left(L_{\mathrm{dyn}} \|\Delta_{\mathbf{u},j}\| + M_{\mathrm{dyn}} \|\Delta_{\mathbf{u},j}\|^2\right) \\
&\leq 2B_{\mathrm{stab}} \left(M_{\mathrm{dyn}} \mathrm{Err}_{\mathbf{u}}^2 + L_{\mathrm{dyn}} \cdot \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \|\Delta_{\mathbf{u},j}\| \right) \\
&\leq 2B_{\mathrm{stab}} \left(M_{\mathrm{dyn}} \mathrm{Err}_{\mathbf{u}}^2 + L_{\mathrm{dyn}} \cdot (\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j})^{1/2} (\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \|\Delta_{\mathbf{u},j}\|)^{1/2} \right) \\
&\leq 2B_{\mathrm{stab}} \left(M_{\mathrm{dyn}} \mathrm{Err}_{\mathbf{u}}^2 + L_{\mathrm{dyn}} \cdot (\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j})^{1/2} \mathrm{Err}_{\mathbf{u}} \right) \\
&\leq 2B_{\mathrm{stab}} \left(M_{\mathrm{dyn}} \mathrm{Err}_{\mathbf{u}}^2 + L_{\mathrm{dyn}} \cdot (\underbrace{\eta \frac{1}{\beta_{\mathrm{stab}}^{-1} - 1}}_{=L_{\mathrm{stab}}})^{1/2} \mathrm{Err}_{\mathbf{u}} \right) \\
&= 2B_{\mathrm{stab}} \mathrm{Err}_{\mathbf{u}} \left(M_{\mathrm{dyn}} \mathrm{Err}_{\mathbf{u}} + \sqrt{L_{\mathrm{stab}}} L_{\mathrm{dyn}} \right)
\end{aligned}
$$

□

Proof of Claim G.11.

$$
\begin{aligned}
L_{\mathrm{dyn}}^{-1} (\mathrm{Term}_{2,k} - 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{k} \|\Delta_{\mathbf{x},1}\|) &= \eta \sum_{j=1}^{k} 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{k-j} \|\Delta_{\mathbf{K},j}\| \|\bar{\Delta}_{\mathbf{x},j}\| \\
&= \eta \sum_{j=1}^{k} 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{k-j} \beta_{\mathrm{stab}}^{j-1} \|\Delta_{\mathbf{K},j}\| \|\bar{\Delta}_{\mathbf{x},1}\| \\
&= 2B_{\mathrm{stab}} \|\bar{\Delta}_{\mathbf{x},1}\| \cdot \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-1} \|\Delta_{\mathbf{K},j}\| \\
&= 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{\frac{k}{2}-1} \|\bar{\Delta}_{\mathbf{x},1}\| \cdot \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{(k-j)/2} \beta_{\mathrm{stab}}^{j/2} \|\Delta_{\mathbf{K},j}\| \\
&\leq 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{\frac{k}{2}-1} \|\bar{\Delta}_{\mathbf{x},1}\| \leq \left(\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{j} \right) \left(\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{j} \|\Delta_{\mathbf{K},j}\|^2 \right)^{1/2} \\
&\leq 2B_{\mathrm{stab}} \beta_{\mathrm{stab}}^{\frac{k}{2}-1} \|\bar{\Delta}_{\mathbf{x},1}\| \leq \left(\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{j} \right)^{1/2} \left(\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{j} \|\Delta_{\mathbf{K},j}\|^2 \right)^{1/2} \\
&\leq 2B_{\mathrm{stab}} L_{\mathrm{stab}}^{1/2} \beta_{\mathrm{stab}}^{\frac{k}{2}-1} \|\bar{\Delta}_{\mathbf{x},1}\| \cdot \underbrace{\left(\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{j} \|\Delta_{\mathbf{K},j}\|^2 \right)^{1/2}}_{=\mathrm{Err}_{\mathbf{K}}} \\
&\leq 4B_{\mathrm{stab}} L_{\mathrm{stab}}^{1/2} \beta_{\mathrm{stab}}^{\frac{k}{2}} \mathrm{Err}_{\mathbf{K}} \|\bar{\Delta}_{\mathbf{x},1}\|. \qquad (\beta_{\mathrm{stab}} \geq 1/2)
\end{aligned}
$$

Thus,

$$\begin{aligned}\mathrm{Term}_{2,k} &\le 4B_{\mathrm{stab}}L_{\mathrm{dyn}}L_{\mathrm{stab}}^{1/2}\beta_{\mathrm{stab}}^{\frac{k}{2}}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| + 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\|\Delta_{\mathbf{x},1}\| \\ &\le \beta_{\mathrm{stab}}^{\frac{k}{2}}\left(4L_{\mathrm{dyn}}B_{\mathrm{stab}}L_{\mathrm{stab}}^{1/2}\mathrm{Err}_{\mathbf{K}}\|\bar{\Delta}_{\mathbf{x},1}\| + 2B_{\mathrm{stab}}\|\Delta_{\mathbf{x},1}\|\right)\end{aligned}$$

□

G.8. Ricatti synthesis of stabilizing gains.

In this section, we show that under a certain *stabilizability* condition, it is always possible to synthesize primitive controllers satisfying Assumption 3.2 with reasonable constants. We begin by defining our notion of stabilizability; we adopt the formulation based on Jacobian linearizations of non-linear systems the discrete analogue of the senses proposed in which is consistent with (Pfrommer et al., 2023; Westenbroek et al., 2021).

Definition G.11 (Stabilizability)**.** A control trajectory $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K}) \in \mathscr{P}_K$ is $L_{\mathcal{V}}$-Jacobian-Stabilizable if $\max_k \mathcal{V}_k(\boldsymbol{\rho}) \le L_{\mathcal{V}}$, where for $k \in [K+1]$, $\mathcal{V}_k(\boldsymbol{\rho})$ is defined by

$$\begin{aligned}\mathcal{V}_k(\boldsymbol{\rho}) := \sup_{\xi:\|\xi\|\le 1}&\left(\inf_{\tilde{\mathbf{u}}_{1:s}} \|\tilde{\mathbf{x}}_{K+1}\|^2 + \eta\sum_{j=k}^{K}\|\tilde{\mathbf{x}}_j\|^2 + \|\tilde{\mathbf{u}}_j\|^2\right)\\ &\text{s.t. } \tilde{\mathbf{x}}_k = \xi, \quad \tilde{\mathbf{x}}_{j+1} = \tilde{\mathbf{x}}_j + \eta\left(\mathbf{A}_j(\boldsymbol{\rho})\tilde{\mathbf{x}}_j + \mathbf{B}_j(\boldsymbol{\rho})\tilde{\mathbf{u}}_j\right),\end{aligned}$$

Here, for simplicity, we use Euclidean-norm costs, though any Mahalanobis-norm cost induced by a positive definite matrix would suffice. We propose to synthesize gain matrices by performing a standard Ricatti update, normalized appropriately to take account of the step size $\eta > 0$ (see, e.g. Appendix F in (Pfrommer et al., 2023)).

Definition G.12 (Ricatti update)**.** Given a path $\boldsymbol{\rho} \in \mathscr{P}_k$ with $\mathbf{A}_k = \mathbf{A}_k(\boldsymbol{\rho})$, $\mathbf{B}_k = \mathbf{B}_k(\boldsymbol{\rho})$ we define

$$\begin{aligned}&\mathbf{P}_{K+1}^{\mathrm{ric}}(\boldsymbol{\rho}) = \mathbf{I}, \quad \mathbf{P}_{k}^{\mathrm{ric}}(\boldsymbol{\rho}) = (\mathbf{I} + \eta\mathbf{A}_{\mathrm{cl},k}(\boldsymbol{\rho}))^\top \mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})(\mathbf{I} + \eta\mathbf{A}_{\mathrm{cl},k}(\boldsymbol{\rho})) + \eta(\mathbf{I} + \mathbf{K}_k(\boldsymbol{\rho})\mathbf{K}_k(\boldsymbol{\rho})^\top)\\ &\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho}) = (\mathbf{I} + \eta\mathbf{B}_k^\top\mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\mathbf{B}_k)^{-1}(\mathbf{B}_k^\top\mathbf{P}_{k+1}(\boldsymbol{\rho}))(\mathbf{I} + \eta\mathbf{A}_k)\\ &\mathbf{A}_{\mathrm{cl},k}^{\mathrm{ric}}(\boldsymbol{\rho}) = \mathbf{A}_k + \mathbf{B}_k\mathbf{K}_k(\boldsymbol{\rho}).\end{aligned}$$

The main result of this section is that the parameters $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$ in Assumption 3.2 can be bounded in terms of L_{dyn} in Assumption 3.1, and the bound $L_{\mathcal{V}}$ defined above.

Proposition G.12 (Instantiating the Lyapunov Lemma)**.** *Let* $L_{\mathrm{dyn}}, L_{\mathcal{V}} \ge 1$, *and let* $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K})$ *be* $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$*-regular and* $L_{\mathcal{V}}$*-Jacobian Stabilizable. Suppose further that* $\eta \le 1/5L_{\mathrm{dyn}}^2 L_{\mathcal{V}}$. *Then,* $(\boldsymbol{\rho}, \mathbf{K}_{1:K}^{\mathrm{ric}})$*-is* $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$*-Jacobian Stable, where*

$$R_{\mathrm{stab}} = \frac{4}{3}L_{\mathcal{V}}L_{\mathrm{dyn}}, \quad B_{\mathrm{stab}} = \sqrt{5}L_{\mathrm{dyn}}L_{\mathcal{V}}, \quad L_{\mathrm{stab}} = 2L_{\mathcal{V}}$$

Proposition G.12 is proven in Appendix G.8.1 below. A consequence of the above proposition is that, given access to a smooth local model of dynamics, one can implement the synthesis oracle by computing linearizations around demonstrated trajectories, and solving the corresponding Ricatti equations as per the above discussions to synthesize the correct gains.

G.8.1. Proof of Proposition G.12 (Ricatti synthesis of gains)

Throughout, we use the shorthand $\mathbf{A}_k = \mathbf{A}_k(\boldsymbol{\rho})$ and $\mathbf{B}_k = \mathbf{B}_k(\boldsymbol{\rho})$, recall that $\|\cdot\|$ denotes the operator norm when applied to matrices. We also recall our assumptions that $L_{\mathrm{dyn}}, L_{\mathcal{V}} \ge 1$. We begin by translating our stabilizability assumption (Definition G.11) into the the $\mathbf{P}$-matrices in Definition G.12. The following statement recalls Lemma F.1 in (Pfrommer et al., 2023), an instantiation of well-known solutions to linear-quadratic dynamic programming (e.g. (Anderson & Moore, 2007)).

Lemma G.13 (Equivalence of stabilizability and Ricatti matrices)**.** *Consider a trajectory* $(\mathbf{x}_{1:K}, \mathbf{u}_{1:K})$, *and define the parameter* $\boldsymbol{\Theta} := (\mathbf{A}_{\mathrm{jac}}(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k), \mathbf{B}_{\mathrm{jac}}(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k))_{k\in[K]}$. *Then, for all* $k \in [K]$,

$$\forall k \in [K], \quad \mathcal{V}_k(\boldsymbol{\rho}) = \|\mathbf{P}_k(\boldsymbol{\Theta})\|_{\mathrm{op}}$$

Hence, if $\boldsymbol{\rho}$ is $L_{\mathcal{V}}$-stabilizable,

$$\max_{k\in[K+1]} \|\mathbf{P}_k(\boldsymbol{\Theta})\|_{\mathrm{op}} \le L_{\mathcal{V}}.$$

Lemma G.14 (Lyapunov Lemma, Lemma F.10 in (Pfrommer et al., 2023))**.** *Let $\mathbf{X}_{1:K}, \mathbf{Y}_{1:K}$ be matrices of appropriate dimension, and let $Q \succeq \mathbf{I}$ and $\mathbf{Y}_k \succeq 0$. Define $\boldsymbol{\Lambda}_{1:K+1}$ as the solution of the recursion*

$$\boldsymbol{\Lambda}_{K+1} = \mathbf{Q}, \quad \boldsymbol{\Lambda}_k = \mathbf{X}_k^\top \boldsymbol{\Lambda}_{k+1}\mathbf{X}_k + \eta\mathbf{Q} + \mathbf{Y}_k$$

Define the operator $\boldsymbol{\Phi}_{j+1,k} = \mathbf{X}_j \cdot \mathbf{X}_{j-1}, \cdots \mathbf{X}_k$, with the convention $\boldsymbol{\Phi}_{k,k} = \mathbf{I}$. Then, if $\max_k \|\mathbf{I} - \mathbf{X}_k\|_{\mathrm{op}} \le \kappa\eta$ for some $\kappa \le 1/2\eta$,

$$\|\boldsymbol{\Phi}_{j,k}\|^2 \le \max\{1, 2\kappa\} \max_{k\in[K+1]} \|\boldsymbol{\Lambda}_k\|(1-\eta\alpha)^{j-k}, \quad \alpha := \frac{1}{\max_{k\in[K+1]} \|\boldsymbol{\Lambda}_{1:K+1}\|}.$$

Claim G.15. *If $\boldsymbol{\rho}$ is $(0, L_{\mathrm{dyn}}, \infty)$-regular, then for all k, $\mathbf{A}_k = \mathbf{A}_k(\boldsymbol{\rho})$ and $\mathbf{B}_k = \mathbf{B}_k(\boldsymbol{\rho})$ satisfy* $\max_{k\in[K]} \max\{\|\mathbf{A}_k\|, \|\mathbf{B}_k\|\} \le L_{\mathrm{dyn}}$.

Proof. For any $k \in [K]$,

$$\max\{\|\mathbf{A}_k\|, \|\mathbf{B}_k\|\} = \max\left\{\left\|\frac{\partial}{\partial x} f(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\right\|, \left\|\frac{\partial}{\partial u} f(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\right\|\right\} \le \|\nabla f(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| \le L_{\mathrm{dyn}},$$

where the last inequality follows by regularity. □

Claim G.16. *Recall $\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho}) = (\mathbf{I} + \eta\mathbf{B}_k^\top \mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\mathbf{B}_k)^{-1}(\mathbf{B}_k^\top \mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho}))(\mathbf{I} + \eta\mathbf{A}_k)$. Then, if $\boldsymbol{\rho}$ is $L_{\mathcal{V}}$-stabilizable and $(0, L_{\mathrm{dyn}}, \infty)$-regular, and if $\eta \le 1/3L_{\mathrm{dyn}}$,*

$$\|\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho})\| \le \frac{4}{3} L_{\mathcal{V}} L_{\mathrm{dyn}}$$

Proof. We bound

$$\begin{aligned}
\|\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho})\| &\le \|\mathbf{B}_k\|\|\mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\|(1 + \eta\|\mathbf{A}_k\|) \\
&\le L_{\mathrm{dyn}}(1 + \eta L_{\mathrm{dyn}})\|\mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\| && \text{(Claim G.15)} \\
&\le L_{\mathcal{V}} L_{\mathrm{dyn}}(1 + \eta L_{\mathrm{dyn}}) && \text{(Lemma G.13, } L_{\mathcal{V}} \ge 1) \\
&\le \frac{4}{3} L_{\mathcal{V}} L_{\mathrm{dyn}} && (\eta \le 1/3L_{\mathrm{dyn}}.)
\end{aligned}$$

□

Proof of Proposition G.12. We want to show that $\mathbf{K}_{1:K}^{\mathrm{ric}}(\boldsymbol{\rho})$ is $(R_{\mathrm{stab}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stabilizing.Claim G.16 has already established that $\max_{k\in[K]} \|\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho})\| \le R_{\mathrm{stab}} = \frac{4}{3} L_{\mathcal{V}} L_{\mathrm{dyn}}$.

To prove the other conditions, we apply Lemma G.14 with $\mathbf{Y}_k = \mathbf{K}_k(\boldsymbol{\Theta})\mathbf{K}_k(\boldsymbol{\Theta})$, $\mathbf{Q} = \mathbf{I}$, and $\mathbf{X}_k = \mathbf{I} + \eta\mathbf{A}_{\mathrm{cl},k}(\boldsymbol{\Theta})$. From Definition G.12, let have that the term $\boldsymbol{\Lambda}_k$ in Lemma G.14 is precise equal to $\mathbf{P}_k(\boldsymbol{\Theta})$. From Lemma G.13,

$$\max_{k\in[K+1]} \|\mathbf{P}_k(\boldsymbol{\Theta})\|_{\mathrm{op}} = \max_{k\in[K+1]} \mathcal{V}_k(\boldsymbol{\rho}) \le L_{\mathcal{V}}.$$

This implies that if $\max_k \|\mathbf{X}_k - \mathbf{I}\| \le \kappa\eta \le 1/2$, we have

$$\|\boldsymbol{\Phi}_{\mathrm{cl},j,k}(\boldsymbol{\Theta})\|^2 = \|(\mathbf{X}_j \cdot \mathbf{X}_{j-1} \cdot \ldots \mathbf{X}_k)\| \le \max\{1, 2\kappa\} L_{\mathcal{V}} \left(1 - \frac{\eta}{L_{\mathcal{V}}}\right)^{j-k}.$$

It suffices to find an appropriate upper bound κ. We have

$$\begin{aligned}
\|\mathbf{X}_k - \mathbf{I}\| = \|\eta \mathbf{A}_{\mathrm{cl},k}(\mathbf{\Theta})\| &\leq \eta(\|\mathbf{A}_k\| + \|\mathbf{B}_k\|\|\mathbf{K}_k(\mathbf{\Theta})\|) \\
&\leq \eta L_{\mathrm{dyn}}(1 + \|\mathbf{K}_k(\mathbf{\Theta})\|) \\
&\leq \eta L_{\mathrm{dyn}}(1 + \frac{4}{3} L_{\mathrm{dyn}} L_{\mathcal{V}}) && \text{(Claim G.16)} \\
&\leq \frac{7}{3} \eta L_{\mathrm{dyn}}^2 L_{\mathcal{V}} && (L_{\mathcal{V}}, L_{\mathrm{dyn}} \geq 1)
\end{aligned}$$

Setting $\kappa = \frac{7}{3} L_{\mathrm{dyn}}^2 L_{\mathcal{V}}$., we have that as $\eta \leq \frac{1}{5L_{\mathrm{dyn}}^2 L_{\mathcal{V}}} \leq \min\{\frac{3}{14L_{\mathrm{dyn}}^2 L_{\mathcal{V}}}, \frac{1}{3L_{\mathrm{dyn}}}\}$ (recall $L_{\mathrm{dyn}}, L_{\mathcal{V}} \geq 1$), we can bound

$$\max\{1, 2\kappa\} \leq \max\left\{1, \frac{14}{3} L_{\mathrm{dyn}}^2 L_{\mathcal{V}}\right\} \leq \max\left\{1, 5L_{\mathrm{dyn}}^2 L_{\mathcal{V}}\right\} = 5L_{\mathrm{dyn}}^2 L_{\mathcal{V}}^2,$$

where again recall $L_{\mathcal{V}}, L_{\mathrm{dyn}} \geq 1$. In sum, for $\eta \leq \frac{1}{5L_{\mathrm{dyn}}^2 L_{\mathcal{V}}}$, we have

$$\|\mathbf{\Phi}_{\mathrm{cl},j,k}\|^2 \leq 5L_{\mathrm{dyn}}^2 L_{\mathcal{V}}^2 \left(1 - \frac{\eta}{L_{\mathcal{V}}}\right)^{j-k}.$$

Hence, using the elementary inequality $\sqrt{1-a} \leq (1 - a/2)$,

$$\|\mathbf{\Phi}_{\mathrm{cl},j,k}\| \leq \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{V}} \left(1 - \frac{\eta}{L_{\mathcal{V}}}\right)^{(j-k)/2} \leq \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{V}} \left(1 - \frac{\eta}{2L_{\mathcal{V}}}\right)^{j-k},$$

which shows that we can select $B_{\mathrm{stab}} = \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{V}}$ and $L_{\mathrm{stab}} = 2L_{\mathcal{V}}$. □

G.9. Solutions to recursions

This section contains the solutions to various recursions used in the proof of the two two results in Appendix G.4: Proposition G.5 (whose proof is given in Appendix G.7) and Lemma G.4 (whose proof is given in Appendix G.6).

Lemma G.17 (First Key Recursion)**.** *Let $C_1 > 0, C_2 \geq 1/2$, $\beta_{\mathrm{stab}} \in (0,1)$, and suppose $\varepsilon_1, \varepsilon_2, \ldots$ is a sequence satisfying $\varepsilon_1 \leq \bar{\varepsilon}_1$, and*

$$\varepsilon_{k+1} \leq C_2 \beta_{\mathrm{stab}}^k \bar{\varepsilon}_1 + C_1 \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \varepsilon_j^2$$

Then, as long as $C_1 \leq \beta(1-\beta)/2\eta$, it holds that $\varepsilon_k \leq 2C_2 \beta_{\mathrm{stab}}^{k-1} \bar{\varepsilon}_1$ for all k.

Proof. Consider the sequence $\nu_k = 2C_2 \beta_{\mathrm{stab}}^{k-1} \bar{\varepsilon}_1$. Since $C_2 \geq 1/2$, we have $\nu_1 \geq \bar{\varepsilon}_1 \geq \varepsilon_1$. Moreover,

$$\begin{aligned}
C_2 \beta_{\mathrm{stab}}^k \bar{\varepsilon}_1 + C_1 \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \nu_j &= C_2 \beta_{\mathrm{stab}}^k \bar{\varepsilon}_1 + 2C_1 C_2 \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k+j-2} \bar{\varepsilon}_1 \\
&= C_2 \beta_{\mathrm{stab}}^k \bar{\varepsilon}_1 \left(1 + \frac{2C_1}{\beta} \sum_{j=0}^{k-1} \beta_{\mathrm{stab}}^j\right) \\
&\leq C_2 \beta_{\mathrm{stab}}^k \bar{\varepsilon}_1 \left(1 + \frac{2C_1 \eta}{\beta(1-\beta)}\right)
\end{aligned}$$

Hence, for $C_1 \leq \beta(1-\beta)/2\eta$, we have $C_2 \beta_{\mathrm{stab}}^k \bar{\varepsilon}_1 + C_1 \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \nu_j \leq 2C_2 \bar{\varepsilon}_1 \beta_{\mathrm{stab}}^k \leq \nu_{k+1}$. This shows that the (ν_k) sequence dominates the (ε_k) sequence, as needed. □

Lemma G.18 (Second Key Recursion)**.** *Let $c, \Delta, \eta > 0$, $\beta_{\text{stab}} \in (0,1)$ and let $\varepsilon_1, \varepsilon_2, \ldots$ satisfy $\varepsilon_1 \le c$ and*

$$\varepsilon_{k+1} \le c\beta_{\text{stab}}^k + c\eta\Delta\beta_{\text{stab}}^{k-1}\sum_{j=1}^k \varepsilon_j.$$

Then, if $\Delta \le \frac{\beta(1-\beta)}{2c\eta}$, $\varepsilon_{k+1} \le 2c\beta_{\text{stab}}^k$ for all k.

Proof. Consider the sequence $\nu_k = 2c\beta_{\text{stab}}^{k-1}$. Since $\varepsilon_1 \le c$, $\nu_1 \ge \varepsilon_1$. Moreover,

$$\begin{aligned}
c\beta_{\text{stab}}^k + c\eta\Delta\beta_{\text{stab}}^{k-1}\sum_{j=1}^k \nu_j &\le c\beta_{\text{stab}}^k + 2c^2\eta\Delta\beta_{\text{stab}}^{k-1}\sum_{j=1}^k \beta_{\text{stab}}^{j-1} \\
&\le c\beta_{\text{stab}}^k + 2c^2\eta\Delta\beta_{\text{stab}}^{k-1}\frac{1}{1-\beta} \\
&\le c\beta_{\text{stab}}^k\left(1 + 2c\Delta\frac{\eta}{\beta(1-\beta)}\right).
\end{aligned}$$

Hence, for $\Delta \le \frac{\beta(1-\beta)}{2c\eta}$, the above is at most $2c\beta_{\text{stab}}^k \le \nu_{k+1}$. This shows that the (ν_k) sequence dominates the (ε_k) sequence, as needed. ☐

Lemma G.19 (Third Key Recursion)**.** *Let $C_1, C_2 > 0$, $\alpha \ge 0$, $\beta_{\text{stab}} \in (1/2, 1)$, and let $\varepsilon_1, \varepsilon_2, \ldots,$ and $\delta_1, \delta_2, \ldots,$, and $\bar{\varepsilon}_1 \ge \varepsilon_1$ and be a sequence of real numbers satisfying*

$$\varepsilon_{k+1} \le \alpha + \eta\sum_{j=1}^k \beta_{\text{stab}}^{k-j}(C_1\varepsilon_j^2 + C_2\varepsilon_j\delta_j) + \beta_{\text{stab}}^{k/3}\bar{\varepsilon}_1$$

Defin, $\mathrm{Err}_\delta := \max_k \eta\sum_{j=1}^k \beta_{\text{stab}}^{(k-j)}\delta_j^2$ and $M = \eta/(1-\beta)$. Then, as long as

$$\alpha \le \frac{1}{18C_1M}, \quad \bar{\varepsilon}_1 \le \frac{1}{108C_1M}, \quad \mathrm{Err}_\delta \le \frac{1}{12\sqrt{M}\max\{C_2, 1\}},$$

the following holds for all $k \ge 0$:

$$\varepsilon_{k+1} \le 3\alpha + 3\bar{\varepsilon}_1\beta_{\text{stab}}^{k/3}.$$

Proof of Lemma G.19. Consider a sequence

$$\nu_{k+1} = \alpha_\star + c_\star\beta_\star^k\bar{\varepsilon}_1, \quad \alpha_\star = 3\alpha, c_\star = 3, \beta_\star = \beta_{\text{stab}}^{1/3}$$

defined for $k \ge 0$, for some $\alpha_\star \ge \alpha$, $\beta_\star \in (\beta, 1)$, and $c_\star \ge 1$. Then, $\nu_1 \ge \bar{\varepsilon}_1, \ge \varepsilon_1$. Let us define the term B_k via

$$B_k = \alpha + \eta\sum_{j=1}^k \beta_{\text{stab}}^{k-j}(C_1\nu_j^2 + C_2\nu_j\delta_j) + \beta_{\text{stab}}^{k/3}\bar{\varepsilon}_1.$$

It suffices to show $B_k \le \nu_{k+1}$ for all k. Introduce $\mathrm{Term}_{\nu,k} = \left(\eta\sum_{j=1}^k \beta_{\text{stab}}^{k-j}\nu_j^2\right)^{1/2}$ and $\mathrm{Err}_\delta = \max_k\left(\eta\sum_{j=1}^k \beta_{\text{stab}}^{k-j}\delta_j^2\right)^{1/2}$ Then, by Cauch-Schwartz,

$$\begin{aligned}
B_k &= \alpha + \eta\sum_{j=1}^k \beta_{\text{stab}}^{k-j}(C_1\nu_j^2 + C_2\nu_j\delta_j) + \beta_{\text{stab}}^{k/3}\bar{\varepsilon}_1 \\
&\le \alpha + C_1\mathrm{Term}_{\nu,k}^2 + C_2\mathrm{Term}_{\nu,k}\mathrm{Err}_\delta + \beta_{\text{stab}}^{k/3}\bar{\varepsilon}_1.
\end{aligned}$$

We now bound

$$\begin{aligned}
\mathrm{Term}^2_{\nu,k} &= \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \nu_j^2 \\
&= \eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} (\alpha_\star + c_\star \bar{\varepsilon}_1 \beta_\star^{j-1})^2 \\
&\le 2\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \alpha_\star^2 + 2\eta c_\star^2 \bar{\varepsilon}_1^2 \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \beta_\star^{2(j-1)} \\
&\le \frac{2\eta\alpha_\star^2}{1-\beta} + 2\eta c_\star^2 \bar{\varepsilon}_1^2 \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \beta_\star^{2(j-1)}.
\end{aligned}$$

Now, recalling $\beta_\star = \beta_{\mathrm{stab}}^{1/3}$, we have

$$\begin{aligned}
\sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \beta_\star^{j-1} = \sum_{j=1}^{k} \beta_\star^{3k-3j} \beta_\star^{2(j-1)} &= \sum_{j=1}^{k} \beta_\star^{3k-j-2} \\
&= \beta_\star^{2k-2} \sum_{j=1}^{k} \beta_\star^{k-j} = \beta_\star^{2k-2} \sum_{j\ge 0} \beta_\star^{j} \\
&\le 3\beta_\star^{2k-2} \sum_{j\ge 0} \beta_\star^{3j} = 3\beta_\star^{2k} \beta_{\mathrm{stab}}^{-2/3} \sum_{j\ge 0} \beta_{\mathrm{stab}}^{j} \\
&= \frac{3}{1-\beta} \beta_\star^{2k} \beta_{\mathrm{stab}}^{-2/3} \le \frac{3}{\beta(1-\beta)} \beta_\star^{2k}.
\end{aligned}$$

Thus, adopting the shorthand $M = \eta/(1-\beta)$, and using the assumption $\beta_{\mathrm{stab}} \ge 1/2$,

$$\mathrm{Term}^2_{\nu,k} \le 2\alpha_\star^2 M + 12 M c_\star^2 \bar{\varepsilon}_1^2 \beta_\star^{2k}.$$

Thus,

$$\begin{aligned}
B_k &\le \alpha + C_1 \mathrm{Term}^2_{\nu,k} + C_2 \mathrm{Term}_{\nu,k} \mathrm{Err}_\delta + \beta_{\mathrm{stab}}^{k} \bar{\varepsilon}_1 \\
&\le \alpha + 2C_1 \alpha_\star^2 M + 12 C_1 M c_\star^2 \bar{\varepsilon}_1^2 \beta_\star^{2k} + \mathrm{Err}_\delta C_2 \sqrt{2M} \alpha_\star + \mathrm{Err}_\delta C_2 \sqrt{12M} c_\star \bar{\varepsilon}_1 \beta_\star^k + \beta_{\mathrm{stab}}^{k/3} \bar{\varepsilon}_1 \\
&= \alpha \left(1 + 2C_1 \frac{\alpha_\star^2}{\alpha} M + \frac{\alpha_\star}{\alpha} \mathrm{Err}_\delta C_2 \sqrt{2M}\right) + \beta_\star^k \bar{\varepsilon}_1 \left(12 C_1 M c_\star^2 \beta_\star^k \bar{\varepsilon}_1 + E_\delta \sqrt{12M} c_\star + \beta_{\mathrm{stab}}^{k/3} \beta_\star^{-k}\right) \\
&\le \alpha \left(1 + 2C_1 \frac{\alpha_\star^2}{\alpha} M + \frac{\alpha_\star}{\alpha} \mathrm{Err}_\delta C_2 \sqrt{2M}\right) + \beta_\star^k \bar{\varepsilon}_1 \left(12 C_1 M c_\star^2 \bar{\varepsilon}_1 + E_\delta \sqrt{12M} c_\star\right)
\end{aligned}$$

where in the last line, we use $\beta_\star = \beta_{\mathrm{stab}}^{1/3} \le 1$. Recalling $\alpha_\star = 3\alpha$ and $c = 3$, we have $B_k \le \alpha_\star + c_\star \bar{\varepsilon}_1 \beta_\star^k = \nu_{k+1}$ as soon as

$$\begin{aligned}
1 &\ge 2C_1 \frac{\alpha_\star^2}{\alpha} M \vee \frac{\alpha_\star}{\alpha} \mathrm{Err}_\delta C_2 \sqrt{2M} \vee 12 C_1 M c_\star^2 \bar{\varepsilon}_1 + E_\delta \sqrt{12M} c_\star \\
&= 18\alpha C_1 M \vee 3\mathrm{Err}_\delta C_2 \sqrt{2M} \vee 108 C_1 M \bar{\varepsilon}_2 + 3E_\delta \sqrt{12M} \\
&= 18\alpha C_1 M \vee 108 C_1 M \bar{\varepsilon}_1 \vee \mathrm{Err}_\delta (3C_2 \sqrt{2M} \vee 3\sqrt{12M}).
\end{aligned}$$

Thus, it suffices that

$$\alpha \le \frac{1}{18 C_1 M}, \quad \bar{\varepsilon}_1 \le \frac{1}{108 C_1 M}, \quad \mathrm{Err}_\delta \le \frac{1}{12\sqrt{M} \max\{C_2, 1\}},$$

as needed.

$\square$

Lemma G.20 (Matrix Product Perturbation)**.** *Define matrix products*

$$\mathbf{\Phi}_{k,j} = \mathbf{X}_{k-1} \cdot \mathbf{X}_{k-2} \cdots \mathbf{X}_j, \quad \mathbf{\Phi}'_{k,j} = \mathbf{X}'_{k-1} \cdot \mathbf{X}'_{k-2} \cdots \mathbf{X}'_j.$$

Let $\eta, \Delta, c > 0$ and $\beta_{\mathrm{stab}} \in (0, 1)$. If (a) $\mathbf{\Phi}_{k,j} \leq \beta_{\mathrm{stab}}^{k-j}$ for all $j \leq k$, (b) $\|\mathbf{X}_j - \mathbf{X}'_j\| \leq \eta\Delta\beta_{\mathrm{stab}}^{j-1}$ for all $j \geq 1$ and (c) $\Delta \leq \frac{\beta(1-\beta)}{2c\eta}$, then, for all $j \leq k$, $\|\mathbf{\Phi}'_{k,j}\| \leq 2c\beta_{\mathrm{stab}}^{k-j}$.

Proof. Without loss of generally, take $j = 1$. Then, letting $\mathbf{\Delta}_k = (\mathbf{X}'_k - \mathbf{X}_k)$,

$$\begin{aligned}
\mathbf{\Phi}'_{k+1,1} &= \mathbf{X}'_k \cdot \mathbf{X}'_{k-2} \cdots \mathbf{X}'_1 \\
&= \mathbf{X}'_k \cdot \mathbf{\Phi}'_{k,1} \\
&= \mathbf{\Delta}_k \mathbf{\Phi}'_{k,1} + \mathbf{X}_k \mathbf{\Phi}'_{k,1} \\
&= \mathbf{\Delta}_k \mathbf{\Phi}'_{k,1} + \mathbf{X}_k \mathbf{\Delta}_{k-1} \mathbf{\Phi}'_{k-2,1} + \mathbf{X}_k \mathbf{X}_{k-1} \mathbf{\Phi}'_{k-2,1} \\
&= \mathbf{\Phi}_{k+1,k+1} \mathbf{\Delta}_k \mathbf{\Phi}'_{k,1} + \mathbf{\Phi}_{k+1,k} \mathbf{\Delta}_{k-1} \mathbf{\Phi}'_{k-2,1} + \mathbf{\Phi}_{k+1,k} \mathbf{\Phi}'_{k-2,1} \\
&= \sum_{j=1}^{k} \mathbf{\Phi}_{k+1,j+1} \mathbf{\Delta}_j \mathbf{\Phi}'_{j,1} + \mathbf{\Phi}_{k+1,1}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\|\mathbf{\Phi}'_{k+1,1}\|_{\mathrm{op}} &\leq c\eta \sum_{j=1}^{k} \beta_{\mathrm{stab}}^{k-j} \|\mathbf{X}_j - \mathbf{X}'_j\| \|\mathbf{\Phi}'_{j,1}\| + c\beta_{\mathrm{stab}}^{k} \\
&\leq c\eta\beta_{\mathrm{stab}}^{k-1} \Delta \sum_{j=1}^{k} \|\mathbf{\Phi}'_{j,1}\| + c\beta_{\mathrm{stab}}^{k}. && (\|\mathbf{X}_j - \mathbf{X}'_j\| \leq \eta\Delta\beta_{\mathrm{stab}}^{j-1})
\end{aligned}$$

Define $\varepsilon_j = \|\mathbf{\Phi}'_{j,1}\|$. Then, $\varepsilon_1 = 1 \leq c$, so Lemma G.18 implies that for $\Delta \leq \frac{(1-\beta)\beta}{2\eta}$, $\|\mathbf{\Phi}'_{k,1}\| := \varepsilon_k \leq 2c\beta_{\mathrm{stab}}^{k}$ for all k. □

H. Sampling and Score Matching

In this section, we provide a rigorous guarantee on the quality of sampling from the learned DDPM under Assumption 3.3. We organize the section as follows:

- In Definition H.1 we provide the main notion of function class complexity, a vectorized Rademacher complexity that also appears in some form in Block et al. (2020a); Maurer (2016).
- We then state the main result of the section, Theorem 6, which provides a high probability upper bound on the number of samples n required in order to sample from `DDPM` trained on a given score estimate such that the sample is close in our optimal transport metric to the target distribution.
- In particular, in (H.1), we give the exact polynomial dependence of the sampling parameters α and J on the parameters of the problem.
- We break the proof of Theorem 6 into two sections. First, in Appendix H.1, we recall a result of Chen et al. (2022); Lee et al. (2023) that shows that it suffices to accurately learn the score in the sense that if the score estimate is accurate in the appropriate sense, then the `DDPM` will adequately sample from a distribution close to the target.
- In Remark H.5, we emphasize the conditions that would be required to sample in total variation and explain why they do not hold in our setting.
- Then, in Appendix H.2, we apply statistical learning techniques, similar to those in Block et al. (2020a), to show that with sufficiently many samples, we can effectively learn the score. We detail in Remark H.7 how the realizability part of Assumption 3.3 can be relaxed.

- Finally, we conclude the proof of Theorem 6 by combining the two intermediate results detailed above.

To begin, we define our notion of statistical complexity:

Definition H.1 (Complexity of Θ Complexity)**.** Define the vector- Rademacher complexity of a function class $\{\mathbf{s}_\theta | \theta \in \Theta_j\}$ by:

$$\mathcal{R}_n(\Theta_j) = \mathbb{E}\left[\sup_{\theta\in\Theta_j} \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{d} \varepsilon_{k,i}\mathbf{s}_\theta^{(i)}(\mathsf{a}_k, \boldsymbol{\rho}_{\mathrm{m},k}, j)\right],$$

where $\mathbf{s}_\theta^{(i)}$ denotes the i^{th} coordinate of $\mathbf{s}_\theta$ and the expectation is over $(\kappa, \boldsymbol{\rho}_{\mathrm{m},h}) \sim q_{[t]}$ and independent Rademacher random variables $\varepsilon_{k,i}$, with $q_{[t]}$ as in Section 2.

We now state the main result of this section.

Theorem 6. *Fix $1 \le h \le H$ and suppose that $(\mathsf{a}_i, \boldsymbol{\rho}_{\mathrm{m},h,i}) \sim q$ are independent for $1 \le i \le n$ Suppose that the projection of q onto the first coordinate has support (as defined in Definition C.3) contained in the euclidean ball of radius $R \ge 1$ in $\mathbb{R}^d$. For $\varepsilon > 0$, set*

$$J = c\frac{d^3R^4(R+\sqrt{d})^4\log\left(\frac{dR}{\varepsilon}\right)}{\varepsilon^{20}}, \qquad \alpha = c\frac{\varepsilon^8}{d^2R^2(R+\sqrt{d})^2}. \tag{H.1}$$

for some universal constant $c > 0$. Suppose that for all $1 \le j \le J$, the following hold:

- *There exists a function class Θ_j containing some θ_j^* such that $\mathbf{s}_\star(\cdot,\cdot,j\alpha) = \mathbf{s}_{\theta_j^*}(\cdot,\cdot,j\alpha) = \nabla\log q_{[j\alpha]}(\cdot|\cdot)$, where $q_{[\cdot]}$ is defined in Section 2.*

- *The following holds for some $\delta > 0$:*

$$\sup_{\substack{\theta,\theta'\in\Theta_j \\ ||\mathsf{a}||\vee||\mathsf{a}'||\le R+\sqrt{d\log\left(\frac{2nd}{\delta}\right)} \\ \boldsymbol{\rho}_{\mathrm{m},h}}} \left|\left|\mathbf{s}_\theta(\mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \mathbf{s}_{\theta'}(\mathsf{a}', \boldsymbol{\rho}_{\mathrm{m},h}, t)\right|\right| \le c\frac{d^2(R+\sqrt{d\log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8}.$$

- *Assumption 3.3 holds and thus, for all $j \in [J]$, it holds that $\mathcal{R}_n(\Theta_j) \le C_\Theta n^{-1/\nu}$ for some $\nu \ge 2$ and all $n \in \mathbb{N}$.*

- *The parameter $\widehat{\theta} = \widehat{\theta}_{1:J}$ is defined to be the empirical minimizer of $\mathcal{L}_{\mathrm{DDPM}}$ from Section 3.*

If

$$n \ge c\left(\frac{C_\Theta dR(R\vee\sqrt{d})\log(dn)}{\varepsilon^4}\right)^{4\nu} \vee \left(\frac{d^6(R^4\vee d^2\log^3\left(\frac{ndR}{\delta\varepsilon}\right))}{\varepsilon^{24}}d^2\right)^{4\nu},$$

then with probability at least $1-\delta$, it holds that

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h}\sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}}\left[\inf_{\mu\in\mathscr{C}(\mathtt{DDPM}(\mathbf{s}_{\widehat{\theta}},\boldsymbol{\rho}_{\mathrm{m},h}),q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h}))} \mathbb{P}_{(\widehat{\mathsf{a}},\mathsf{a}^*)\sim\mu}\left(||\widehat{\mathsf{a}}-\mathsf{a}^*|| \ge \varepsilon\right)\right] \le 3\varepsilon.$$

Remark H.1. We emphasize that the exact value of the polynomial dependence (and in particular its pessimism) stem from the guarantees of Chen et al. (2022); Lee et al. (2023) regarding the quality of sampling with DDPMs. We remark below that the learning process itself does not incur such poor polynomial dependence except via these guarantees. Furthermore, we do not expect the sampling guarantees of those two works to be tight in any sense and such a poor polynomial dependence is not observed in practice. Rather, we include the bounds of Chen et al. (2022); Lee et al. (2023) so as to provide a fully rigorous end-to-end guarantee showing that polynomially many samples suffice to do imitation learning under our assumptions.

Remark H.2. A subtle difference between the presentation in the body and that here is the dependence of the complexity of Θ on the parameter α. We phrase the complexity guarantee as we did in the body in order to emphasize the dependence on the algorithmic parameter. If we let C'_Θ denote a constant such that $\mathcal{R}_n(\Theta) \le C'_\Theta (\alpha/n)^{-1/\nu}$, then the sample complexity above becomes

$$n \ge c \left(\frac{C'_\Theta \log(dn)}{\alpha} \right)^{4\nu} \vee \left(\frac{d^2 (R^2 \vee d^2 \log^3 \left(\frac{ndR}{\varepsilon\delta} \right))}{\alpha^2 \varepsilon^{16}} \right)^{4\nu}.$$

Remark H.3. We observe that while at first it may seem that the upper bound on the osculation of $\mathbf{s}_\theta$ is limiting, and, indeed, it is not obvious that this assumption does not contradict the realizability assumption immediately preceding it, it follows immediately from Lemma H.2 that if the preceding assumptions are satisfied, then the true score function $\mathbf{s}_\star$ automatically satisfies the bound on osculation. Moreover, the boundedness of the function class is only assumed for the sake of convenience and could be substantially relaxed to an assumption requiring finiteness of moments of the envelope of the class (Wainwright, 2019; Rakhlin et al., 2017). For the sake of simplicity, we do not further remark on this.

Critically, the guarantee of the quality of our DDPM is not in TV, but rather an optimal transport distance tailored to the problem at hand. As remarked in Section 3, it is precisely this weaker guarantee that makes the problem challenging.

We begin by recalling the basic motivation for Denoising Diffusion Probabilistic Models (DDPMs) and explain how we train them. We then apply results from Chen et al. (2022) to show that if we have learned the conditional score function, then sampling can be done efficiently. While Block et al. (2020a) demonstrated that unconditional score learning can be learned through standard statistical learning techniques, we generalize these results to the case of conditional score learning and conclude the section by proving that with sufficiently many samples, we can efficiently sample from a distribution close to our target. In this section, we drop the subscript h for clarity, as our theoretical analysis treats each $\mathbf{s}_{\theta,h}$ separately; while empirically one sees better success in training the score estimates jointly, the focus of this paper is not on sampling and score estimation and so we make the simplifying assumption for the sake of convenience.

H.1. Denoising Diffusion Probabilistic Models

We begin by motivating the sampling procedure described in (2.2), which is derived by fixing a horizon T and considering the continuum limit as $\alpha \downarrow 0$ and $J = \frac{T}{\alpha}$. More precisely, consider a forward process satisfying the stochastic differential equation (SDE) for $0 \le t \le T$:

$$d\mathsf{a}^t = -\mathsf{a}^t dt + \sqrt{2} dB_t, \quad \mathsf{a}^0 \sim q,$$

where B_t is a Brownian motion on $\mathbb{R}^d$ and a^0 is sampled from the target density. Applying the standard time reversal to this process results in the following SDE:

$$d\mathsf{a}_{\leftarrow}^{T-t} = \left(\mathsf{a}_{\leftarrow}^t + 2\nabla \log q_{T-t}(\mathsf{a}_{\leftarrow}^t) \right) dt + \sqrt{2} dB_t, \quad \mathsf{a}_{\leftarrow}^0 \sim q_T,$$

where q_t is the law of a^t. Because the forward process mixes exponentially quickly to a standard Gaussian, in order to approximately sample from q, the learner may sample $\widetilde{\mathsf{a}}_{\leftarrow}^0 \sim \mathcal{N}(0, \mathbf{I})$ and evolving $\widetilde{\mathsf{a}}_{\leftarrow}^t$ according to the SDE above. Note that the classical Euler-Maruyama discretization of the above procedure is exactly (2.2), but with the true score $\nabla \log q_{T-t}$ replaced by score estimates $\mathbf{s}_\theta(\cdot, T-t) : \mathbb{R}^d \to \mathbb{R}^d$; we may hope that if $\mathbf{s}_\theta(\cdot, T-t) \approx \nabla \log q_{T-t}$ as functions, then the procedure in (2.2) produces a sample close in law to q. Indeed, the following result provides a quantitative bound:

Theorem 7 (Corollary 4, Chen et al. (2022))**.** *Suppose that a distribution q on $\mathbb{R}^d$ is supported on some ball of radius $R \ge 1$. Let C be a universal constant, fix $\varepsilon > 0$, and let α, J be set as in* (H.1)*. If we have a score estimator $\mathbf{s}_\theta : \mathbb{R}^d \times [\tau] \to \mathbb{R}^d$ such that*

$$\max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, j) - \nabla \log q_{[\alpha j]}(\mathsf{a}) \right|\right|^2 \right] \le \varepsilon^4,$$

then

$$\sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \le 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathrm{Law}(\mathsf{a}^J)} \left[f(\widehat{\mathsf{a}}) \right] - \mathbb{E}_{\mathsf{a}^* \sim q} \left[f(\mathsf{a}^*) \right] \le \varepsilon^2,$$

where a^J is sampled as in (2.2)*.*

Remark H.4. As a technical aside, we note that Chen et al. (2022, Corollary 4) applies to an "early stopped" DDPM, in the sense that the denoising is stopped in slightly fewer than J steps. On the other hand, for the choice of α given above, Chen et al. (2022, Lemma 20 (a)) demonstrates that this distribution is ε^2-close in Wasserstein distance to the sample produced by using all J steps and so by multiplying C above by a factor of 2 the guarantee is preserved. Because in practice we do not stop the DDPM early, we phrase Theorem 7 in the way above as opposed to the more complicated version with the early stopping.

Remark H.5. While (Chen et al., 2022; Lee et al., 2023) show that if $\mathbf{s}_\theta$ is close to the $\mathbf{s}_{\star,h}$ in $L^2(q_{[\alpha j]})$ and q satisfies mild regularity properties, then the law of a_h^J will be close in total variation to q. Unfortunately, the required regularity of q, that the score is Lipschitz, is too strong to hold in many of our applications, such as when the data lie close to a low-dimensional manifold. In such cases, Chen et al. (2022) provided guarantees in a weaker metric on distributions. We emphasize that even with full dimensional support, the Lipschitz constant of $\nabla \log q$ is likely large and thus the dependence on this constant appearing in Chen et al. (2022, Theorem 2) is unacceptable. In particular, this subtle point is what necessitates the intricate construction of our paper; as remarked in Section 3, if we could expect the score to be sufficiently regular and producing a sample close in total variation to the target distribution were feasable, the problem would be trivial.

While Theorem 7 applies to unconditional sampling, it is easy to derive conditional sampling guarantees as a corollary.

Corollary H.1. *Suppose that q is a joint distribution on actions a and observations $\boldsymbol{\rho}_{\mathrm{m},h} \in \mathbb{R}^{d'}$. Further assume that the marginals over $\mathbb{R}^d$ are fully supported in a ball of radius $R \geq 1$. Then there exists a universal constant C such that for all small $\varepsilon > 0$, if J and α are set as in (H.1) and*

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, j, \boldsymbol{\rho}_{\mathrm{m},h}) - \nabla \log q_{[\alpha j]}(\mathsf{a} | \boldsymbol{\rho}_{\mathrm{m},h}) \right|\right|^2 \right] \right] \leq \varepsilon^4, \tag{H.2}$$

then

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\inf_{\mu \in \mathscr{C}(\mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h}), q(\cdot | \boldsymbol{\rho}_{\mathrm{m},h}))} \mathbb{P}_{(\widehat{\mathsf{a}}, \mathsf{a}^*) \sim \mu} \left(||\widehat{\mathsf{a}} - \mathsf{a}^*|| \geq \varepsilon \right) \right] \leq 3\varepsilon$$

Proof. We begin by showing an intermediate result,

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \leq 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\widehat{\mathsf{a}}) \right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\mathsf{a}^*) \right] \right] \leq 3\varepsilon^2. \tag{H.3}$$

using Theorem 7. Let

$$\mathcal{A} = \left\{ \max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, j, \boldsymbol{\rho}_{\mathrm{m},h}) - \nabla \log q_{[\alpha j]}(\mathsf{a} | \boldsymbol{\rho}_{\mathrm{m},h}) \right|\right|^2 \right] \leq \varepsilon^2 \right\}.$$

By Markov's inequality and (H.2), it holds that

$$\mathbb{P}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left(\mathcal{A}^c \right) \leq \frac{\varepsilon^4}{\varepsilon^2} = \varepsilon^2$$

and thus

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \leq 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\widehat{\mathsf{a}}) \right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\mathsf{a}^*) \right] \right] \\
&= \mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\mathbf{I}[\mathcal{A}] \sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \leq 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\widehat{\mathsf{a}}) \right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\mathsf{a}^*) \right] \right] \\
&+ \mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\mathbf{I}[A^c] \sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \leq 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\widehat{\mathsf{a}}) \right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[f(\mathsf{a}^*) \right] \right] \\
&\leq \mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\mathbf{I}[\mathcal{A}] \inf_{q' \in \Delta(\mathbb{R}^d)} W_2(q(\cdot | \boldsymbol{\rho}_{\mathrm{m},h}), q') + \mathrm{TV}\left(q', \mathrm{Law}(\pi^\tau)\right) \right] + 2\varepsilon^2.
\end{aligned}$$

For each $\boldsymbol{\rho}_{\mathrm{m},h}$, we may apply Theorem 7 and observe that for $\boldsymbol{\rho}_{\mathrm{m},h} \in \mathcal{A}$,

$$\sup_{f\colon ||f||_\infty \vee ||||\nabla f||||_\infty \le 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h})} [f(\widehat{\mathsf{a}})] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h})} [f(\mathsf{a}^*)] \le \varepsilon^2,$$

which proves (H.3). Now, for any fixed $\boldsymbol{\rho}_{\mathrm{m},h}$, by Markov's inequality and the definition of Wasserstein distance,

$$\inf_{\mu \in \mathscr{C}(\mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h}), q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h}))} \mathbb{P}_{(\widehat{\mathsf{a}}, \mathsf{a}^*) \sim \mu} (||\widehat{\mathsf{a}} - \mathsf{a}^*|| \ge \varepsilon) \le \frac{W_1(\mathtt{DDPM}(\mathbf{s}_\theta, \boldsymbol{\rho}_{\mathrm{m},h}), q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h}))}{\varepsilon}.$$

The result follows. □

Note that the guarantee in Corollary H.1 is precisely what we need to control the one step imitation error in Theorem 2; thus, the problem of conditional sampling has been reduced to estimating the score. In the subsequent section, we will apply standard statistical learning techniques to provide a nonasymptotic bound on the quality of a score estimator.

H.2. Score Estimation

In the previous section we have shown that conditional sampling can be reduced to the problem of learning the conditional score. While there exist non-asymptotic bounds for learning the unconditional score (Block et al., 2020a), they apply to a slightly different score estimator than is typically used in practice. Here we upper bound the estimation error in terms of the complexity of the space of parameters Θ.

Observe that in order to apply Corollary H.1, we need a guarantee on the error of our score estimate in $L^2(q_{[\alpha j]})$ for all $j \in [J]$. Ideally, then, for fixed $\boldsymbol{\rho}_{\mathrm{m},h}$ and $t = \alpha j$, we would like to minimize $\mathbb{E}_{\mathsf{a} \sim q_{[t]}} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}|\boldsymbol{\rho}_{\mathrm{m},h}) \right|\right|^2 \right]$, where the inner norm is the Euclidean norm on $\mathbb{R}^d$. Unfortunately, because $q_{[t]}$ itself is unkown, we cannot even take an empirical version of this loss. Instead, through a now classical integration by parts (Hyvärinen & Dayan, 2005; Vincent, 2011; Song & Ermon, 2019), this objective can be shown to be equivalent to minimizing

$$\mathcal{L}_{\mathrm{DDPM}}(\theta, \mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},}, t) = \mathbb{E}_{\mathsf{a} \sim q_{[t]}} \left[\left|\left| \mathbf{s}_\theta \left(e^{-t}\mathsf{a} + \sqrt{1 - e^{-2t}} \boldsymbol{\gamma}, \boldsymbol{\rho}_{\mathrm{m},h}, t \right) + \frac{1}{\sqrt{1 - e^{-2t}}} \boldsymbol{\gamma} \right|\right|^2 \right].$$

Because we are really interested in the expectation over the joint distribution $(\mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},h})$, we may take the expectation over $\boldsymbol{\rho}_{\mathrm{m},h}$ and recover (3.1) as the empirical approximation. We now prove the following result for a single time step t:

Proposition H.1. *Suppose that q is a distribution such that $q(\cdot|\boldsymbol{\rho}_{\mathrm{m},i})$ is supported on a ball of radius R for q-almost every $\boldsymbol{\rho}_{\mathrm{m},h}$. For fixed $j \in [J]$ and α from (H.1), let $t = j\alpha$ and suppose that there is some $\theta^* \in \Theta_j$ such that $\mathbf{s}_\star(\cdot, \cdot, t) = \mathbf{s}_{\theta^*}(\cdot, \cdot, t) = \nabla \log q_{[t]}(\cdot|\cdot)$, i.e., $\mathbf{s}_\theta$ is rich enough to represent the true score at time t. Suppose further that the class of functions $\{\mathbf{s}_\theta | \theta \in \Theta_j\}$ satisfies for all $\theta \in \Theta_j$,*

$$\sup_{\substack{\theta, \theta' \in \Theta_j \\ ||\mathsf{a}|| \vee ||\mathsf{a}'|| \le R \\ \boldsymbol{\rho}_{\mathrm{m},h}}} \left|\left| \mathbf{s}_\theta(\mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \mathbf{s}_{\theta'}(\mathsf{a}', \boldsymbol{\rho}_{\mathrm{m},h}, t) \right|\right| \le c \frac{d^2 (R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8}$$

for some universal constant $c > 0$. Recall the Rademacher term $\mathcal{R}_n(\Theta_j)$ defined in Definition H.1, and let

$$\widehat{\theta} \in \operatorname*{arg\,min}_{\theta \in \Theta} \sum_{i=1}^n \mathcal{L}_{\mathrm{DDPM}}(\theta, \mathsf{a}_i, \boldsymbol{\rho}_{\mathrm{m},i}, t)$$

for independent and identically distributed $(\mathsf{a}_i, \boldsymbol{\rho}_{\mathrm{m},i}) \sim q$. Then it holds with probability at least $1 - \delta$ over the data that

$$\begin{aligned} &\mathbb{E}_{(\mathsf{a}_t, \boldsymbol{\rho}_{\mathrm{m},h}) \sim q_{[t]}} \left[\left|\left| \mathbf{s}_{\widehat{\theta}}(\mathsf{a}_t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}_t | \boldsymbol{\rho}_{\mathrm{m},h}) \right|\right|^2 \right] \\ &\le c \cdot \sqrt{\frac{\log(dn)}{1 - e^{-2t}}} \left(\mathcal{R}_n(\Theta) + \frac{d^2 (R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8} \cdot \sqrt{\frac{d \log\left(\frac{4dn}{\delta}\right)}{n}} \right). \end{aligned}$$

Before we provide a proof, we recall the following result:

Lemma H.2. *Suppose that q is supported in a ball of radius R and let $t \geq \alpha$ for α as in (H.1). Then $\nabla \log q_{[t]}(\cdot|\cdot)$ is L-Lipschitz with respect to the first parameter for*

$$L = \frac{dR^2(R \vee \sqrt{d})^2}{\varepsilon^8}.$$

In particular,

$$\sup_{\substack{||\mathsf{a}|| \vee ||\mathsf{a}'|| \leq R \\ \boldsymbol{\rho}_{\mathrm{m},h}}} \left|\left|\nabla \log q_{[t]}(\mathsf{a}|\boldsymbol{\rho}_{\mathrm{m},h}) - \nabla \log q_{[t]}(\mathsf{a}'|\boldsymbol{\rho}_{\mathrm{m},h})\right|\right| \leq 2LR$$

and there exists some assignment of Θ and $\mathbf{s}_\theta$ that satisfies the boundedness condition in Proposition H.1.

Proof. The first statement follows from eplacing the ε in Chen et al. (2022, Lemma 20 (c)) by ε^2. The second statement follows immediately from the first. □

We also require the following standard result:

Lemma H.3. *If $\mathcal{R}_n(\Theta_j)$ is defined as in Definition H.1, then*

$$\mathbb{E}_{\boldsymbol{\gamma}_1,\dots,\boldsymbol{\gamma}_n}\left[\sup_{\substack{\theta \in \Theta_j \\ 1 \leq j \leq J}} \frac{1}{n} \cdot \sum_{i=1}^n \left\langle \mathbf{s}_\theta(\mathsf{a}, \boldsymbol{\rho}_{\mathrm{m},i}, j), \boldsymbol{\gamma}_i \right\rangle\right] \leq \sqrt{\pi \log(dn)} \cdot \mathcal{R}_n(\Theta_j)$$

Proof. This statement is classical and follows immediately from the fact that the norm of a Gaussian is independent from its sign as well as the fact that $\mathbb{E}\left[\max_{i,j}(\boldsymbol{\gamma}_i)_j\right] \leq \sqrt{\pi \log(dn)}$ by classical Gaussian concentration. See Van Handel (2014) for more details. □

Proof of Proposition H.1. Let P_n denote the empirical measure on n independent samples $\left\{\left(\mathsf{a}_i, \boldsymbol{\rho}_{\mathrm{m},i}, \boldsymbol{\gamma}_i\right)\right\}$ and let $\mathsf{a}_i^t = e^{-t}\mathsf{a}_i + \sqrt{1-e^{-2t}}\boldsymbol{\gamma}_i$. Let $C_t = \sqrt{1-e^{-2t}}$ and observe that by definition and realizability,

$$P_n\left(\left|\left|C_t \mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \boldsymbol{\gamma}\right|\right|^2\right) \leq \cdot P_n\left(\left|\left|C_t \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h}) - \boldsymbol{\gamma}\right|\right|^2\right). \tag{H.4}$$

We emhasize that by Lemma H.2, realizability does not make the result vaccuous. Adding and subtracting $C_t \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h})$ from the left hand inequality, expanding and rearranging, we see that

$$\begin{aligned} C_t^2 P_n\left(\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h})\right|\right|^2\right) &\leq 2C_t \cdot P_n\left(\left\langle \mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h}), \boldsymbol{\gamma}\right\rangle\right) \\ &\leq 2C_t \cdot P_n\left(\sup_{\theta \in \Theta_j} \left\langle \mathbf{s}_\theta(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h}), \boldsymbol{\gamma}\right\rangle\right). \end{aligned}$$

We now claim that with probability at least $1-\delta$, it holds that

$$\begin{aligned} P_n\left(\sup_{\theta \in \Theta} \left\langle \mathbf{s}_\theta(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h}), \boldsymbol{\gamma}\right\rangle\right) &\leq \mathbb{E}\left[P_n\left(\sup_{\theta \in \Theta_j} \left\langle \mathbf{s}_\theta(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\boldsymbol{\rho}_{\mathrm{m},h}), \boldsymbol{\gamma}\right\rangle\right)\right] \\ &\quad + B \cdot \sqrt{\frac{d \log\left(\frac{2d}{\delta}\right)}{n}}, \end{aligned}$$

where

$$B = c\frac{d^2(R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8} \tag{H.5}$$

for some universal constant $c > 0$. To see this, we claim that with probability at least $1 - \frac{\delta}{2}$, it holds that $||\mathsf{a}_i^t|| \leq c\left(R + \sqrt{d\log\left(\frac{2nd}{\delta}\right)}\right)$ for all $1 \leq i \leq n$. Indeed, this follows by Gaussian concentration in Jin et al. (2019, Lemmata 1 & 2). We may now apply Lemma H.2 to a bound on the osculation of $\mathbf{s}_\theta - \nabla \log q_{[t]}$ in the ball of the above radius. Conditioning on the event that $||\mathsf{a}_i^t||$ is bounded by the above, we may argue as in Wainwright (2019, Theorem 4.10) that if we let the function

$$G = G(\mathsf{a}_1, \boldsymbol{\rho}_{\mathrm{m},1}, \ldots, \mathsf{a}_n, \boldsymbol{\rho}_{\mathrm{m},n}) = P_n\left(\sup_{\theta \in \Theta_j} \left\langle \mathbf{s}_\theta(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t | \boldsymbol{\rho}_{\mathrm{m},h}), \boldsymbol{\gamma} \right\rangle\right),$$

then for any i, on the event of bounded norm, replacing $(\mathsf{a}_i, \boldsymbol{\rho}_{\mathrm{m},i})$ with $(\mathsf{a}_i', \boldsymbol{\rho}_{\mathrm{m},i}')$ and leaving other terms unchanged changes ensures that $|G - G'| \leq \frac{2B}{n}\boldsymbol{\gamma}_i$. Thus by Jin et al. (2019, Corollary 7) and a union bound, the claim holds. Because $\boldsymbol{\gamma}$ is mean zero, we have

$$\begin{aligned}\mathbb{E}\left[P_n\left(\sup_{\theta \in \Theta} \left\langle \mathbf{s}_\theta(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t | \boldsymbol{\rho}_{\mathrm{m},h}), \boldsymbol{\gamma} \right\rangle\right)\right] &\leq \mathbb{E}\left[P_n\left(\sup_{\theta \in \Theta} \left\langle \mathbf{s}_\theta(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t), \boldsymbol{\gamma} \right\rangle\right)\right] \\ &\leq \sqrt{\pi \log(dn)} \cdot \mathcal{R}_n(\Theta_j),\end{aligned}$$

where the last inequality follows by Lemma H.3 and the fact that $t = jJ$. Summing up the argument until this point and rearranging tells us that with probability at least $1 - \delta$, it holds that

$$P_n\left(\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t | \boldsymbol{\rho}_{\mathrm{m},h})\right|\right|^2\right) \leq \frac{2}{C_t}\sqrt{\pi \log(nd)} \cdot \mathcal{R}_n(\Theta) + \frac{B}{C_t} \cdot \sqrt{\frac{d\log\left(\frac{2nd}{\delta}\right)}{n}},$$

with B given in (H.5). We now use a uniform norm comparison between population and empirical norms to conclude the proof. Indeed, it holds by Rakhlin et al. (2017, Lemma 8.i & 9) that there exists a critical radius

$$r_n \leq cB\log^3(n)\mathcal{R}_n(\Theta_j)^2$$

such that with probability at least $1 - \delta$,

$$\begin{aligned}&\mathbb{E}_{(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}) \sim q_{[t]}}\left[\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t | \boldsymbol{\rho}_{\mathrm{m},h})\right|\right|^2\right] \\ &\qquad \leq 2 \cdot P_n\left(\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t | \boldsymbol{\rho}_{\mathrm{m},h})\right|\right|^2\right) + cr_n + c\frac{\log\left(\frac{1}{\delta}\right) + \log\log n}{n},\end{aligned}$$

where again c is some universal constant. Combining this with our earlier bound on the empirical distance and a union bound, after rescaling δ, we have that

$$\begin{aligned}\mathbb{E}_{(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}) \sim q_{[t]}}\left[\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \boldsymbol{\rho}_{\mathrm{m},h}, t) - \nabla \log q_{[t]}(\mathsf{a}^t | \boldsymbol{\rho}_{\mathrm{m},h})\right|\right|^2\right] &\leq \frac{4}{C_t}\sqrt{\pi \log(nd)} \cdot \mathcal{R}_n(\Theta_j) + \frac{2B}{C_t} \cdot \sqrt{\frac{d\log\left(\frac{4nd}{\delta}\right)}{n}} \\ &\quad + cB\log^3(n) \cdot \mathcal{R}_n^2(\Theta_j) + c\frac{\log\left(\frac{2}{\delta}\right) + \log\log(n)}{n}\end{aligned}$$

with probability at least $1 - \delta$. This concludes the proof. □

Remark H.6. For the sake of simplicity, in the proof of Proposition H.1 we applied uniform deviations and recovered the "slow rate" of $\mathcal{R}_n(\Theta)$, up to logarithmic factors. Indeed, if we were to further assume that the score function class is star-shaped around the true score, we could recover a faster rate, as was done in the case of unconditional sampling in Block et al. (2020a) with a slightly different loss. While in our proof the appeal to Rakhlin et al. (2017) to control the population norm by the empirical norm could be replaced with a simpler uniform deviations argument because we have already given up on the fast rate, such an argument is necessary in the more refined analysis. As the focus of this paper is not on the sampling portion of the end-to-end analysis, we do not include a rigorous proof of the case of fast rates for the sake of simplicity and space.

Remark H.7. While we assumed for simplicity that the score was realizable with respect to our function class for every time $t = \alpha j$, this condition can be relaxed to approximate realizability in a standard way. In particular, if the score is ε-far away from some function representable by our class in a pointwise sense, then we can add an ε to the right hand side of (H.4) with minimal modification to the proof.

With Proposition H.1, and a union bound, we recover the following result:

Proposition H.4. *Suppose that the conditions on $\mathbf{s}_\theta$ in Proposition H.1 continue to hold. Let J and α be as in (H.1) and suppose that $\alpha \le \frac{1}{2}$. Then, with probability at least $1-\delta$ over $\mathcal{D}'$, it holds that*

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[\left| \left| \mathbf{s}_\theta(\mathsf{a}, j, \boldsymbol{\rho}_{\mathrm{m},h}) - \nabla \log q_{[\alpha j]}(\mathsf{a} | \boldsymbol{\rho}_{\mathrm{m},h}) \right| \right|^2 \right] \right]$$
$$\le c \frac{dR(R \vee \sqrt{d}) \log(dn)}{\varepsilon^4} \mathcal{R}_n(\Theta) + c \frac{d^3 \left(R^2 + d \log\left(\frac{ndR}{\delta\varepsilon}\right)\right)}{\varepsilon^{12}} \sqrt{\frac{d \log\left(\frac{4dnR}{\delta\varepsilon}\right)}{n}}$$

In particular if

$$\mathcal{R}_n(\Theta_j) \le C_\Theta n^{-1/\nu}$$

for some $\nu \ge 2$ and all $j \in [J]$, then for

$$n \ge c \left(\frac{C_\Theta dR(R \vee \sqrt{d}) \log(dn)}{\varepsilon^4} \right)^{4\nu} \vee \left(\frac{d^6 (R^4 \vee d^2 \log^3\left(\frac{ndR}{\delta\varepsilon}\right))}{\varepsilon^{24}} d^2 \right)^{4\nu}$$

it holds that with probability at least $1-\delta$,

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}} \left[\max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}(\cdot | \boldsymbol{\rho}_{\mathrm{m},h})} \left[\left| \left| \mathbf{s}_\theta(\mathsf{a}, j, \boldsymbol{\rho}_{\mathrm{m},h}) - \nabla \log q_{[\alpha j]}(\mathsf{a} | \boldsymbol{\rho}_{\mathrm{m},h}) \right| \right|^2 \right] \right] \le \varepsilon^4.$$

Proof. We begin by noting that

$$1 - e^{-2t} \ge 1 - e^{-2\alpha} \ge \alpha$$

because $2\alpha \le 1$. We now apply Proposition H.1 and take a union bound over $j \in [J]$. The result follows. □

We note that in our simplified analysis, we have assumed that $N_{\mathtt{aug}} = 1$, i.e., for each sample, we take a single noise level from the path. In practice, we use many augmentations per sample. Again, as the focus of our paper is not on score estimation and sampling, we treat this as a simple convenience and leave open to future work the problem of rigorously demonstrating that multiple augmentations indeed help with learning. Finally, for a discussion on bounding $\mathcal{R}_n(\Theta)$, see Wainwright (2019).

Proof of Theorem 6. We note that the proof follows immediately from combining Corollary H.1 with Proposition H.4. □

I. End-to-end Guarantees and the Proof of Theorem 1

In this section, we provide a number of end-to-end guarantees for the learned imitation policy under various assumptions. The core of the section is Theorem 8, which provides the basis for the final proof of Theorem 1 in the body by uniting the analysis in the composite MDP from Appendix E, the control theory from Appendix G, and the sampling guarantees from Appendix H. We now summarize the organisation of the appendix:

- In Appendix I.1, we recall the association between the control setting and the composite MDP presented in Section 4, as well as rigorously instantiating the direct decomposition and the expert policy.
- In Appendix I.2, we state a reduction from imitation learning to conditional sampling, which we then use to derive a proof of Theorem 1.

- In Appendix I.3, we demonstrate that if the demonstrator policy is assumed to be TVC, then we can recover stronger guarantees than those provided in Theorem 1 without this assumption; in particular, we show that we can bound the *joint* imitation loss as well as the marginal and final versions.
- In Appendix I.4, we show that if we were able to produce samples from a distribution close in *total variation* to the expert policy distribution, as opposed to the weaker optimal transport metric that we consider in the rest of the paper, then without any further assumptions, imitation learning is easily achievable.
- In Appendix I.5, we show that if we remove the data augmentation from TODA, i.e., we set $\sigma = 0$, then we can recover similar guarantees under the assumption that the imitator policy $\hat{\pi}$ is TVC. In this way, we show that in some sense, total variation continuity is the important property imparted by smoothing.
- In Appendix I.6, we demonstrate the utility of our imitation losses, showing that for Lipschitz cost functions decomposing in natural ways, our imitation losses as defined in Definition 2.2 provide control over the difference in expected cost under expert and imitated distributions.
- Finally, in Appendix I.7, we collect a number of useful lemmata that we use throughout the appendix.

I.1. Preliminaries

Here, we state various preliminaries to the end-to-end theorems. For simplicity, to avoid complications with the boundary effects at $h = 1$, we re-define $h = 1$-memory chunks $\rho_{m,1}$ as elements $\mathscr{P}_{\tau_m - 1}$ by prepending the necessary zeros – i.e. $\rho_{m,1} = (0, 0, \ldots, 0, \mathbf{x}_1)$– and similarly modifying $\rho_{c,1} \in \mathscr{P}_{\tau_c}$ by prepending zeros. We first recall the definitions of the composite-states and -actions from Section 4. The prepending of zeros in the $h = 1$ case is mentioned above. For $h > 1$, recall that $\mathsf{s}_h = (\mathbf{x}_{t_{h-1}:t_h}, \mathbf{u}_{t_{h-1}:t_h - 1})$ and that $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$, where we again emphasize that a_h begins at the same t that s_{h+1} does. We further recall that $\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_{t \in [t_{h-1}:t_h]} ||\mathbf{x}_t - \mathbf{x}'_t|| \vee \max_{t \in [t_{h-1}:t_h - 1]} ||\mathbf{u}_t - \mathbf{u}'_t||$, $\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_{t \in [t_h - \tau_m:t_h]} ||\mathbf{x}_t - \mathbf{x}'_t|| \vee \max_{t \in [t_h - \tau_m:t_h - 1]} ||\mathbf{u}_t - \mathbf{u}'_t||$, and $\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) = \left|\left|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\right|\right|$. Finally, for $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_c}, \bar{\mathbf{x}}_{1:\tau_c}, \bar{\mathbf{K}}_{1:\tau_c})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_c}, \bar{\mathbf{x}}'_{1:\tau_c}, \bar{\mathbf{K}}'_{1:\tau_c})$, recall from (4.2) that

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') := c_1 \max_{1 \le k \le \tau_c} (\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|) + \mathbf{I}_{0,\infty}\{\mathcal{E}\},$$

where we $\mathcal{E} = \{\max_{1 \le k \le \tau_c} \max\{\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\|, \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\|, \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|\} \le c_2\}$, $\mathbf{I}_{0,\infty}$ is the indicator taking infinite value when the event fails to hold, and c_1 and c_2 are given in Definition G.5.

Direct Decomposition and Smoothing Kernel. This section will invoke the generalizations Theorem 2 which requires TVC only subspace of the state space. This invokes the direct decomposition explained in Appendix E.

Definition I.1 (Direct Decomposition and Smoothing Kernel)**.** We consider the decomposition of $\mathcal{S} = \mathcal{Z} \oplus \mathcal{V}$, where $\mathcal{Z} = \mathscr{P}_{\tau_m - 1}$ are the coordinates of $\rho_{c,h}$ corresponding to the memory chunk $\rho_{m,h}$, and $\mathcal{V}$ are all remaining coordinates We let $\phi_{\mathcal{Z}} : \mathcal{S} \to \mathcal{Z}$ denote the projection onto the coordinates in $\mathcal{Z}$. We instantiate the smoothing kernel W_σ as follows: For $\mathsf{s} = \rho_{c,h} \in \mathscr{P}_{\tau_c}$, we let

$$\mathsf{W}_\sigma(\mathsf{s}) = \mathcal{N}\left(\rho_{c,h}, \begin{bmatrix} \sigma^2 \mathbf{I}_{\mathcal{Z}} & 0 \\ 0 & 0 \end{bmatrix}\right),$$

where $\mathbf{I}_{\mathcal{Z}}$ denotes the identity supported on the coordinates in $\mathcal{Z}$ as described above.

We note that the above direct decomposition satisfies the requiste compatibility assumptions explained in Appendix E. Note also that $\mathsf{d}_{\mathrm{IPS}}$ and W_σ are compatible with the above direct decomposition.

Chunking Policies. We continue by centralizing a definition of chunking policies.

Definition I.2 (Policy and Initial-State Distributions)**.** Given an *chunking policy* $\pi = (\pi_h)_{h=1}^H$ with $\pi_h : \mathscr{P}_{\tau_m - 1} \to \Delta(\mathcal{A})$, we let $\mathcal{D}_\pi$ denote the distribution over ρ_T and $\mathsf{a}_{1:H}$ induced by selecting $\mathsf{a}_h \sim \pi_h(\rho_{m,h})$, and rolling out the dynamics as described in Section 2. We extend chunking policies to maps $\pi_h : \mathcal{S} = \mathscr{P}_{\tau_c} \to \Delta(\mathcal{A})$ by expressing $\pi_h = \pi_h \circ \phi_{\mathcal{Z}}$ (i.e., projection $\rho_{c,h}$ onto its $\rho_{m,h}$-components). Further, we let $\mathsf{P}_{\mathrm{init}}$ denote the distribution of $\mathbf{x}_1$ under $\rho_T \sim \mathcal{D}_{\mathrm{exp}}$.

Remark I.1. The notation $\mathcal{D}_\pi$ denotes the special case of chunking policies in the control setting of Section 2, whereas we reserve the seraf font D_π for the distribution induced by policies in the abstract MDP. For composite MDPs instantiated as in Section 4.1, the two exactly coincide.

Construction of $\pi^\star$ for composite MDP. We now explain how to extract $\pi^\star$ from $\mathcal{D}_{\mathrm{exp}}$ in the composite MDP.

Definition I.3 (Policies corresponding to $\mathcal{D}_{\mathrm{exp}}$)**.** Define the following sequence kernels $\pi^\star = (\pi^\star_h)_{h=1}^H$ and $\pi^\star_{\mathrm{dec}} = (\pi^\star_{\mathrm{dec},h})_{h=1}^H$ via the following process. Let $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$, and let $\mathsf{a}_{1:H} = \mathtt{synth}(\boldsymbol{\rho}_T)$; further, let $\boldsymbol{\rho}_{\mathrm{m},1:H}$ be the corresponding memory-chunks from $\boldsymbol{\rho}_T$. Let

- $\pi^\star_h(\cdot) : \mathscr{P}_{\tau_{\mathrm{m}}-1} \to \mathcal{A}$ denote a regular conditional probability corresponding to the distribution over a_h given $\boldsymbol{\rho}_{\mathrm{m},h}$ in the above construction.
- $\pi^\star_{\mathrm{dec},h}(\cdot) : \mathscr{P}_{\tau_{\mathrm{m}}-1} \to \mathcal{A}$ denote a regular conditional probability corresponding to the distribution over a_h given an augmented $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} \sim \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2\mathbf{I})$.

Finally, for $\pi^\star$ as constructed above, $\mathsf{P}^\star_h$ denotes the distribution over $\boldsymbol{\rho}_{\mathrm{c},h}$ under $\mathcal{D}_{\pi^\star}$. By Lemma I.6, this is in fact equal to the distribution over $\boldsymbol{\rho}_{\mathrm{c},h}$ under $\mathcal{D}_{\mathrm{exp}}$. Notice further, therefore, that $\phi_{\mathcal{Z}} \circ \mathsf{P}^\star_h$ is precisely the distribution of $\boldsymbol{\rho}_{\mathrm{m},h}$ under $\mathcal{D}_{\mathrm{exp}}$.

Remark I.2. We remark that by Theorem 3, $\pi^\star_h$ is unique up to a measure zero set of $\boldsymbol{\rho}_{\mathrm{m},h}$ as distributed as above, and $\pi^\star_{\mathrm{dec},h}$ is unique almost surely for $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}$ distributed as above. In particular, since the latter has density with respect to the Lebesgue measure and infinite support, $\pi^\star_{\mathrm{dec},h}$ is unique in a Lebesgue almost everywhere sense.

Instantiation of the distance $\mathsf{d}_{\mathcal{A}}$ for pairs of actions. We recall the instantiation of the distance $\mathsf{d}_{\mathcal{A}}$:

Definition I.4 (Instantiation of $\mathsf{d}_{\mathcal{A}}$)**.** We recall $\mathsf{d}_{\mathcal{A}} : \mathcal{A} \times \mathcal{A} \to \mathbb{R}_{\ge 0}$ as defined in (4.2):

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') := c_1 \max_{1 \le k \le \tau_{\mathrm{c}}} (\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|) + \mathbf{I}_{0,\infty}\{\mathcal{E}\},$$

where we define $\mathcal{E} := \{\max_{1 \le k \le \tau_{\mathrm{c}}} \max\{\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\|, \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\|, \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|\} \le c_2\}$, $\mathbf{I}_{0,\infty}$ is the indicator taking infinite value when the event fails to hold, and c_1 and c_2 are constants depending polynomially on all problem parameters, given in Definition G.7.

I.1.1. Preliminaries for joint-distribution imitation.

This section introduces a further *joint imitation gap*, which we can make small under a stronger bounded-memory assumption on $\mathcal{D}_{\mathrm{exp}}$ stated below.

Definition I.5 (Joint Imitation Gap)**.** Given a chunking polcy π', we let

$$\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi) := \inf_{\mu} \mathbb{P}_{\mu}\left[\max_{t \in [T]} \max\left\{\|\mathbf{x}^{\mathrm{exp}}_{t+1} - \mathbf{x}^{\pi}_{t+1}\|, \|\mathbf{u}^{\mathrm{exp}}_t - \mathbf{u}^{\pi}_t\|\right\} > \varepsilon\right],$$

where the infimum is over all couplings between the distribution of $\boldsymbol{\rho}_T$ under $\mathcal{D}_{\mathrm{exp}}$ and that induced by the policy π.

Controlling $\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi)$ requires various additional stronger assumptions (*which we do not require in Theorem 1*), one of which is that the demonstrator has bounded memory:

Definition I.6. We say that the demonstration distribution, synthesis oracle pair $(\mathcal{D}_{\mathrm{exp}}, \mathtt{synth})$ have τ-bounded memory if under $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T}) \sim \mathcal{D}_{\mathrm{exp}}$ and $\mathsf{a}_{1:H} = \mathtt{synth}(\boldsymbol{\rho}_T)$, the conditional distribution of a_h and $\mathbf{x}_{1:t_h-\tau}, \mathbf{u}_{1:t_h-\tau}$ are conditionally independence given $(\mathbf{x}_{t_h-\tau+1:t_h}, \mathbf{u}_{t_h-\tau+1:t_h-1})$.

We note that enforcing Definition I.6 can be relaxed to a mixing time assumption (see Remark I.4). Moreover, we stress that we *do not* need the condition in Definition I.6 if we only seek imitation of marginal distributions (as captured by $\mathcal{L}_{\mathrm{marg},\varepsilon}$ and $\mathcal{L}_{\mathrm{fin},\varepsilon}$), as in Theorem 1.

I.1.2. Translating Control Imitation Losses to Composite-MDP Imitation Gaps

Lemma I.1. *Recall the imitation losses Definitions 2.2 and I.5, and the compsite-MDP imitation gaps Definition 4.1. Further consider, the substitutions defined in Section 4.1, with $\pi^\star$ instantiated as in Definition I.3. Given policies $\pi = (\pi_h)$ with $\pi_h : \mathscr{P}_{\tau_{\mathrm{m}}-1} \to \mathcal{A}$, we can extend $\pi_h : \mathcal{S} = \mathscr{P}_{\tau_{\mathrm{c}}} \to \mathcal{A}$ by the natural embedding of $\mathscr{P}_{\tau_{\mathrm{m}}-1}$ into $\mathscr{P}_{\tau_{\mathrm{c}}}$. Then, for any $\varepsilon > 0$,*

$$\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi) \le \Gamma_{\mathrm{marg},\varepsilon}(\pi \parallel \pi^\star).$$

If we instead consider the the substitutions defined in Section 4.1, but set $\mathsf{d}_{\mathcal{S}}$ to equal d_{IPS}, which only measures distance in the final coordinate of each trajectory chunk $\boldsymbol{\rho}_{\mathrm{c},h}$,

$$\mathcal{L}_{\mathrm{fin},\varepsilon}(\pi) \le \Gamma_{\mathrm{marg},\varepsilon}(\pi \parallel \pi^\star), \quad \mathsf{d}_{\mathcal{S}}(\cdot,\cdot) \leftarrow \mathsf{d}_{\text{IPS}}(\cdot,\cdot) \tag{I.1}$$

Finally, if $\mathcal{D}_{\exp}$ has $\tau \le \tau_{\mathrm{m}}$-bounded memory,

$$\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi) \le \Gamma_{\mathrm{joint},\varepsilon}(\pi \parallel \pi^\star).$$

Proof. Let's start with the first bound, let superscript $\exp$ denote objects from $\mathcal{D}_{\exp}$ and superscript π from $\mathcal{D}_\pi$, the distribution induced by chunking policy π. Letting $\inf_\mu$ denote infima over couplings between the two, we have

$$\begin{aligned}
\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi) &:= \max_{t\in[T]} \inf_\mu \left\{ \mathbb{P}_\mu\left[\|\mathbf{x}^{\exp}_{t+1} - \mathbf{x}^\pi_{t+1}\| > \varepsilon\right], \mathbb{P}_\mu\left[\|\mathbf{u}^{\exp}_t - \mathbf{u}^\pi_t\| > \varepsilon\right]\right\} \\
&:= \max_{t\in[T]} \inf_\mu \left\{ \mathbb{P}_\mu\left[\|\mathbf{x}^{\exp}_{t+1} - \mathbf{x}^\pi_{t+1}\| \vee \|\mathbf{u}^{\exp}_t - \mathbf{u}^\pi_t\| > \varepsilon\right]\right\} \\
&\le \max_{h\in[H]} \inf_\mu \left\{ \mathbb{P}_\mu\left[\max_{0\le i\le \tau_{\mathrm{c}}} \|\mathbf{x}^{\exp}_{t_h-i} - \mathbf{x}^\pi_{t_h-i}\| \vee \max_{1\le i\le \tau_{\mathrm{c}}} \|\mathbf{u}^{\exp}_{t_h-i} - \mathbf{u}^\pi_{t_h-i}\|\right]\right\} \\
&\le \max_{h\in[H]} \inf_\mu \left\{ \mathbb{P}_\mu\left[\mathsf{d}_{\mathcal{S}}(\boldsymbol{\rho}^{\exp}_{\mathrm{c},h}, \boldsymbol{\rho}^\pi_{\mathrm{c},h})\right]\right\},
\end{aligned}$$

From Lemma I.6, $\boldsymbol{\rho}^{\exp}_{\mathrm{c},h}$ has the same marginal distribution as $\boldsymbol{\rho}^{\pi^\star}_{\mathrm{c},h}$, the distribution induced by $\pi^\star$ in Definition I.3. Note the subtlety that the joint distribution of these may defer because $\pi^\star$ has limited trajectories. Still, letting $\inf_{\mu'}$ denote infimum over couplings between $\mathcal{D}_\pi$ and $\mathcal{D}_{\pi^\star}$, equality of marginals suffices to ensure

$$\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi) = \max_{h\in[H]} \inf_\mu \left\{ \mathbb{P}_{\mu'}\left[\mathsf{d}_{\mathcal{S}}(\boldsymbol{\rho}^{\pi^\star}_{\mathrm{c},h}, \boldsymbol{\rho}^\pi_{\mathrm{c},h})\right]\right\},$$

which is at most $\Gamma_{\mathrm{marg},\varepsilon}(\pi \parallel \pi^\star)$ by definition Definition 4.1.

For the final-state imitation loss,

$$\begin{aligned}
\mathcal{L}_{\mathrm{fin},\varepsilon}(\pi) &:= \inf_\mu \mathbb{P}_\mu\left[\|\mathbf{x}^{\exp}_{T+1} - \mathbf{x}^\pi_{T+1}\| > \varepsilon\right] \\
&\le \max_{h\in[H]} \inf_\mu \left\{ \mathbb{P}_\mu\left[\mathsf{d}_{\text{IPS}}(\boldsymbol{\rho}^{\exp}_{\mathrm{c},h}, \boldsymbol{\rho}^\pi_{\mathrm{c},h})\right]\right\},
\end{aligned}$$

where again d_{IPS} only measures error in the final state of $\boldsymbol{\rho}_{\mathrm{c},h}$. The corresponding bound in (I.1) follows similarly.

Finally, we have

$$\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi) := \inf_\mu \mathbb{P}_\mu\left[\max_{t\in[T]} \max\left\{\|\mathbf{x}^{\exp}_{t+1} - \mathbf{x}^\pi_{t+1}\|, \|\mathbf{u}^{\exp}_t - \mathbf{u}^\pi_t\|\right\} > \varepsilon\right],$$

When $\mathcal{D}_{\exp}$ has $\tau \le \tau_{\mathrm{m}}$-bounded memory, then, the expert and $\pi^\star$-induced trajectories are identically distributed. Therefore, directly from this observation and Definition 4.1,

$$\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi) = \inf_\mu \mathbb{P}_\mu\left[\max_{t\in[T]} \max\left\{\|\mathbf{x}^{\pi^\star}_{t+1} - \mathbf{x}^\pi_{t+1}\|, \|\mathbf{u}^{\pi^\star}_t - \mathbf{u}^\pi_t\|\right\} > \varepsilon\right] \le \Gamma_{\mathrm{joint},\varepsilon}(\pi \parallel \pi^\star).$$

□

I.2. Proof of Theorem 1 and a general reduction

We now state a reduction from which Theorem 1 is readily derived from our statistical learning analysis of score estimation.

Theorem 8 (Reduction from trajectory imitation to conditional sampling)**.** *Consider applying* TODA *with $\sigma > 0$, and let $\delta \in (0,1)$, and define*

$$\Delta_h(\varepsilon) := \mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h}\sim\mathcal{D}_{\exp}} \mathbb{E}_{\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}\sim\mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h},\sigma^2\mathbf{I})} \inf_{\mu\in\mathscr{C}(\pi^\star_{\mathrm{dec},h}(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}),\hat{\pi}_h(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}))} \mathbb{P}_{(\mathsf{a},\mathsf{a}')\sim\mu}[\mathsf{d}_{\mathcal{A}}(\mathsf{a},\mathsf{a}')]. \tag{I.2}$$

where $\mathsf{d}_{\mathcal{A}}$ is as in Definition I.4, $\boldsymbol{\rho}_{\mathrm{m},h} \sim \mathcal{D}_{\exp}$ is shorthand for $\boldsymbol{\rho}_T \sim \mathcal{D}_{\exp}$, and $\boldsymbol{\rho}_{\mathrm{m},h}$ denotes the corresponding h-th memory chunk of $\mathcal{D}_{\exp}$. Consider the following setup:

- *Suppose that Assumptions 3.1 and 3.2 hold.*
- *Let $c_1, \ldots, c_5$ be the constants defined Definition G.7, which we recall are polynomial in the terms in Assumptions 3.1 and 3.2*
- *Define $d = \tau_c(d_u + d_x + d_u d_x)$.*
- *Suppose that $\tau_c \geq c_3/\eta$*
- *The parameters $\varepsilon, \sigma > 0$ satisfy $5d_x + \log\left(\frac{4\sigma}{\varepsilon}\right) \leq c_4^2/(16\sigma^2)$,*

For all we have

$$\begin{aligned}
&\mathcal{L}_{\mathrm{marg},\varepsilon_1}(\hat{\pi}_\sigma) \vee \mathcal{L}_{\mathrm{fin},\varepsilon_2}(\hat{\pi}_\sigma) \\
&\leq H\left(\frac{2\varepsilon}{\sigma} + 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}e^{-\frac{\eta(\tau_c - \tau_m)}{L_{\mathrm{stab}}}}\right) + \sum_{h=1}^{H} \Delta_h(\varepsilon)
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_1 &= \varepsilon + 4c_5\sigma \cdot \sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)} \\
\varepsilon_2 &= \varepsilon + 4c_5\sigma e^{-\frac{\eta\tau_c}{L_{\mathrm{stab}}}} \cdot \sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}
\end{aligned} \tag{I.3}$$

We first demonstrate how Theorem 1 follows from Theorem 8 and Theorem 6:

Proof of Theorem 1. From Theorem 8, it suffices to show that with probability at least $1-\delta$, it holds that $\Delta_h \leq \frac{3\varepsilon}{\sigma}$ for all $h \in [H]$. Note that by Assumption 3.2 it holds $\mathcal{D}_{\mathrm{exp}}$-almost surely that $||\mathsf{a}_h|| \leq R_{\mathrm{stab}}$ and thus the condition on q in Theorem 6 holds for $R = R_{\mathrm{stab}}$. Moreover, for $d = \tau_c(d_x + d_u + d_x d_u)$, we have that $\mathsf{a} \in \mathbb{R}^d$. By Assumption 3.3, the conditions on the score class $\mathbf{s}_\theta$ hold for us to apply Theorem 6. Note that by assumption,

$$N_{\mathrm{exp}} \geq c\left(\frac{C_\Theta dR(R \vee \sqrt{d})\log(dn)}{(\varepsilon/\sigma)^4}\right)^{4\nu} \vee \left(\frac{d^6(R^4 \vee d^2\log^3\left(\frac{HndR\sigma}{\delta\varepsilon}\right))}{(\varepsilon/\sigma)^{24}}d^2\right)^{4\nu},$$

where we note that the right hand side is $\mathrm{poly}\left(C_\Theta, \varepsilon/\sigma, R_{\mathrm{stab}}, d, \log(H/\delta)\right)^\nu$, and J and α are set as in (H.1). Taking a union bound over $h \in [H]$ and applying Theorem 6 tells us that with probability at least $1-\delta$, for all $h \in [H]$, it holds that

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim q_{\boldsymbol{\rho}_{\mathrm{m},h}}}\left[\inf_{\mu \in \mathscr{C}(DDPM(\mathbf{s}_{\hat{\theta}}, \boldsymbol{\rho}_{\mathrm{m},h}), q(\cdot|\boldsymbol{\rho}_{\mathrm{m},h}))} \mathbb{P}_{(\hat{\mathsf{a}},\mathsf{a}^*)\sim\mu}\left(||\hat{\mathsf{a}} - \mathsf{a}^*|| \geq \varepsilon/\sigma\right)\right] \leq \frac{3\varepsilon}{\sigma}.$$

Thus it holds that with probability at least $1-\delta$,

$$\sum_{h=1}^{H} \Delta_h(\varepsilon/\sigma) \leq \frac{3H\varepsilon}{\sigma}.$$

Plugging this in to Theorem 8 concludes the proof. □

Proof of Theorem 8. Lets begin by bounding $\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi)$. Recall the definitions of $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ in Section 4, and let $\mathsf{s}^\star_{1:H+1}$ and $\mathsf{s}_{1:H+1}$ denote the composite states corresponding to a trajectory $(\mathbf{x}^{\pi^\star}_{1:T+1}, \mathbf{u}^{\pi^\star}_{1:T})$ under $\pi^\star$ and $(\mathbf{x}^{\pi}_{1:T+1}, \mathbf{u}^{\pi}_{1:T})$, respectively, under the instantiation of the composite MDP in Section 4.1. We can view $\pi^\star$ and π (which depend only on memory chunks $\boldsymbol{\rho}_{\mathrm{m},h}$) as policies in the composite MDP which are compatible with the decomposition Definition E.1. We make the following points:

- In light of Lemma I.1,

$$\mathcal{L}_{\mathrm{marg},\varepsilon_1}(\pi \parallel \pi^\star) \le \Gamma_{\mathrm{marg},\varepsilon_1}(\pi \parallel \pi^\star).$$

- By Lemma I.8, a consequence of Pinsker's inequality, it holds that the Gaussian kernel W_σ used in TODA is γ_σ-TVC (w.r.t. $\mathsf{d}_{\mathrm{TVC}}$) with $\gamma_\sigma(u) = \frac{u\sqrt{\tau_{\mathrm{m}}+1}}{2\sigma}$.

- Note that $\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) = \|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\|$ measures Euclidean distance between the last $\mathbf{x}$-coordinates of $\mathsf{s}_h, \mathsf{s}'_h$. Moreover, if $\mathsf{s}'_h \sim \mathsf{W}_\sigma(\mathsf{s}_h)$ the last coordinate $\mathbf{x}'_{t_h}$ of s' is distributed as $\mathcal{N}(\mathbf{x}_{t_h}, \sigma^2 I)$. By Lemma I.7 with $d = d_x$, that for $r = 2\sigma \cdot \sqrt{5d_x + 2\log\left(\frac{1}{p}\right)}$

$$\mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}, \mathsf{s}') > r] \le p.$$

- As (a) $\mathsf{s}^\star_h$ corresponds to $\boldsymbol{\rho}_{\mathrm{c},h}$ from $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$, (b) as $\hat{\pi}, \pi^\star_{\mathrm{dec}}$ are functions of $\boldsymbol{\rho}_{\mathrm{m},h}$, and (c) by recalling the definition of $\mathsf{d}_{\mathrm{os},\varepsilon}$ in Definition 4.1,

$$\begin{aligned}&\mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}(\pi_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h)) \\ &= \mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \sim \mathcal{D}_{\mathrm{exp}}} \mathbb{E}_{\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} \sim \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2 \mathbf{I})} \inf_{\mu \in \mathscr{C}(\pi^\star_{\mathrm{dec}}(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}), \hat{\pi}(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}))} \mathbb{P}_{(\mathsf{a},\mathsf{a}') \sim \mu}[\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') \ge \varepsilon],\end{aligned}$$

  which is at most $\Delta_h(\varepsilon_0)$ by assumption.

- Finally, Proposition 4.1 ensures that under our assumption $\tau_{\mathrm{c}} \ge c_3/\eta$, and let $r_{\mathrm{IPS}} = c_4$, $\gamma_{\mathrm{IPS},1}(u) = c_5 u \exp(-\eta(\tau_{\mathrm{c}} - \tau_{\mathrm{m}})/L_{\mathrm{stab}})$, $\gamma_{\mathrm{IPS},2}(u) = c_5 u$ for c_3, c_4, c_5 given in Definition G.7. Then, for $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ as above, we have that $\pi^\star$ is $(\gamma_{\mathrm{IPS},1}, \gamma_{\mathrm{IPS},2}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-IPS.

Consequently, for $r = 2\sigma \cdot \sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)} \in (0, \frac{1}{2}r_{\mathrm{IPS}})$, Theorem 5 (which, we recall, generalizes Theorem 2 to account for the direct decomposition structure) implies

$$\begin{aligned}\mathcal{L}_{\mathrm{marg},\varepsilon+2rc_5}(\hat{\pi}_\sigma) &= \mathcal{L}_{\mathrm{marg},\varepsilon+2rc_5}(\hat{\pi}_\sigma \parallel \pi^\star) \le \Gamma_{\mathrm{marg},\varepsilon+2rc_5}(\hat{\pi}_\sigma \parallel \pi^\star) \\ &\le H\sqrt{2\tau_{\mathrm{m}} - 1}\left(\frac{\varepsilon}{2\sigma} + \frac{3}{2\sigma}\left(\max\left\{\varepsilon, 2rc_5 e^{-\frac{\eta(\tau_{\mathrm{c}}-\tau_{\mathrm{m}})}{L_{\mathrm{stab}}}}\right\}\right)\right) \\ &\quad + \textstyle\sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}(\pi_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h)) \\ &\le H\sqrt{2\tau_{\mathrm{m}} - 1}\left(\frac{2\varepsilon}{\sigma} + 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)} e^{-\frac{\eta(\tau_{\mathrm{c}}-\tau_{\mathrm{m}})}{L_{\mathrm{stab}}}}\right) + \sum_{h=1}^H \Delta_h(\varepsilon)\end{aligned}$$

Substituting in $\varepsilon_1 = \varepsilon + 2rc_5 = \varepsilon + 4c_5\sigma \cdot \sqrt{5d + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}$ the bound on $\mathcal{L}_{\mathrm{marg},\varepsilon_1}$ is proved.

To show $\mathcal{L}_{\mathrm{fin},\varepsilon_2}(\hat{\pi}_\sigma)$ satisfies the same bound, we replace $\mathsf{d}_{\mathcal{S}}$ in the above argument (as defined in Section 4.1) with $\mathsf{d}_{\mathcal{S}}(\cdot,\cdot) \leftarrow \mathsf{d}_{\mathrm{IPS}}(\cdot,\cdot)$, where again we recall that $\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_s, \mathsf{s}'_s) = \|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\|$ measures differences in the final associated control state. From Corollary G.1, which is a a generalization of Proposition 4.1, it follows that we can replace $\gamma_{\mathrm{IPS},2}(u) = c_5 u$ as used above with the considerable smaller quantity $\gamma_{\mathrm{IPS},2}(u) = c_5 u e^{-\frac{\eta\tau_{\mathrm{c}}}{L_{\mathrm{stab}}}}$. Thus, we can replace ε_1 above with $\varepsilon_2 := \varepsilon + 4c_5 e^{-\eta\tau_{\mathrm{c}}/L_{\mathrm{stab}}}\sigma \cdot (5d_x + 2\log\left(\frac{1}{\varepsilon}\right))^{1/2}$. This concludes the proof that

$$\mathcal{L}_{\mathrm{marg},\varepsilon_2}(\hat{\pi}_\sigma) \le H\sqrt{2\tau_{\mathrm{m}} - 1}\left(6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)} e^{-\frac{\eta(\tau_{\mathrm{c}}-\tau_{\mathrm{m}})}{L_{\mathrm{stab}}}} + \frac{2\varepsilon}{\sigma}\right),$$

as needed. $\square$

I.3. Imitation of the joint trajectory under total variation continuity of demonstrator policy

Here, we show that if the demonstrator policy satisfies a certain continuity property in total variation distance, then we can imitate the *joint distribution* over trajectories, not just marginals. Recall the joint imitation loss from $\mathcal{L}_{\mathrm{joint},\varepsilon}$ from Definition I.5.

Theorem 9. *Consider the setting Theorem 8, with $\Delta_h(\varepsilon)$ as in (I.2), and suppose all the assumptions of that theorem are met. Suppose that, in addition, there is a strictly increasing function $\gamma(\cdot)$ such that for all $\rho_{\mathrm{m},h}, \rho'_{\mathrm{m},h} \in \mathscr{P}_{\tau_{\mathrm{m}}-1}$,*

$$\mathsf{TV}(\pi^\star(\rho_{\mathrm{m},h}), \pi^\star(\rho'_{\mathrm{m},h})) \le \gamma(\|\rho_{\mathrm{m},h} - \rho'_{\mathrm{m},h}\|),$$

where $\pi^\star$ is defined is the conditional in Definition I.3. Further, suppose that $\mathcal{D}_{\mathrm{exp}}$ has $\tau \le \tau_{\mathrm{m}}$ bounded memory (Definition I.6). Then, with $\varepsilon_1 := \varepsilon + 4c_5\sigma \cdot \sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}$ as in (I.3),

$$\begin{aligned}\mathcal{L}_{\mathrm{joint},\varepsilon_1}(\hat{\pi}_\sigma) &\le H \cdot \textsc{ErrTvc}(\sigma,\gamma) \\ &+ H\sqrt{2\tau_{\mathrm{m}}-1}\left(\frac{2\varepsilon}{\sigma} + 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}e^{-\frac{\eta(\tau_{\mathrm{c}}-\tau_{\mathrm{m}})}{L_{\mathrm{stab}}}}\right) + \sum_{h=1}^{H}\Delta_h(\varepsilon),\end{aligned}$$

where we define $d_0 = \tau_{\mathrm{m}} d_x + (\tau_{\mathrm{m}}-1)d_u$ and $u_0 = \gamma(8\sigma\sqrt{d_0\log(9)})$, and

$$\textsc{ErrTvc}(\sigma,\gamma) = \begin{cases} 2c\sigma\sqrt{d_0} & \textit{linear } \gamma(u) = c\cdot u, c>0 \\ u_0 + \int_{u_0}^{\infty} e^{-\frac{\gamma^{-1}(u)^2}{64\sigma^2}}\mathrm{d}u & \textit{general } \gamma(\cdot)\end{cases}. \tag{I.4}$$

In particular, under Assumption 3.3, if

$$N_{exp} \ge c\left(\frac{C_\Theta dR(R\vee\sqrt{d})\log(dn)}{(\varepsilon/\sigma)^4}\right)^{4\nu} \vee \left(\frac{d^6(R^4\vee d^2\log^3\left(\frac{HndR\sigma}{\delta\varepsilon}\right))}{(\varepsilon/\sigma)^{24}}d^2\right)^{4\nu},$$

then with probability at least $1-\delta$, it holds that

$$\mathcal{L}_{\mathrm{joint},\varepsilon_1}(\hat{\pi}_\sigma) \le H\cdot\textsc{ErrTvc}(\sigma,\gamma) + H\sqrt{2\tau_{\mathrm{m}}-1}\left(\frac{3\varepsilon}{\sigma} + 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)}e^{-\frac{\eta(\tau_{\mathrm{c}}-\tau_{\mathrm{m}})}{L_{\mathrm{stab}}}}\right).$$

Remark I.3. The second term in our bound on $\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi)$ is identical to the bound in Theorem 8. The term $\textsc{ErrTvc}$ captures the additional penalty we pay to strengthen for imitation of marginals to imitation of joint distributions. Notice that if $\lim_{u\to 0}\gamma(u) \to 0$ and $\gamma(u)$ is sufficiently integrable, then, $\lim_{\sigma\to 0}\mathrm{Err}(\sigma,\gamma) = 0$. This is most clear in the linear $\gamma(\cdot)$ case, where $\mathrm{Err}(\sigma,\gamma) = \mathcal{O}(\sigma)$.

The proof is given in Appendix I.3.1; it mirrors that of Theorem 8, but replaces Theorem 2 with the following imitation guarantee in the composite MDP abstraction of Section 4, which bounds the joint imitation gap relative to $\pi^\star$ if $\pi^\star$ is TVC.

Proposition I.2. *Consider the set-up of Section 4, and suppose that the assumptions of Theorem 5, but that, in addition, the expert policy $\pi^\star$ is $\tilde{\gamma}(\cdot)$-TVC with respect to the pseudometric $\mathsf{d}_{\mathrm{TVC}}$, where $\tilde{\gamma}: \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ is strictly increasing. Then, for all parameters as in Theorem 2, and any $\tilde{r} > 0$,*

$$\begin{aligned}\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) &\le H\int_0^\infty \max_{\mathsf{s}} \mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]\mathrm{d}u \\ &+ H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\mathrm{IPS},1}(2r)\})\right) + \textstyle\sum_{h=1}^H \mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star}\mathbb{E}_{\tilde{\mathsf{s}}_h^\star\sim\mathsf{W}_\sigma(\mathsf{s}_h^\star)}\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}_h^\star) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}_h^\star)),\end{aligned}$$

where the term in color on the first line is the only term that differs from the bound in Theorem 2.

Moreover, in the special case where all of the distributions of $\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') \mid \mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})$ are stochastically dominated by a common random variable Z, and further more $\tilde{\gamma}(u) = \tilde{c}\cdot u$ for some constant $\tilde{c}$, then our bound may be simplified to

$$\begin{aligned}\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) &\le 2\tilde{c}H\mathbb{E}[Z] \\ &+ H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\mathrm{IPS},1}(2r)\})\right) + \textstyle\sum_{h=1}^H \mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star}\mathbb{E}_{\tilde{\mathsf{s}}_h^\star\sim\mathsf{W}_\sigma(\mathsf{s}_h^\star)}\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}_h^\star) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}_h^\star)).\end{aligned}$$

Proof Sketch. Proposition I.2 is derived below in Appendix I.3.2. It is corollary of Theorem 2, combined with adjoining the coupling constructed therein to a TV distance coupling between $\pi^\star_{\circlearrowleft\sigma}$ (whose joints we *can always* imitate) and $\pi^\star$. Coupling trajectories induced by $\pi^\star_{\circlearrowleft\sigma}$ and $\pi^\star$ relies on the TVC of $\pi^\star$, as well as concentration of W_σ. $\square$

Using the above proposition, we can derive the following consequences for imitation of the joint distribution.

I.3.1. Proof of Theorem 9

The proof is nearly identical to that of Theorem 8, with the modifications that we replace our use of Theorem 2 with Proposition I.2 taking $\tilde{\gamma} \leftarrow \gamma$. By Lemma I.1 and the assumpton that $\mathcal{D}_{\mathrm{exp}}$ has $\tau \leq \tau_{\mathrm{m}}$-bounded memory, it suffices to bound the joint-gap in the composite MDP:

$$\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi) \leq \Gamma_{\mathrm{joint},\varepsilon}(\pi \parallel \pi^\star).$$

We bound this directly from Proposition I.2. The final statement follows from Theorem 6 in the same way that it does in the proof of Theorem 1.

The only remaining modification, then, is to evaluate the additional additive terms colored in purple in Proposition I.2; we will show that ErrTvc as defined in (I.4) suffices as an upper bound. We have two cases. In both, let $d_0 = \tau_{\mathrm{m}} d_x + (\tau_{\mathrm{m}} - 1) d_u$. As $\mathsf{d}_{\mathrm{TVC}}$ measures the distance between the chunks $\boldsymbol{\rho}_{\mathrm{m},h} = \phi_{\mathcal{Z}}(\mathsf{s}_h)$, $\tilde{\boldsymbol{\rho}}_{\mathrm{m},h} = \phi_{\mathcal{Z}}(\mathsf{s}'_h)$, which have dimension d_0, and since we $\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\cdot) = \mathcal{N}(\cdot, \sigma^2 \mathbf{I}_{d_0})$, we have

$$\mathsf{d}_{\mathrm{TVC}}(\phi_{\mathcal{Z}} \circ \mathsf{s}, \phi_{\mathcal{Z}} \circ \mathsf{s}') \mid \mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s}) \overset{\mathrm{dist}}{=} \|\boldsymbol{\gamma}\|, \quad \boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_0}) \tag{I.5}$$

General $\gamma(\cdot)$. Eq. (I.5) and Lemma I.7 imply that

$$\mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}, \mathsf{s}')] \leq \exp(-r^2/16\sigma^2), \quad r \geq 4\sigma d_0 \log(9).$$

Hence, if $u_0 = \gamma(8\sigma d_0 \log(9))$, then

$$\mathbb{P}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}, \mathsf{s}') > \gamma^{-1}(u)/2] \leq \exp(-\gamma^{-1}(u)^2/64\sigma^2), \quad u \geq u_0.$$

Thus, as probabilities are at most one,

$$\int_0^\infty \max_{\mathsf{s}} \mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}, \mathsf{s}') > \gamma^{-1}(u)/2] \mathrm{d}u \leq u_0 + \int_{u_0}^\infty e^{-\frac{\gamma^{-1}(u)^2}{64\sigma^2}} \mathrm{d}u,$$

as needed.

Linear $\gamma(\cdot)$. In the special case where $\gamma(u) = c(u)$, Eq. (I.5) implies that we can take $Z = \|\boldsymbol{\gamma}\|$ where $\boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{d_0})$ in the second part of Proposition I.2. The corresponding additive term is then $2Hc\mathbb{E}[\|\boldsymbol{\gamma}\|]$. By Jensen's inequality, $\mathbb{E}[\|\boldsymbol{\gamma}\|] \leq \sqrt{\mathbb{E}[\|\boldsymbol{\gamma}\|^2]} = \sqrt{\sigma^2 d_0} = \sigma\sqrt{d_0}$, as needed. $\square$

I.3.2. Proof of Proposition I.2

Define the shorthand

$$B := H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon, \gamma_{\mathrm{IPS},1}(2r)\})\right) + \textstyle\sum_{h=1}^H \mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star} \mathbb{E}_{\tilde{\mathsf{s}}_h^\star \sim \mathsf{W}_\sigma(\mathsf{s}_h^\star)} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}_h^\star) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}_h^\star)),$$

and recall that Theorem 2 ensures $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) \leq B$. Further, recall from Definition 4.1 that

$$\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma}) = \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\max_{h \in [H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\circlearrowleft}_{h+1}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}^{\circlearrowleft}_h, \hat{\mathsf{a}}_h)\} > \varepsilon\right],$$

where the infinum is over all couplings μ_1 of $(\hat{\mathsf{s}}_{1:H+1}, \hat{\mathsf{a}}_{1:H}) \sim \mathsf{D}_{\hat{\pi} \circ \mathsf{W}_\sigma}$ and $(\mathsf{s}^{\circlearrowleft}_{1:H+1}, \mathsf{a}^{\circlearrowleft}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}}$ with $\mathbb{P}_{\mu_1}[\hat{\mathsf{s}}_1 = \mathsf{s}^{\circlearrowleft}_1] = 1$. For any coupling μ_1, we can consider another coupling μ_2 of $(\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H}) \sim \mathsf{D}_{\pi^\star}$ and $(\mathsf{s}^{\circlearrowleft}_{1:H+1}, \mathsf{a}^{\circlearrowleft}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}}$ with $\mathbb{P}_{\mu_2}[\mathsf{s}^\star_1 = \mathsf{s}^{\circlearrowleft}_1] = 1$. By the "gluing lemma" (Lemma C.2), we can construct a combined coupling μ which respects the marginals of μ_1 and μ_2. This combined coupling induces a joint coupling $\tilde{\mu}_1$ of $\mathsf{D}_{\hat{\pi}\circ\mathsf{W}_\sigma}$ and $\mathsf{D}_{\pi^\star}$ which, by a union bound, satisfies $\mathbb{P}_{\tilde{\mu}_1}[\hat{\mathsf{s}}_1 = \mathsf{s}_1^\star] = 1$. Thus, by a union bound, we can bound

$$\begin{aligned}\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) &\le \mathbb{P}_{\tilde{\mu}_1}\left[\max_{h\in[H]}\max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}_{h+1}^\star, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}_h^\star, \hat{\mathsf{a}}_h)\} > \varepsilon\right] \\ &\le \mathbb{P}_{\mu_1}\left[\max_{h\in[H]}\max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\circlearrowleft}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}_h^{\circlearrowleft}, \hat{\mathsf{a}}_h)\} > \varepsilon\right] \\ &\quad + \mathbb{P}_{\mu_2}\left[(\mathsf{s}_{1:H+1}^\star, \mathsf{a}_{1:H}^\star) \ne (\mathsf{s}_{1:H+1}^{\circlearrowleft}, \mathsf{a}_{1:H}^{\circlearrowleft})\right].\end{aligned}$$

Passing to the infinum over μ_1, μ_2,

$$\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) \le \underbrace{\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star_{\circlearrowleft\sigma})}_{\le B} + \inf_{\mu_2}\mathbb{P}_{\mu_2}\left[(\mathsf{s}_{1:H+1}^\star, \mathsf{a}_{1:H}^\star) \ne (\mathsf{s}_{1:H+1}^{\circlearrowleft}, \mathsf{a}_{1:H}^{\circlearrowleft})\right],$$

where again μ_2 quantify couplines of $(\mathsf{s}_{1:H+1}^\star, \mathsf{a}_{1:H}^\star) \sim \mathsf{D}_{\pi^\star}$ and $(\mathsf{s}_{1:H+1}^{\circlearrowleft}, \mathsf{a}_{1:H}^{\circlearrowleft}) \sim \mathsf{D}_{\pi^\star_{\circlearrowleft\sigma}}$ with $\mathbb{P}_{\mu_2}[\mathsf{s}_1^\star = \mathsf{s}_1^{\circlearrowleft}] = 1$. Bounding the infinum over μ_2 with Proposition I.4, we have

$$\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) \le B + \sum_{h=1}^{H}\mathbb{E}_{\mathsf{s}_h^\star}\mathsf{TV}(\pi_h^\star(\mathsf{s}_h^\star), \boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s}_h^\star))$$

To conclude, it suffices to show the following bound:

Claim I.3. *For any* $\mathsf{s} \in \mathcal{S}$, $h \in [H]$, *and* $\tilde{r} \ge 0$, $\mathsf{TV}(\pi_h^\star(\mathsf{s}), \boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s})) \le \int_0^\infty \max_{\mathsf{s}}\max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]$.

Proof. To show this claim, we note that we can represent (via the notation in Appendix E.3) $\boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s}) = \pi_h^\star\circ\mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s})$, where $\mathsf{W}^\star_{\circlearrowleft,h}$ is the replica-kernel defined in Definition E.5. Thus, we can construct a coupling of $\mathsf{a}^\star \sim \pi_h^\star(\mathsf{s})$ and $\mathsf{a}^{\circlearrowleft} \sim \boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s})$ by introducing an intermediate state $\mathsf{s}' \sim \mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s})$ and $\mathsf{a}^{\circlearrowleft} \sim \pi^\star(\mathsf{s}')$. By Lemma C.4, the fact that TV distance is bounded by one, and the assumption that $\pi^\star$ is $\tilde{\gamma}$-TVC, we then have

$$\mathsf{TV}(\pi_h^\star(\mathsf{s}), \boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s})) \le \mathbb{E}_{\mathsf{s}'\sim\mathsf{W}^\star_{\circlearrowleft,h}(\mathsf{s})}\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi_h^\star(\mathsf{s}')).$$

Recall the well-known formula that, for a non-negative random variable X, $\mathbb{E}[X] = \int_0^\infty \mathbb{P}[X > u]\mathrm{d}u$ (Durrett, 2019). From this formula, we find

$$\begin{aligned}\mathsf{TV}(\pi_h^\star(\mathsf{s}), \boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s})) &\le \int_0^\infty \mathbb{P}[\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi_h^\star(\mathsf{s}')) > u]\mathrm{d}u \\ &\overset{(i)}{\le} \int_0^\infty \mathbb{P}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)]\mathrm{d}u\end{aligned}$$

where in (i) we used that $\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi_h^\star(\mathsf{s}')) \le \tilde{\gamma}(\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}'))$ and that, as $\tilde{\gamma}(\cdot)$ is strictly increasing, we have the equality of events $\{\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi_h^\star(\mathsf{s}')) > u\} = \{\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \gamma^{-1}(u)\}$. Arguing as in the proof of Lemma E.5, we have that $\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)] \le \max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]$. Hence, we conclude

$$\mathsf{TV}(\pi_h^\star(\mathsf{s}), \boldsymbol{\pi}^\star_{\circlearrowleft\sigma,h}(\mathsf{s})) \le \int_0^\infty \max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]\mathrm{d}u$$

which proves the first guarantee. □

With the above claim proven, we conclude the proof of the first statement of Proposition I.2. For the second statement, we observe that under the stated stochastic domination assumption by Z, and if $\tilde{\gamma}(u) = \tilde{c}\cdot u$, then $\max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2] \le \mathbb{P}[Z > \frac{u}{2c}]$. Hence, by a change of variables $u = \frac{t}{2c}$,

$$\int_0^\infty \max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]\mathrm{d}u \le \int_0^\infty \mathbb{P}[Z > \frac{u}{2c}] = 2c\int_0^\infty \mathbb{P}[Z > u] = 2c\mathbb{E}[Z],$$

where again we invoke that Z must be nonnegative (to stochastically dominate non-negative random variables), and thus used the expectation formula referenced above. □

I.4. Imitation in total variation distance

Here, we notice that estimating the score in TV distance fascilliates estimation in the composite MDP, with no smoothing:

Theorem 10. *For a chunking policy $\hat{\pi}$, suppose that there are terms $(\bar{\Delta}_h)_{1\le h\le H}$ such that*

$$\bar{\mathbb{E}}_{\boldsymbol{\rho}_{\mathrm{m},h}\sim\mathcal{D}_{\mathrm{exp}}}\mathsf{TV}(\pi^\star(\boldsymbol{\rho}_{\mathrm{m},h}),\hat{\pi}(\boldsymbol{\rho}_{\mathrm{m},h}))\le\bar{\Delta}_h,$$

Then, under no additional assumption (not even those in Section 3), we have

$$\mathcal{L}_{\mathrm{fin},\varepsilon=0}(\hat{\pi})\le\mathcal{L}_{\mathrm{marg},\varepsilon=0}(\hat{\pi})\le\sum_{h=1}^{H}\bar{\Delta}_h$$

In in addition $\pi^\star$ has τ-bounded memory(Definition I.6) for $\tau\le\tau_{\mathrm{m}}$, then for $\mathcal{L}_{\mathrm{joint},\varepsilon}$ as in Definition I.5,

$$\mathcal{L}_{\mathrm{joint},\varepsilon=0}(\hat{\pi})\le\sum_{h=1}^{H}\bar{\Delta}_h$$

The above theorem is a direct consequence of the result below in the composite MDP, together with the correct instantiations for control, and Lemma I.1 to convert $\mathcal{L}_{\mathrm{marg},\varepsilon}$ and $\mathcal{L}_{\mathrm{fin},\varepsilon}$ into $\Gamma_{\mathrm{marg},\varepsilon}\le\Gamma_{\mathrm{joint},\varepsilon}$, and $\Gamma_{\mathrm{joint},\varepsilon}$, respectively.

Proposition I.4. *Consider the composite MDP setting of Section 4. Then, there exists a coupling*

$$\mathsf{TV}(\mathsf{D}_{\hat{\pi}},\mathsf{D}_{\pi^\star})\le\sum_{h=1}^{H}\mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star}\mathsf{TV}(\pi_h^\star(\mathsf{s}_h^\star),\hat{\pi}_h(\mathsf{s}_h^\star))$$

Thus, there exists a a couple $\mu\in\mathscr{C}(\mathsf{D}_{\pi^\star},\mathsf{D}_{\hat{\pi}})$ of $(\mathsf{s}^\star_{1:H+1},\mathsf{a}^\star_{1:H})\sim\mathsf{D}_{\pi^\star}$ and $(\hat{\mathsf{s}}_{1:H+1},\hat{\mathsf{a}}_{1:H})\sim\mathsf{D}_{\hat{\pi}}$ such that $\mathbb{P}_\mu[(\mathsf{s}^\star_{1:H+1},\mathsf{a}^\star_{1:H})\ne(\hat{\mathsf{s}}_{1:H+1},\hat{\mathsf{a}}_{1:H})]$ is bounded by the right-hand side of the above display. Moreover, this coupling can be constructed such that $\mathbb{P}_\mu[\mathsf{s}_1^\star=\hat{\mathsf{s}}_1]$.

Proof of Proposition I.4. This is a direct consequence of Lemma I.9, with $\mathsf{P}_1\leftarrow\mathsf{P}_{\mathrm{init}}$, and Q_{h+1} corresponding to the kernel for sampling $\mathsf{a}_h^\star\sim\pi^\star(\mathsf{s}_h^\star)$ and incrementing the dynamics $\mathsf{s}_{h+1}^\star=F_h(\mathsf{s}_h^\star,\mathsf{a}_h^\star)$, and Q_h' the same for $\hat{\mathsf{a}}_h\sim\hat{\pi}(\hat{\mathsf{s}}_h)$, and similar incrementing of the dynamics. □

I.5. Imitation with no augmentation

Theorem 11. *Let $\hat{\pi}$ be a learner policy, and define*

$$\Delta_h^\star(\varepsilon):=\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h}\sim\mathcal{D}_{\mathrm{exp}}}\mathbb{E}_{\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}\sim\mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h},\sigma^2\mathbf{I})}\inf_{\mu\in\mathscr{C}(\pi_h^\star(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}),\hat{\pi}_h(\tilde{\boldsymbol{\rho}}_{\mathrm{m},h}))}\mathbb{P}_{(\mathsf{a},\mathsf{a}')\sim\mu}[\mathsf{d}_{\mathcal{A}}(\mathsf{a},\mathsf{a}')],$$

which we note defers from $\Delta_h(\varepsilon)$ in Eq. (I.2) in that it measures error with respect to $\pi_h^\star$, rather than $\pi^\star_{\mathrm{dec},h}$. Suppose that there is a non-decreasing function $\gamma(\cdot)$ such that for all $\boldsymbol{\rho}_{\mathrm{m},h},\boldsymbol{\rho}'_{\mathrm{m},h}\in\mathscr{P}_{\tau_{\mathrm{m}}-1}$

$$\mathsf{TV}(\hat{\pi}(\boldsymbol{\rho}_{\mathrm{m},h}),\hat{\pi}(\boldsymbol{\rho}'_{\mathrm{m},h}))\le\gamma(\|\boldsymbol{\rho}_{\mathrm{m},h}-\boldsymbol{\rho}'_{\mathrm{m},h}\|),$$

where $\pi^\star$ is defined is the conditional in Definition I.3. Then, the loss of $\hat{\pi}$, without smoothing, is bounded by

$$\mathcal{L}_{\mathrm{marg},\varepsilon}(\hat{\pi})\le H\gamma(\varepsilon\sqrt{2\tau_{\mathrm{m}}-1})+\sum_{h=1}^{H}\Delta_h^\star(\varepsilon),$$

Further if $\mathcal{D}_{\mathrm{exp}}$ has $\tau\le\tau_{\mathrm{m}}$ bounded memory (Definition I.6), then it also holds that

$$\mathcal{L}_{\mathrm{joint},\varepsilon}(\hat{\pi})\le H\gamma(\varepsilon\sqrt{2\tau_{\mathrm{m}}-1})+\sum_{h=1}^{H}\Delta_h^\star(\varepsilon)$$

Proof. The above is a direct consequence of the following points. First, with our instantition of the composite MDP, we can bound $\mathcal{L}_{\mathrm{marg},\varepsilon}(\hat{\pi}) \le \Gamma_{\mathrm{marg},\varepsilon}(\hat{\pi} \parallel \pi^\star) \le \Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star)$ due to Lemma I.1; and moreover, we have $\mathcal{L}_{\mathrm{joint},\varepsilon}(\hat{\pi}) \le \Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star)$ when $\mathcal{D}_{\mathrm{exp}}$ has $\tau \le \tau_{\mathrm{m}}$-bounded memory.

Next, bounding $\|\boldsymbol{\rho}_{\mathrm{m},h} - \boldsymbol{\rho}'_{\mathrm{m},h}\| \le \sqrt{2\tau_{\mathrm{m}} - 1}\mathsf{d}_{\mathrm{TVC}}(\boldsymbol{\rho}_{\mathrm{m},h}, \boldsymbol{\rho}'_{\mathrm{m},h})$, we see $\hat{\pi}$ is $\tilde{\gamma}(\cdot)$-TVC w.r.t. $\mathsf{d}_{\mathrm{TVC}}$, where $\tilde{\gamma}(u) = \gamma(u\sqrt{2\tau_{\mathrm{m}} - 1})$. The bound now follows from Proposition D.1, and the fact that Proposition 4.1 verifies the input-stability property. □

I.6. Consequence for expected costs

Finally, we prove Proposition I.5, which shows that it is sufficient to control the imitation losses in Definition 2.2 if we wish to control the difference of a Lipschitz cost function between the learned policy and the expert distribution:

Proposition I.5. *Recall the marginal and final imitation losses in Definition 2.2, and also the joint imitation loss in Definition I.5. Consider a cost function $\mathfrak{J} : \mathscr{P}_T \to \mathbb{R}$ on trajectories $\boldsymbol{\rho}_T \in \mathscr{P}_T$. Finally, let $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$, and let $\boldsymbol{\rho}'_T \sim \mathcal{D}_\pi$ be under the distribution induced by π Then,*

(a) If $\max_{\boldsymbol{\rho}_T} |\mathfrak{J}(\boldsymbol{\rho}_T)| \le B$, and $\boldsymbol{\rho}_T$ is L Lipschitz in the Euclidean norm[9] (treating $\boldsymbol{\rho}_T$ as Euclidean vector in $\mathbb{R}^{(T+1)d_x + Td_u}$), then

$$|\mathbb{E}_{\mathcal{D}_{\mathrm{exp}}}[\mathfrak{J}(\boldsymbol{\rho}_T)] - \mathbb{E}_{\mathcal{D}_\pi}[\mathfrak{J}(\boldsymbol{\rho}'_T)]| \le \sqrt{2T}L\varepsilon + 2B\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi).$$

(b) If $\mathfrak{J}$ decomposes into a sum of of costs, $\mathfrak{J}(\boldsymbol{\rho}) = \ell_{T+1,1}(\mathbf{x}_{1+T}) + \sum_{t=1}^T \ell_{t,1}(\mathbf{x}_t) + \ell_{t,2}(\mathbf{u}_t)$, where $\ell_{t,1}(\cdot), \ell_{t,2}(\cdot)$ are L-Lipschitz and bounded in magnitude in B. Then,

$$|\mathbb{E}_{\mathcal{D}_{\mathrm{exp}}}[\mathfrak{J}(\boldsymbol{\rho}_T)] - \mathbb{E}_{\mathcal{D}_\pi}[\mathfrak{J}(\boldsymbol{\rho}'_T)]| \le 4TB\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi) + 2TL\varepsilon.$$

(c) $\mathfrak{J}(\boldsymbol{\rho}) = \ell_{T+1,1}(\mathbf{x}_{T+1})$ depends only on $\mathbf{x}_{T+1}$, then

$$|\mathbb{E}_{\mathcal{D}_{\mathrm{exp}}}[\mathfrak{J}(\boldsymbol{\rho}_T)] - \mathbb{E}_{\mathcal{D}_\pi}[\mathfrak{J}(\boldsymbol{\rho}'_T)]| \le +2B\mathcal{L}_{\mathrm{fin},\varepsilon}(\pi) + L\varepsilon$$

Thus, for our imitation guarantees to apply to most natural cost functions used in practice, it suffices to control the imitation losses defined above.

Proof of Proposition I.5. Let $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T}) \sim \mathcal{D}_{\mathrm{exp}}$, and let $\boldsymbol{\rho}'_T = (\mathbf{x}'_{1:T+1}, \mathbf{u}'_{1:T})$ be under the distribution induced by π.

Part (a). For any coupling μ between the two under which $\mathbf{x}_1 = \mathbf{x}'_1$, and let $\mathcal{E}_\varepsilon := \{\max_t \|\mathbf{x}_{t+1} - \mathbf{x}'_{t+1}\| \vee \|\mathbf{u}_t - \mathbf{u}'_t\| \le \varepsilon\}$.

$$\begin{aligned}
|\mathbb{E}[\mathfrak{J}(\boldsymbol{\rho}_T)] - \mathbb{E}[\mathfrak{J}(\boldsymbol{\rho}'_T)]| &= |\mathbb{E}_\mu[\mathfrak{J}(\boldsymbol{\rho}_T) - \mathfrak{J}(\boldsymbol{\rho}'_T)]| \\
&\le \mathbb{E}_\mu[|\mathfrak{J}(\boldsymbol{\rho}_T) - \mathfrak{J}(\boldsymbol{\rho}'_T)|] \\
&\le 2B\,\mathbb{P}_\mu[\mathcal{E}_\varepsilon^c] + \mathbb{E}_\mu[|\mathfrak{J}(\boldsymbol{\rho}_T) - \mathfrak{J}(\boldsymbol{\rho}'_T)|\mathbf{I}\{\mathcal{E}_\varepsilon\}]
\end{aligned}$$

By passing to an infinum over couplings, $\inf_\mu \mathbb{P}_\mu[\mathcal{E}_\varepsilon^c] \le \mathcal{L}_{\mathrm{joint},\varepsilon}(\pi)$. Moreover, we observe that under μ, $\mathbf{x}_1 = \mathbf{x}'_1$, and the remaining coordinates, $(\mathbf{x}_{2:T+1}, \mathbf{u}_{1:T})$ and $(\mathbf{x}'_{2:T+1}, \mathbf{u}'_{1:T})$ are the concatentation of $2T$ vectors, so the Euclidean norm of the concatenations $\|\boldsymbol{\rho}_T - \boldsymbol{\rho}'_T\|$ is at most $\sqrt{2T}\max_t \|\mathbf{x}_{t+1} - \mathbf{x}'_{t+1}\| \vee \|\mathbf{u}_t - \mathbf{u}'_t\|$, which on $\mathcal{E}_\varepsilon$ is at most $\sqrt{2T}\varepsilon$. Using Lipschitz-ness of $\mathfrak{J}$ concludes.

Part (b) . Using the adaptive discomposition of the cost and the fact that $\mathbf{x}_1$ and $\mathbf{x}'_1$ have the same distributions,

$$\begin{aligned}
|\mathbb{E}[\mathfrak{J}(\boldsymbol{\rho}_T)] - \mathbb{E}[\mathfrak{J}(\boldsymbol{\rho}'_T)]| &= |\sum_{t=1}^T (\mathbb{E}[\ell_{t,1}(\mathbf{x}_{t+1})] - \mathbb{E}[\ell_{t,1}(\mathbf{x}'_{t+1})) + (\mathbb{E}[\ell_{t,2}(\mathbf{u}_t)] - \mathbb{E}[\ell_{t,2}(\mathbf{u}'_1))| \\
&\le \sum_{t=1}^T |\mathbb{E}[\ell_{t,1}(\mathbf{x}_{t+1})] - \mathbb{E}[\ell_{t,1}(\mathbf{x}'_{t+1})| + |\mathbb{E}[\ell_{t,2}(\mathbf{u}_t)] - \mathbb{E}[\ell_{t,2}(\mathbf{u}'_1)|
\end{aligned}$$

[9] Of course, Lipschitznes in other norms can be derived, albeit with different T dependence

Applying similar arguments as in part (a) to each term, we can bound

$$\max\left\{|\mathbb{E}[\ell_{t,1}(\mathbf{x}_{t+1})] - \mathbb{E}[\ell_{t,1}(\mathbf{x}'_{t+1})|, |\mathbb{E}[\ell_{t,2}(\mathbf{u}_t)] - \mathbb{E}[\ell_{t,2}(\mathbf{u}'_1)|\right\} \le 2B\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi) + L\varepsilon.$$

Summing over the $2T$ terms concludes.

Part (c). Follows similar to part (b). □

I.7. Useful Lemmata

I.7.1. On the trajectories induced by $\pi^\star$ from $\mathcal{D}_{\mathrm{exp}}$

The key step in all of our proofs is to relate the expert distribution over trajectories $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$ to the distribution induced by the chunking policy $\pi^\star$ in Definition I.3

Lemma I.6. *There exists a sequence of probability kernels $\pi^\star_h$ mapping $\boldsymbol{\rho}_{\mathrm{m},h} \to \Delta(\mathcal{A})$ such that the chunking policy $\pi^\star = (\pi^\star_h)_{1 \le h \le H}$ satisfies the following:*

(a) $\pi^\star_h(\boldsymbol{\rho}_{\mathrm{m},h})$ is equal to the almost-sure conditional probability of a_h conditioned on $\boldsymbol{\rho}_{\mathrm{m},h}$ under $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$ and $\mathsf{a}_{1:H} = \mathtt{synth}(\boldsymbol{\rho}_T)$.

(b) The marginal distribution over each $\boldsymbol{\rho}_{\mathrm{c},h}$ is the same as the marginals of each $\mathbf{x}_t$ and $\mathbf{u}_t$ under $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$.

(c) If $\mathcal{D}_{\mathrm{exp}}$ has τ-bounded memory (Definition I.6) and if $\tau \le \tau_{\mathrm{m}}$, then the joint distribution of $\boldsymbol{\rho}_T$ induced by $\pi^\star$ is equal to the joint distribution over $\boldsymbol{\rho}_T$ under $\mathcal{D}_{\mathrm{exp}}$.

Remark I.4 (Replacing τ-bounded memory with mixing)**.** We can replace that τ-bounded memory condition to the following mixing assumption. Define the chunk $\boldsymbol{\rho}_{i \le j} = (\mathbf{x}_{i:j}, \mathbf{u}_{i:j-1})$. Define the measures

$$\begin{aligned}
\mathsf{Q}_h(\boldsymbol{\rho}_{\mathrm{m},h}) &= \mathbb{P}_{\mathsf{a}_{1:h-1},\boldsymbol{\rho}_{1:t_h-\tau_{\mathrm{m}}-1},\mathsf{a}_{h:H},\boldsymbol{\rho}_{t_h:T+1}|\boldsymbol{\rho}_{\mathrm{m},h}}\\
\mathsf{Q}^{\otimes}_h(\boldsymbol{\rho}_{\mathrm{m},h}) &= \mathbb{P}_{\mathsf{a}_{1:h-1},\boldsymbol{\rho}_{1:t_h-\tau_{\mathrm{m}}-1}|\boldsymbol{\rho}_{\mathrm{m},h}} \otimes \mathbb{P}_{\mathsf{a}_{h:H},\boldsymbol{\rho}_{t_h:T+1}|\boldsymbol{\rho}_{\mathrm{m},h}}.
\end{aligned}$$

which describes the conditional distribution of the whole trajectory without $\boldsymbol{\rho}_{\mathrm{m},h}$ and the product-distribution of the conditional distributions of the before-$\boldsymbol{\rho}_{\mathrm{m},h}$ part of the trajectory, and after $\boldsymbol{\rho}_{\mathrm{m},h}$-part. Under the condition

$$\mathbb{E}_{\boldsymbol{\rho}_{\mathrm{m},h} \text{ from } \boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}} \mathsf{TV}\left(\mathsf{Q}_h(\boldsymbol{\rho}_{\mathrm{m},h}), \mathsf{Q}^{\otimes}_h(\boldsymbol{\rho}_{\mathrm{m},h})\right) \le \varepsilon_{\mathrm{mix}}(\tau_{\mathrm{m}}),$$

which measures how close the before- and after-$\boldsymbol{\rho}_{\mathrm{m},h}$ parts of the trajectory are to being conditionally independent, one can leverage Lemma I.9 to show that

$$\mathsf{TV}(\mathcal{D}_{\pi^\star}, \mathcal{D}_{\mathrm{exp}}) \le H\varepsilon_{\mathrm{mix}}(\tau_{\mathrm{m}})$$

Lemma I.6 corresponds to the special when when $\varepsilon_{\mathrm{mix}} = 0$.

Proof of Lemma I.6. We prove each part in sequence

Part (a). follows from the fact that all random variables are in real vector spaces, and thus Polish spaces. Hence, we can invoke the existence of regular conditional probabilities by Theorem 3.

Part (b). This follows by marginalization and Markovianity of the dynamics. Specifically, let $(\boldsymbol{\rho}^\star_T, \mathsf{a}^\star_{1:H}$ be a trajectory and composite actions induced by the chunking policy $\pi^\star$, and let $(\boldsymbol{\rho}_T, \mathsf{a}_{1:H})$ be the same induced by $\mathcal{D}_{\mathrm{exp}}$. Let $\boldsymbol{\rho}^\star_{\mathrm{m},h}$ denote memory chunks of $\boldsymbol{\rho}^\star_T$, and let $\boldsymbol{\rho}_{\mathrm{m},h}$ memory chunks of $\boldsymbol{\rho}_T$ (length $\tau_{\mathrm{m}} - 1$); similarly, denote by $\boldsymbol{\rho}^\star_{\mathrm{c},h}$ and $\boldsymbol{\rho}_{\mathrm{c},h}$ the respective trajectory chunks (length $\tau_{\mathrm{c}} \ge \tau_{\mathrm{m}}$).

We argue inductively that the trajectory chunks $\boldsymbol{\rho}^\star_{\mathrm{c},h}$ and $\boldsymbol{\rho}_{\mathrm{c},h}$ are identically distributed for each h. For $h = 1$, $\boldsymbol{\rho}^\star_{\mathrm{c},1}$ and $\boldsymbol{\rho}_{\mathrm{c},1}$ are identically distributed according to $\mathcal{D}_{\mathbf{x}_1}$. Now assume we have show that $\boldsymbol{\rho}^\star_{\mathrm{c},h}$ and $\boldsymbol{\rho}_{\mathrm{c},h}$ are identically distributed. As memory chunks are sub-chunks of trajectory chunks, this means that $\boldsymbol{\rho}^\star_{\mathrm{m},h}$ and $\boldsymbol{\rho}_{\mathrm{m},h}$ are identically distributed. By part (a), it follows that $(\boldsymbol{\rho}^\star_{\mathrm{m},h}, \mathsf{a}^\star_h)$ and $(\boldsymbol{\rho}_{\mathrm{m},h}, \mathsf{a}_h)$ are identically distributed. In particular, $(\mathbf{x}^\star_{t_h}, \mathsf{a}^\star_h)$ and $(\mathbf{x}_{t_h}, \mathsf{a}_h)$ are identically distributed, where $\mathbf{x}^\star_{t_h}$ (resp $\mathbf{x}_{t_h}$) these denote the t_h-th control state under $\pi^\star$ (resp. $\mathcal{D}_{\mathrm{exp}}$). By Markovianity of the dynamics, $\boldsymbol{\rho}^\star_{\mathrm{c},h+1}$ and $\boldsymbol{\rho}_{\mathrm{c},h+1}$ are functions of $(\mathbf{x}^\star_{t_h}, \mathsf{a}^\star_h)$ and $(\mathbf{x}_{t_h}, \mathsf{a}_h)$, respectively, $\boldsymbol{\rho}^\star_{\mathrm{c},h+1}$ and $\boldsymbol{\rho}_{\mathrm{c},h+1}$ are identically distributed, as needed.

Part (c). When $\mathcal{D}_{\mathrm{exp}}$ has τ-bounded memory and $\tau \le \tau_{\mathrm{m}}$, then we have the almost-sure equality

$$\mathbb{P}_{\mathcal{D}_{\mathrm{exp}}}[\mathsf{a}_h \in \cdot \mid \mathbf{x}_{1:t_h}, \mathbf{u}_{1:t_h}] = \mathbb{P}_{\mathcal{D}_{\mathrm{exp}}}[\mathsf{a}_h \in \cdot \mid \boldsymbol{\rho}_{\mathrm{m},h}] = \pi^\star_h(\boldsymbol{\rho}_{\mathrm{m},h})[\mathsf{a}_h \in \cdot].$$

Finally, $\mathbf{x}_{t_h+1:t_{h+1}}, \mathbf{u}_{t_h:t_{h+1}-1}$ are determined by $\mathbf{x}_{t_h}$ and a_h, this inductively establishes equality of the joint-trajectory distributions.

□

I.7.2. Concentration and TVC of Gaussian Smoothing.

We now include two easy lemmata necessary for the proof. The first shows that p_r is small when r is $\Theta(\sigma)$ by elementary Gaussian concentration:

Lemma I.7. *Suppose that $\boldsymbol{\gamma} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ is a centred Gaussian vector with covariance $\sigma^2\mathbf{I}$ in $\mathbb{R}^d$ for some $\sigma > 0$. Then for all $p > 0$, it holds with probability at least $1 - p$ that*

$$||\boldsymbol{\gamma}|| \le 2\sigma \cdot \sqrt{2d\log(9) + 2\log\left(\frac{1}{p}\right)} \le 2\sigma \cdot \sqrt{5d + 2\log\left(\frac{1}{p}\right)}$$

Moreover, for $r \ge 4\sigma\sqrt{d\log(9)}$, $\mathbb{P}[||\boldsymbol{\gamma}|| \ge r] \le \exp(-r^2/16\sigma^2)$.

Proof. We apply the standard covering based argument as in, e.g., Vershynin (2018, Section 4.2). Note that

$$||\boldsymbol{\gamma}|| = \sup_{\mathbf{w} \in \mathcal{S}^{d-1}} \langle \boldsymbol{\gamma}, \mathbf{w} \rangle,$$

where $\mathcal{S}^{d-1}$ is the unit sphere in $\mathbb{R}^d$. Let $\mathcal{U}$ denote a minimal $(1/4)$-net on $\mathcal{S}^{d-1}$ and observe that a simple computation tells us that

$$\sup_{\mathbf{w} \in \mathcal{S}^{d-1}} \langle \boldsymbol{\gamma}, \mathbf{w} \rangle \le 2 \cdot \max_{\mathbf{w} \in \mathcal{U}} \langle \mathbf{w}, \boldsymbol{\gamma} \rangle.$$

A classical volume argument (see for example, Vershynin (2018, Section 4.2)) tells us that $|\mathcal{U}| \le 9^d$. A classical Gaussian tail bound tells us that for any $\mathbf{w} \in \mathcal{S}^{d-1}$, it holds that for any $r > 0$,

$$\mathbb{P}\left(\langle \mathbf{w}, \boldsymbol{\gamma} \rangle > r\right) \le e^{-\frac{r^2}{2\sigma^2}}.$$

Thus by a union bound, we have

$$\mathbb{P}\left(||\boldsymbol{\gamma}|| > r\right) \le |\mathcal{U}| \cdot \max_{\mathbf{w} \in \mathcal{U}} \mathbb{P}\left(||\boldsymbol{\gamma}|| > \frac{r}{2}\right) \le 9^d \cdot e^{-\frac{r^2}{8\sigma^2}}.$$

Inverting concludes the proof.

□

The second lemma shows that the relevant smoothing kernel is TVC:

Lemma I.8. *For any $\sigma > 0$, let $\phi_{\mathcal{Z}}$ and W_σ be as in Definition I.1 kernel, then W_σ is γ_{TVC}-TVC for with respect to $\mathsf{d}_{\mathrm{TVC}}$ (as defined in Section 4.1)*

$$\gamma_{\mathrm{TVC}}(u) = \frac{u\sqrt{2\tau_{\mathrm{m}} - 1}}{2\sigma}.$$

Proof. Recall that $\phi_{\mathcal{Z}}$ denotes projection onto the $\mathcal{Z}$-component of the direct decomposition in Definition E.1, i.e. projects onto the memory chunk $\boldsymbol{\rho}_{\mathrm{m},h}$. We apply Pinsker's inequality (Polyanskiy & Wu, 2022+): Then, for for $\mathsf{s}, \mathsf{s}' \in \mathbb{R}^p$, we have

$$\mathrm{TV}\left(\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}), \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}')\right) \le \sqrt{\frac{1}{2} \cdot \mathrm{D}_{\mathrm{KL}}\left(\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}) \parallel \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}')\right)}.$$

Note that for $\mathsf{s} = \boldsymbol{\rho}_{\mathrm{c},h}$ with corresponding memory chunk $\boldsymbol{\rho}_{\mathrm{m},h}$ $\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}) \sim \mathcal{N}(\boldsymbol{\rho}_{\mathrm{m},h}, \sigma^2 \mathbf{I})$. Similarly, for $\boldsymbol{\rho}'_{\mathrm{m},h}$ corresponding to s', $\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}') \sim \mathcal{N}(\boldsymbol{\rho}'_{\mathrm{m},h}, \sigma^2 \mathbf{I})$. Hence,

$$\mathrm{D}_{\mathrm{KL}}\left(\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}) \parallel \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}')\right) \le \frac{\left|\left|\boldsymbol{\rho}_{\mathrm{m},h} - \boldsymbol{\rho}'_{\mathrm{m},h}\right|\right|^2}{2\sigma^2}.$$

Thus, we conclude $\mathrm{TV}\left(\phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}), \phi_{\mathcal{Z}} \circ \mathsf{W}_\sigma(\mathsf{s}')\right) \le \frac{||\boldsymbol{\rho}_{\mathrm{m},h} - \boldsymbol{\rho}'_{\mathrm{m},h}||}{2\sigma}$. Finally, we upper bound the Euclidean norm $\left|\left|\boldsymbol{\rho}_{\mathrm{m},h} - \boldsymbol{\rho}'_{\mathrm{m},h}\right|\right|$ of vectors consistening of $2\tau_{\mathrm{m}} - 1$ sub-vectors via $\mathsf{d}_{\mathrm{TVC}}$ (which is the maximum Euclidean norm of these subvectors) via $\left|\left|\boldsymbol{\rho}_{\mathrm{m},h} - \boldsymbol{\rho}'_{\mathrm{m},h}\right|\right| \le \sqrt{2\tau_{\mathrm{m}} - 1}\,\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}, \mathsf{s}')$. □

I.7.3. Total Variation Telescoping

Lemma I.9 (Total Variation Telescoping)**.** *Let $\mathcal{Y}_1, \dots, \mathcal{Y}_H, \mathcal{Y}_{H+1}$ be Polish spaces. Let $\mathsf{P}_1 \in \Delta(\mathcal{Y}_1)$, and let $\mathsf{Q}_h, \mathsf{Q}'_h \in \Delta(\mathcal{Y}_h \mid \mathcal{X}, \mathcal{Y}_{1:h-1})$, $h > 1$. Define $\mathsf{P}'_1 = \mathsf{P}_1$, and recursively define*

$$\mathsf{P}_h = \mathrm{law}(\mathsf{Q}_h; \mathsf{P}_{h-1}), \quad \mathsf{P}'_h = \mathrm{law}(\mathsf{Q}'_h; \mathsf{P}'_{h-1}), \quad h > 1.$$

Then,

$$\mathsf{TV}(\mathsf{P}_{H+1}, \mathsf{P}'_{H+1}) \le \sum_{h=1}^{H} \mathbb{E}_{Y_{1:h} \sim \mathsf{P}_h} \mathsf{TV}(\mathsf{Q}_{h+1}(\cdot \mid Y_{1:h}), \mathsf{Q}'_{h+1}(\cdot \mid Y_{1:h}))$$

Moreover, there exists a coupling of $\mu \in \mathscr{C}(\mathsf{P}_{H+1}, \mathsf{P}'_{H+1})$ over $Y_{1:H+1} \sim \mathsf{P}_{H+1}$ and $Y_{1:H+1}] \sim \mathsf{P}'_{H+1}$ such that

$$\mathbb{P}_\mu[Y_1 = Y'_1] = 1, \quad \mathbb{P}_\mu[Y_{1:H+1} \ne Y'_{1:H+1}] \le \sum_{h=1}^{H} \mathbb{E}_{Y_{1:h} \sim \mathsf{P}_h} \mathsf{TV}(\mathsf{Q}_{h+1}(\cdot \mid Y_{1:h}), \mathsf{Q}'_{h+1}(\cdot \mid Y_{1:h})).$$

Proof. To prove the first part of the lemma, define $\mathsf{Q}'_{i,j}$ for $2 \le i \le j \le H+1$ by $\mathsf{Q}'_{i,i} = \mathsf{Q}_i$ define $\mathsf{Q}'_{i,j}$ by appending $\mathsf{Q}'_{i,j}$ to $\mathsf{Q}'_{i,j-1}$. and $\mathrm{law}(\mathsf{Q}'_{i,j}; (\cdot)) = \mathrm{law}(\mathsf{Q}'_j; \mathrm{law}(\mathsf{Q}_{i,j-1}; (\cdot))')$. We now define

$$\mathsf{P}^{(i)} = \mathrm{law}(Q'_{i+1,H+1}; \mathsf{P}_i),$$

with the convenction $\mathrm{law}(Q'_{H+2,H+1}; \mathsf{P}_{H+1}) = \mathsf{P}_{H+1}$. Note that $\mathsf{P}^{(H+1)} = \mathsf{P}_{H+1}$, and $\mathsf{P}^{(1)} = \mathsf{P}'_{H+1}$. Then, because TV distance is a metric,

$$\mathsf{TV}(\mathsf{P}_{H+1}, \mathsf{P}'_{H+1}) \le \sum_{h=1}^{H} \mathsf{TV}(\mathsf{P}^{(i)}, \mathsf{P}^{(i+1)})$$

Moreover, we can write $\mathsf{P}^{(i)} = \mathrm{law}(Q'_{i+2,H+1}; \mathrm{law}(\mathsf{Q}'_{i+1}; \mathsf{P}_i))$ and $P_{i+1} = \mathrm{law}(\mathsf{Q}_{i+1}; \mathsf{P}_i)$. Thus,

$$\begin{aligned}
\mathsf{TV}(\mathsf{P}^{(i)}, \mathsf{P}^{(i+1)}) &= \mathsf{TV}(\mathrm{law}(Q'_{i+2,H+1}; \mathrm{law}(\mathsf{Q}'_{i+1}; \mathsf{P}_i)), \mathrm{law}(Q'_{i+2,H+1}; \mathrm{law}(\mathsf{Q}_{i+1}; \mathsf{P}_i)) && \text{(Lemma C.4)} \\
&= \mathsf{TV}(\mathrm{law}(\mathsf{Q}'_{i+1}; \mathsf{P}_i), \mathrm{law}(\mathsf{Q}_{i+1}; \mathsf{P}_i) \\
&= \mathbb{E}_{Y_{1:i} \sim \mathsf{P}_i} \mathsf{TV}(\mathsf{Q}'_i(Y_{1:i}), \mathsf{Q}_i(Y_{1:i})). && \text{(Corollary C.1)}
\end{aligned}$$

This completes the first part of the demonstration (noting symmetry of TV). The second part follows from Corollary C.1, by letting $Y \leftarrow Y_1$, and $X \leftarrow Y_{2:H+1}$ in that lemma. □

J. Extensions and Further Results

J.1. Noisy Dynamics

We can directly extend our imitation guarantees in the composite MDP to settings with noise:

$$\mathsf{s}_{h+1} \sim F_h^{\mathrm{noise}}(\mathsf{s}_h, \mathsf{a}_h, \mathsf{w}_h), \quad \mathsf{w}_h \sim \mathsf{P}_{\mathrm{noise},h}, \tag{J.1}$$

where the noises are idependent of states and of each other. Indeed, (J.1) can be directly reduced to the no-noise setting by lifting "actions" to pairs $(\mathsf{a}_h, \mathsf{w}_h)$, and policies π to encompass their distribution of actions, and over noise.

Another approach is instead to condition on the noises $\mathsf{w}_{1:H}$ first, and treat the noise-conditioned dynamics as deterministic. Then one can take expectation over the noises and conclude. The advantage of this approach is that the couplings constructed thereby is that the trajectories experience identical sequences of noise with probability one.

Extending the control setting to incorporate noise is doable but requires more effort:

- If the *demonstrations are noiseless*, then one can still appeal to the synthesis oracle to synthesis stabilizing gains. However, one needs to (ever so slightly) generalize the proofs of the various stability properties (e.g. IPS in Proposition 4.1) to accomodate system noise.
- If the demonstrations themselves have noise, one may need to modify the synthesis oracle setup somewhat. This is because the synthesis oracle, if it synthesizes stabilizing gains, will attempt to get the learner to stabilize to a noise-perturbed trajectory. This can perhaps be modified by synthesizing controllers which stabilize to smoothed trajectories, or by collecting demonstrations of desired trajectories (e.g. position control), and stabilizing to the these states than than to actual states visited in demonstrations.

J.2. Robustness to Adversarial Perturbations

Our results can accomodate an even more general framework where there are both noises as well adversarial perturbations. We explain this generalization in the composite MDP.

Specifical, consider a space $\mathcal{E}$ of adversarial perturbations, as well as $\mathcal{W}$ of noises as above. We may posite a dynamics function $F^{\mathsf{adv}} : \mathcal{S} \times \mathcal{A} \times \mathcal{W} \times \mathcal{A} \to \mathcal{S}$, and consider the evolution of an imitator policy $\hat{\pi}$ under the adversary

$$\begin{aligned}
\hat{\mathsf{s}}_{h+1} = F_h^{\mathsf{adv}}(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h, \mathsf{w}_h, \mathsf{e}_h), \quad & \mathsf{w}_h \sim \mathsf{P}_{\mathrm{noise},h} \\
\hat{\mathsf{a}}_h \sim \hat{\pi}_h(\mathsf{s}_h) & \\
\mathsf{e}_h \sim \pi_h^{\mathsf{adv}}(\hat{\mathsf{s}}_{1:h}, \mathsf{a}_{1:h}, \mathsf{w}_{1:h}, \mathsf{e}_{1:h-1}), & \\
\hat{\mathsf{s}}_1 \sim \pi_0^{\mathsf{adv}}(\mathsf{s}_1), \quad \mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}. &
\end{aligned}$$

By constrast, we can model the demonstrator trajectory as arising from noisy, but otherwise unperturbed trajectories:

$$\mathsf{s}_{h+1}^\star \sim F_h^{\mathsf{adv}}(\mathsf{s}_h^\star, \mathsf{a}_h^\star, \mathsf{w}_h, 0), \quad \mathsf{w}_h \sim \mathsf{P}_{\mathrm{noise},h}, \quad \mathsf{a}_h^\star \sim \pi_h^\star(\mathsf{s}_h^\star), \quad \mathsf{s}_1^\star \sim \mathsf{P}_{\mathrm{init}}.$$

To reduce the composite-MDP in Section 4, we can view the combination of adverary π^{adv} and imitator $\hat{\pi}$ as a combined policy, and the $\pi^\star$ with zero augmentation as another policy; here, we would them treat actions as $\tilde{\mathsf{a}} = (\mathsf{a}, \mathsf{e})$. Then, one can consider modified senses of stability which preserve trajectory tracking, as well as a modification of $\mathsf{d}_{\mathcal{A}}$ to a function measuring distances between $\tilde{\mathsf{a}} = (\mathsf{a}, \mathsf{e})$ and $\tilde{\mathsf{a}}' = (\mathsf{a}', \mathsf{e}')$. The extension is rather mechanical, and we fit details. Note further that, by including a $\pi_0^{\mathsf{adv}}(\mathsf{s}_1)$, we can modify the analysis to allow for subtle differences in initial state distribution. This would in turn require strengthening our stability asssumptions to allow stability to initial state (e.g., the definition of incremental stability as exposited by (Pfrommer et al., 2022)).

J.3. Deconvolution Policies and Total Variation Continuity

While our strongest guarantees hold for the replica policies, where we add noise both as a data augmentation at training time *and* at test time, many practitioners have seen some success with the deconvolution policies where noise is only added at training time. We note that Proposition D.1 holds when the learned policy is TVC; without noise at training time this certainly will not hold when the expert policy is not TVC. We show here that the deconvolution expert policy is TVC under mild assumptions, which lends some credence to the empirical success of deconvolution policies.

Precisely, we show that, under reasonable conditions, deconvolution is total variation continuous. In particular, suppose that $\mu \in \Delta(\mathbb{R}^d)$ is a Borel probabilty measure and p is a density with respect to μ. Further suppose that Q is a density with respect to the Lebesgue measure on $\mathbb{R}^d$. Suppose that $\mathbf{x} \sim p$, $\mathbf{w} \sim Q$, and let $\widetilde{\mathbf{x}} = \mathbf{x} + \mathbf{w}$. We will show that the deconvolution measure $p(\mathbf{x}|\widetilde{\mathbf{x}})$ is continuous in TV.

Proposition J.1. *Let $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^d$ be fixed, let $p : \mathbb{R}^d \to \mathbb{R}$ denote a probability density, and let $Q : \mathbb{R}^d \to \mathbb{R}$ denote a function such that $\nabla^2 Q$ and $\nabla \log Q$ exist and are continuous on the set*

$$\mathcal{X} = \{(1-t)\widetilde{\mathbf{x}} + t\widetilde{\mathbf{x}}' - x | \mathbf{x} \in \operatorname{supp} p \text{ and } t \in [0,1]\}$$

Then it holds that

$$\mathrm{TV}\left(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}')\right) \leq ||\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}};|| \cdot \sup_{\mathbf{x} \in \mathcal{X}} ||\nabla \log Q(\mathbf{x})|| \,.$$

By Lemma C.4, any policy composed with the total variation kernel is thus total variation continuous with a linear γ_{TVC}; moreover, the Lipschitz constant is given by the maximal norm of the score of the noise distribution. For example, if Q is the density of a Gaussian with variance σ^2, then $\gamma_{\mathrm{TVC}}(u) \leq \frac{\sup_{\mathcal{X}} ||\mathbf{x}||}{\sigma^2}$ is dimension independent.

Remark J.1. Note that our notation is intentionally different from that in the body to emphasize that this is a general fact about abstract probability measures. We may intantiate the guarantee in the control setting of interest by letting $\mathbf{x} = \boldsymbol{\rho}_{\mathrm{m},h}$ and consider Q to be a Gaussian (for example) kernel. In this case, we see that the deconvolution policy of Definition 3.1 is automatically TVC.

To prove Proposition J.1, we begin with the following lemma:

Lemma J.2. *Let $\widetilde{\mathbf{x}} \in \mathbb{R}^d$ be fixed and suppose that $\nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x})$ exists for all $\mathbf{x} \in \operatorname{supp} p$. Then, for all $\mathbf{x} \in \operatorname{supp} p$, it holds that $\nabla_{\widetilde{\mathbf{x}}} p(\mathbf{x}|\widetilde{\mathbf{x}})$ exists. Furthermore,*

$$\int ||\nabla p(\mathbf{x}|\widetilde{\mathbf{x}})|| \, d\mu(\mathbf{x}) \leq 2 \sup_{\mathbf{x} \in \operatorname{supp} p} ||\nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x})|| \,,$$

where the gradient above is with respect to $\widetilde{\mathbf{x}}$.

Proof. We begin by noting that if $\nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x})$ exists, then so does $\nabla Q(\widetilde{\mathbf{x}} - \mathbf{x})$. By Bayes' rule,

$$p(\mathbf{x}|\widetilde{\mathbf{x}}) = \frac{p(\mathbf{x}) Q(\widetilde{\mathbf{x}} - \mathbf{x})}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')}.$$

We can then compute directly that

$$\nabla p(\mathbf{x}|\widetilde{\mathbf{x}}) = \frac{p(\mathbf{x}) \nabla Q(\widetilde{\mathbf{x}} - \mathbf{x})}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')} - \frac{p(\mathbf{x}) Q(\widetilde{\mathbf{x}} - \mathbf{x}) \cdot \int \nabla Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')}{\left(\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')\right)^2},$$

where the exchange of the gradient and the integral is justified by Lebesgue dominated convergence and the assumption of differentiability of Q and thus existence is ensured. We have now that

$$\begin{aligned}
||\nabla p(\mathbf{x}|\widetilde{\mathbf{x}})|| &= \frac{p(\mathbf{x}) Q(\widetilde{\mathbf{x}} - \mathbf{x})}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')} \cdot \left|\left| \nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x}) - \frac{\int \nabla Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')} \right|\right| \\
&= \frac{p(\mathbf{x}) Q(\widetilde{\mathbf{x}} - \mathbf{x})}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')} \cdot \left|\left| \nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x}) - \frac{\int (\nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x}')) \cdot Q(\widetilde{\mathbf{x}} - \mathbf{x}) p(\mathbf{x}') d\mu(\mathbf{x}')}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')} \right|\right| \\
&\leq \left(\sup_{\mathbf{x} \in \operatorname{supp} p} ||\nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x})|| \right) \cdot \frac{p(\mathbf{x}) Q(\widetilde{\mathbf{x}} - \mathbf{x})}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')} \cdot \left(1 + \frac{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}) p(\mathbf{x}') d\mu(\mathbf{x}')}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}) p(\mathbf{x}') d\mu(\mathbf{x}')}\right) \\
&= \left(2 \sup_{\mathbf{x} \in \operatorname{supp} p} ||\nabla \log Q(\widetilde{\mathbf{x}} - \mathbf{x})|| \right) \cdot \frac{p(\mathbf{x}) Q(\widetilde{\mathbf{x}} - \mathbf{x})}{\int Q(\widetilde{\mathbf{x}} - \mathbf{x}') p(\mathbf{x}') d\mu(\mathbf{x}')}.
\end{aligned}$$

Now, integrating over $\mathbf{x}$ makes the second factor 1, concluding the proof. $\square$

We will now make use of the theory of Dini derivatives ((Hagood & Thomson, 2006)) to prove a bound on total variation.

Lemma J.3. *For fixed $\widetilde{\mathbf{x}}, \widetilde{\mathbf{x}}'$ and $0 \le t \le 1$, let the upper Dini derivative*

$$D^+ \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)) = \limsup_{h \downarrow 0} \frac{\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))}{h},$$

where

$$\widetilde{\mathbf{x}}_t = (1-t)\widetilde{\mathbf{x}} + t\widetilde{\mathbf{x}}'.$$

If $\nabla \log Q(\widetilde{\mathbf{x}}_t - \mathbf{x})$ exists and is finite for all $\mathbf{x} \in \operatorname{supp} p$ and $t \in [0,1]$, then

$$\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}')) \le \int_0^1 D^+ \operatorname{TV}\left(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)\right) dt. \tag{J.2}$$

Proof. We compute:

$$\begin{aligned} 2\left|\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))\right| &= \left|\int |p(\mathbf{x}|\widetilde{\mathbf{x}}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h})| - |p(\mathbf{x}|\widetilde{\mathbf{x}}) - p(\widetilde{\mathbf{x}}_t)|\, d\mu(\mathbf{x})\right| \\ &\le \int |p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t)|\, d\mu(\mathbf{x}). \end{aligned} \tag{J.3}$$

Observe that by the assumption on Q and Lemma J.2, $p(\mathbf{x}|\widetilde{\mathbf{x}}_t)$ is differentiable and thus continuous in $\widetilde{\mathbf{x}}_t$. We therefor see that the function

$$t \mapsto \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))$$

is continuous as $\widetilde{\mathbf{x}}_t$ is linear in t. By Hagood & Thomson (2006, Theorem 10), (J.2) holds. □

We now bound the Dini derivatives:

Lemma J.4. *Let $\widetilde{\mathbf{x}}, \widetilde{\mathbf{x}}' \in \mathbb{R}^d$ such that for all $t \in [0,1]$it holds that*

$$\sup_{\mathbf{x} \in \operatorname{supp} p} \left|\frac{d^2}{dt^2}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right)\right| = C < \infty,$$

where the derivative is applied on $\widetilde{\mathbf{x}}_t$. If the assumptions of Lemmas J.2 and J.4 hold, then

$$D^+ \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)) \le ||\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}'|| \cdot \sup_{\substack{\mathbf{x} \in \operatorname{supp} p \\ t \in [0,1]}} ||\nabla \log Q(\widetilde{\mathbf{x}}_t - x)||\,.$$

Proof. By definition,

$$D^+ \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)) = \limsup_{h \downarrow 0} \frac{\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))}{h}.$$

Fix some t and some small h. By (J.3), it holds that

$$\left|\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))\right| \le \frac{1}{2} \cdot \int |p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t)|\, d\mu(\mathbf{x}).$$

By Taylor's theorem, it holds that

$$p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t) = h \cdot \frac{d}{dt}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right) + h^2 \cdot \frac{d^2}{dt^2}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_{t'})\right)$$

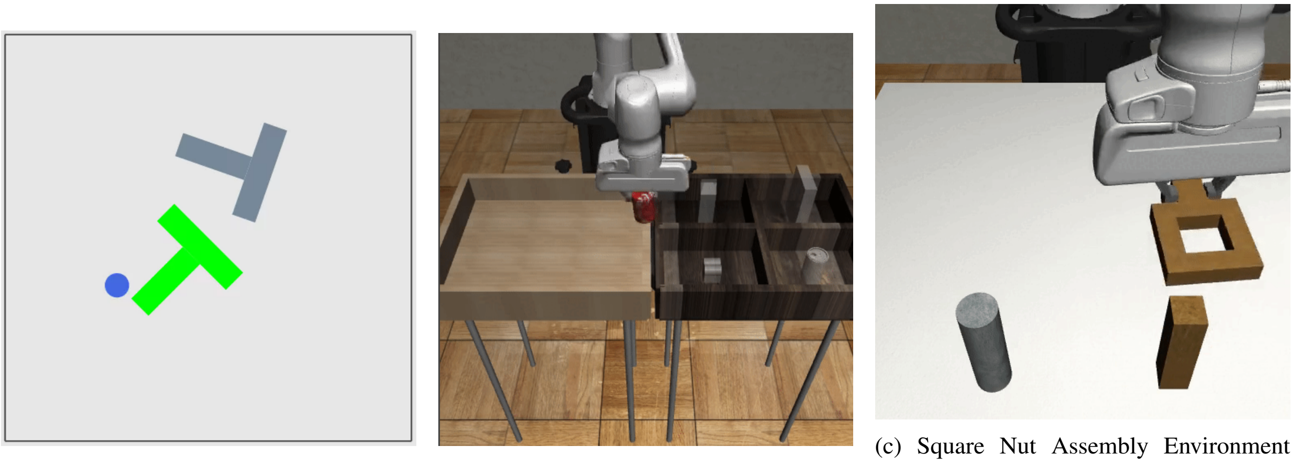
for some $t' \in [0,1]$. By the chain rule, we have

$$\frac{d}{dt}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right) = \left\langle \widetilde{\mathbf{x}}' - \widetilde{\mathbf{x}}, \nabla p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right\rangle,$$

and thus,

$$|p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t)| \le h \cdot ||\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}'|| \cdot ||\nabla p(\mathbf{x}|\widetilde{\mathbf{x}}_t)|| + h^2 C$$

Now, applying Lemma J.2 and plugging into the previous computation concludes the proof. □



(a) PushT Enviroment (Chi et al., 2023). The blue circle is the manipulation agent, while the green area is the target position which the agent must push the blue T block into.

(b) Can Pick-and-Place Environment (Mandlekar et al., 2021). The grasper must pick up a can from the left bin and place it into the correct bin on the right side.

(c) Square Nut Assembly Environment (Mandlekar et al., 2021). The grasper must pick up the square nut (the position of which is randomized) and place it over the square peg.

Figure 6. Environment Visualizations.

We are finally ready to state and prove our main result:

Proof of Proposition J.1. Note that

$$\frac{d^2}{dt^2}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right) = \left(\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}'\right)^T \nabla^2 p(\mathbf{x}|\widetilde{\mathbf{x}}_t)(\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}')$$

and thus is bounded if and only if $\nabla^2 p(\mathbf{x}|\widetilde{\mathbf{x}}_t)$ is bounded. An elementary computation shows that if $\nabla^2 Q$ exists and is continuous on $\mathcal{X}$, then $\nabla^2 p(\mathbf{x}|\widetilde{\mathbf{x}}_t)$ is bounded in operator norm on $\mathcal{X}$. Thus the assumption in Lemma J.4 holds. Applying Lemma J.3 then concludes the proof. □

K. Experiment Details

K.1. Compute and Codebase Details

Code. For our experiments we build on the existing PyTorch-based codebase and standard environment set provided by Chi et al. (2023) as well as the robomimic demonstration dataset Mandlekar et al. (2021). [10]

Compute. We ran all experiments using 4 Nvidia V100 GPUs on an internal cluster node. For each environment running all experiments depicted in Figure 2 took 12 hours to complete with 20 workers running simultaneously for a total of approximately 10 days worth of compute-hours. Between all 20 workers, peak system RAM consumption totaled about 500 GB.

K.2. Environment Details

For simplicity the stabilizatin oracle `synth` is built into the environment so that the diffusion policy effectively only performs positional control. See Appendix K for visualizations of the environments.

PushT. The PushT environment introduced in (Chi et al., 2023) is a 2D manipulation problem simulated using the PyMunk physics engine. It consists of pushing a T-shaped block from a randomized start position into a target position using a controllable circular agent. The synthesis oracle runs a low-level feedback controller at a 10 times higher to stabilize the

[10] The modified codebase with instructions for running the experiments is available at the following anonymous link: `https://www.dropbox.com/s/vzw0gvk1fd3yadw/diffusion_policy.zip?dl=0`. We will provide a public github repository for the final release.

agent's position towards a desired target position at each point in time via acceleration control. Similar to Chi et al. (2023), we use a position-error gain of $k_p = 100$ and velocity-error gain of $k_v = 20$. The observation provided to the DDPM model consists of the x,y oordinates of 9 keypoints on the T block in addition to the x,y coordinates of the manipulation agent, for a total observation dimensionality of 20.

For rollouts on this environment we used trajectories of length $T = 300$. Policies were scored based on the maximum coverage between the goal area and the current block position, with > 95 percent coverage considered an "successful" (score = 1) demonstration and the score linearly interpolating between 0 and 1 for less coverage. A total of 206 human demonstrations were collected, out of which we use a subset of 90 for training.

Can Pick-and-Place. This environment is based on the Robomimic (Mandlekar et al., 2021) project, which in turn uses the MuJoCo physics simulator. For the low-level control synthesis we use the feedback controller provided by the Robomimic package. The position-control action space is 7 dimensional, including the desired end manipulator position, rotation, and gripper position, while the observation space includes the object pose, rotation in addition to position and rotation of all linkages for a total of 23 dimensions. Demonstrations are given a score of 1 if they successfully complete the pick-and-place task and a score of 0 otherwise. We roll out 400 timesteps during evaluation and for training use a subset of up to 90 of the 200 "proficient human" demonstrations provided.

Square Nut Assembly. For Square Nut Assembly, which is also Robomimic-based (Mandlekar et al., 2021), we use the same setup as the Can Pick and Place task in terms of training data, demonstration scoring, and low-level positional controller. The observation, action spaces are also equivalent to the Can Pick-and-Place task with 23 and 7 dimensions respectively.

K.3. DDPM Model and Training Details.

For our DDPM we use the same 1-D convolutional UNet-style (Ronneberger et al., 2015) architecture employed by (Chi et al., 2023), which is in turn adapted from Janner et al. (2022). This principally consists of 3 sets of downsampling 1-dimensional convolution operations using Mish activation functions (Misra, 2019), Group Normalization (with 8 groups) (Wu & He, 2018), and skip connections with 64, 128, and 256 channels followed by transposed convolutions and activations in the reversed order. The observation and timestep were provided to the model with Feature-wise Linear Modulation (FiLM) (Perez et al., 2018), with the timestep encoded using sin-positional encoding into a 64 dimensional vector.

During training and evaluation we utilize a squared cosine noise schedule (Nichol & Dhariwal, 2021) with 100 timesteps. For training we use the AdamW optimizer with linear warmup of 500 steps, followed by an initial learning rate of 1×10^{-4} combined with cosine learning rate decay over the rest of the training horizon. For PushT models we train for 800 epochs and evaluate test trajectories every 200 epochs while for Can Pick-and-Place and Square Nut Assembly we evaluate performance every 250 epochs and train for a total of 1500 epochs.

In both environments the diffusion models are conditioned on the previous two observations trained to predict a sequence of 16 target manipulator positions, starting at the first timestep in the conditional observation sequence. The 2rd (corresponding to the target position for the current timestep) through 9th generated actions are emitted as the $\tau_c = 8$ length action sequence and the rest is discarded. Extracting a subsequence of a longer prediction horizon in this manner has been shown to improve performance over just predicting the $H = 8$ action sequence directly (Chi et al., 2023).

For $\sigma > 0$ we generate new perturbed observations per training iteration, effectively using $N_{\text{aug}} = N_{\text{epoch}}$ augmentations. We find this to be easier than generating and storing N_{aug} augmentations with little impact on the training and validation error. Noise is injected after the observations have been normalized such that all components lie within $[-1, 1]$ range. Performing noise injection post normalization ensures that the magnitude of noise injected is not affected by different units or magnitudes. .