Contents

Part I

General Results and Discussion

A Notation and Organization of Appendix

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_{chunk} and observation length τ_{obs} satisfying $0 \le \tau_{\mathrm{obs}} \le \tau_{\mathrm{chunk}}$. We recall the definition of the trajectory-chunk s_h and observation-chunk in o_h in Section 2, which introduced the indexing h, such that $t_h = (h-1)\tau_{\mathrm{chunk}} + 1$. Recall also the composite actions $\mathsf{a}_h = (\kappa_{t_h:t_{h+1}-1}) \in \mathcal{A} = \mathcal{K}^{\tau_{\mathrm{chunk}}}$ as the concatenation of τ_{chunk} 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 D.1. We refer the reader to Appendix D for the relevant terminology, and to Appendix D.2 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 E, we generalize some of our results to accomodate general incrementally stabilizing primitive controllers. In Appendix E.2, 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 D.5 with those found in prior work.

After the preliminaries on organization, notation, and related work, we divide our appendices into three parts.

Part I: General Results and Discussion In this section, we describe major results and discussion omitted from the main body in the interest of space. This present appendix contains notation and organization. Subsequently, Appendix B provides a comprehensive discussion of related work. Appendix C provides discussion for Theorem 1, a proof sketch of Theorem 2, and the requisite assumptions and formal statement of our guarantee for HINT. Appendix D provides a detailed overview of the analysis, and Appendix E extends our results from affine primitive controllers to general ones.

Part II: The Composite MDP. In the first part of the Appendix, we expand on and provide rigorous proofs of statements and results pertaining to the composite MDP as considered in Appendix D. We begin by providing a detailed background in Appendix F 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 G, we provide the full proof of Proposition D.2, as warm-up to the proof of Theorem 4. In particular, the argument substantially simplifies if we consider the case of no added augmentation (when $\sigma = 0$ in HINT) 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 H, where we rigorously prove a generalization of Theorem 4 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 I, which demonstrate the tightness of our arguments in Appendix H.

Part III: Instantiations We continue our appendices in the second part, which is concerned with the instantiation of the composite MDP in with incremental stability. The heart of the second part of our appendices is Appendix J, which provides the final, end-to-end guarantees and the proof of Theorem 1 and Theorems 2 and 3. We also provide a number of variations on this result, including stronger guarantees on imitation of the joint trajectories (Appendix J.4), guarantees on HINT under the aassumption that sampling is close in total variation (Appendix J.5). We also show in Proposition J.5 that most natural cost funtions have similar expected values on imitator and demonstrator trajectories assuming that the imitation losses are small. In Appendix K, 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 D.1, and Appendix K.6 in particular explains how to synthesize stabilizing gains, as assumed in Section 2. With the stability properties thus proven, we proceed in Appendix L to instantiate our conditional sampling guarantees with DDPMs. In particular, by applying earlier work, we state and prove Theorem 13, 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 L.4 why stronger total variation guarantees on sampling are unrealistic in our setting.

Appendix M generalizes the proofs in Appendix J to the generic primitive controllers considered in Appendix E. We then provide a number of extensions of our main results in Appendix N, including to the setting of noisy dynamics (Appendix N.2). Finally, in Appendix O, 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 [30, 15, 10], visuomotor policies [23, 78], and navigation tasks [31]. 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 [57]. This distribution shift has been shown to result in the imitator making incorrect judgements regarding observation-action causality, often with catastrophic consequences [21]. 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 [58, 57, 40]. Among this class of methods, the DAgger algorithm in particular has seen widespread adoption [57, 65, 37]. 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 [16, 50]. Our work is similarly focused on the offline setting, but is capable of handling stochastic, non-Markovian demonstrators. Unlike [50], 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 and other Generative Approaches. Denoising Diffusion Probabilistic Models (DDPMs) [60, 29] and their variant, Annealed Langevin Sampling [62], have seen enourmous empirical success in recent years, especially in state-of-the-art image generation [55, 47, 61]. 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. [33], Chi et al. [19], Pearce et al. [48], Hansen-Estruch et al. [26]. DDPMs rely on learning the score function of the target distribution, which is generally accomplished through some kind of denoised estimation [32, 74, 63]. 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 [53, 13, 14], 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. [18], Lee et al. [41] 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.

Recent work has also shown that transformer architectures [80, 17, 59] can also serve as probabilistic models for predicting sequences of robotic actions, and can represent multi-modality to varying degrees. Notably, these approaches also rely heavily on the action-chunking which we consider in this paper [80].

Hierarchical Planning. Hierarchy has long been applied in robotic learning and planning to abstract away low-level primitives. Task-and-motion planning (TAMP) [35] reduces robotic motion represents planning via sequences of discrete primitives — a "mode sequence" — constrained to which the optimization problem is continous. LQR trees [66] proposes using linear quadratic regulator (LQR) trees to efficiently cover a control state space with local feedback laws so as to compute a motion plan that reaches a desired goal or behavior, subject to stability guarantees. Graph of Convex Sets (GCS), a more recent innovation, decomposes constraint sets into convex regions, each of which represents nodes in a planning graph [43]. More recent work has used hierarchy to leverage the power of large learned models for solving tasks that contain multiple types of data inputs[4] [4], and as modules for generating multiple forms of supervision [79].

[4]These tasks are also called multimodal, where modes here refer to types of data. In distinction, multi-modality in this paper refers to multiple modes within the distribution over expert demonstrations

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. [40]), 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 [36]. Our work gives an interpretation of this decision as enforcing that the learned policy obey the distributional continuity property we term TVC (Definition D.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 [39], adaptive meta-learning [3], in more traditional supervised learning as well [28].

C Further Main Results and Discussion

We begin with a general remark.

Remark C.1 (Do we need state observation or time-varying policies?)**.** In practical applications, behavior cloning policies respond not to measurements of physical system state, but rather visual observations and/or tactile feedback. Additionally, learned policies to not explicit take a time index h as input. This allows these policies to perform flexibily across tasks with varying time horizons, and to automatically reset after encountered obstacles.

In contrast, our formulation requires policies to be (a) functions of system state and (b) vary with the chunk index h. If visual observations or tactile measurements are *sufficient* to recover system state, then we can view these data are redundant states, and thus (a) is not a restriction. Moreover, as described in Remark C.6, a limited portion of theoretical results do hold for policies which do not vary with h. In general, restrictions (a) and (b) are necessary for our analysis because they allow us to analyze the imitator behavior in a Markovian fashion. Without these restrictions, one would have to reason about (a′) uncertainty over state given observation, or (b′) variation in expert behavior across different time steps h. Removing these restrictions is an exciting direction for future work.

C.1 Discussion surround Theorem 1

Remark C.2 (The $\varepsilon = 0$ case)**.** If we were able to bound the policy error $\Delta_{(\varepsilon)}$ with $\varepsilon = 0$ – which corresponds to estimating $\mathsf{a}_h \mid \mathsf{o}_h$ in *total variation* distance – the imitation learning problem would be trivialized, and neither the TVC condition above or the noise-injection based smoothing in the section below would not be needed (see Appendix J). Appendix L 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.

Remark C.3 (On the TVC assumption)**.** It is true than any $\hat{\pi}$ implemented as a DDPM with a Lipschitz activation with bounded-magnitude parameters is indeed TVC. Unfortunately, these Lipschitz constants can be too large to be meaningful in practical scenarios, scaling exponentially with network depth. In addition, the absence of smoothing σ may make the corresponding DDPM learning problem more challenging. Hence, in what follows, we shall require the additional sophistication of smoothing with Gaussian noise of variance $\sigma^2 > 0$ for meaningful guarantees.

Furthermore, we show in Appendix G.1 that the TVC assumption, which measures total variation distance between nearby $\hat{\pi}_h(\cdot)$ at nearby observations, can be relaxed to variant which measures the probability (under a minimal coupling) that actions differ by some tolerance. However, this tolerance has to be quite small, and as we argue, any reasonable notion of Wasserstein continuity is unlikely to suffice.

Remark C.4 (Imitation of the joint distribution)**.** Suppose the expert distribution $\mathcal{D}_{\mathrm{exp}}$ has at most τ_{obs}-bounded memory (defined formally in Definition J.5). Then $\mathcal{L}_{\mathrm{joint},\varepsilon}(\hat{\pi})$ satisfies the same upper bound (E.2), where $\mathcal{L}_{\mathrm{joint},\varepsilon}(\hat{\pi})$, formally defined in Definition J.4, measures an optimal transport distance between the *joint distribution* of the expert trajectory and the one induced by $\hat{\pi}$.

Remark C.5 (Is chunking necessary?)**.** In Appendix N.1, we show that we can remove the required lower bound on τ_{chunk} — allowing, in particular, the choice of $\tau_{\mathrm{chunk}} = 1$ — under the slightly stronger condition that our synthesis oracle ensures that the entire sequence of primitive controllers $\kappa_{1:T}$ on the whole horizon T are incrementally stabilizing. However, chunking is known to yield empirical benefits [80], and training models to predict action-chunks of longer duration than the agent acts on is also observed to improve performance [19].

Remark C.6 (Are time-varying policies necessary? (continuing Remark C.1))**.** In practice, time-invariant policies $\hat{\pi}$ which do depend on the h-index are preferred because they are more resilient to varying-horizon tasks, and can automatically "reset" when they encounter an obstacle. Here, we note that if $\pi_h^\star(\mathsf{o}_h)$, the conditional distribution of $\pi_h^\star$ given o_h, is independent of h — that is, the expert is Markov and time-invariant given o_h — then the term $\Delta_{(\varepsilon/c_1)}\left(\pi_h^\star(\mathsf{o}_h), \hat{\pi}_h(\mathsf{o}_h)\right)$ on the right-hand side of (E.2) can be made small by choosing a time-invariant $\hat{\pi}$. Thus, certain expert behavior can indeed be imitated by time-invariant policies. However, we do require time-varying policies to imitate *arbitrary experts*. And, in addition, the data smoothing strategy described below requires a time-varying $\hat{\pi}$. Extending our results to time-invariant $\hat{\pi}$ is an interesting direction for future inquiry, and we suspect that this may require some further notion of cost-to-go to made the formulation feasible.

C.2 Proof sketch of Theorem 2

As with Theorem 1, the key ideas of the proof are given Appendix D, expressed in terms of a general abstraction for behavior cloning we call the "composite MDP". This template is instantiated with a details in Appendix J. Moreso than Theorem 1, the proof of Theorem 2 requires sophisticated couplings between expert and learner trajectories, and in particular. The intuition is based on the observation that $\hat{\pi}_{\sigma,h}$ mimic $\pi^\star_{\mathrm{rep},\sigma,h} := (\pi^\star_{\mathrm{dec},\sigma,h})_\sigma$, the smoothing of the deconvolution policy. Inspired by replica pairs in statistical physics, we call $\pi^\star_{\mathrm{rep},\sigma}$ the "replica" policy because actions from $\pi^\star_{\mathrm{rep},\sigma,h}$ can be thought of as actions from $\pi_h^\star$ that have been noised and deconvolved. This implies:

Fact C.1. Let $\mathsf{o}_h \sim \mathcal{D}_{\mathrm{exp},h}$. Then, the distributions of $\mathsf{a}_h \sim \pi_h^\star(\mathsf{o}_h)$, and $\mathsf{a}_h' \sim \pi^\star_{\mathrm{rep},\sigma,h}(\mathsf{o}_h)$, marginalized over o_h, are identical.

This observation can be interpreted as meaning that smoothing and deconvolution are inverse operations at the distributional level. In particular, for a moment, consider an idealized environment where a_h, and not o_h, perfectly determined the dynamics (e.g. by teleportation), then $\pi^\star_{\mathrm{rep},\sigma} = (\pi^\star_{\mathrm{rep},\sigma,h})$ and $\pi^\star = (\pi_h^\star)$ would induce the same dynamics (and, as remarked above, $\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi^\star) = 0$).

For non-idealized environments, the argument goes as follows: We couple $\mathsf{a}_h \sim \pi_h^\star(\mathsf{o}_h)$, and $\mathsf{a}_h' \sim \pi^\star_{\mathrm{rep},\sigma,h}(\mathsf{o}_h)$ so that a_h' has the distribution of $\mathsf{a}_h' \sim \pi^\star(\mathsf{o}_h')$, where o_h' is distributed as o_h and, with high probability, o_h' and o_h are $\tilde{O}(\sigma)$-close in Euclidean norm. This coupling is depicted Figure 4. We then argue that the dynamics induced by $\pi^\star_{\mathrm{rep},\sigma}$ track those of $\pi^\star$ by roughly a similar margin. Moreover, by smoothing $\hat{\pi}_h$ and $\pi^\star_{\mathrm{dec},h}$ and applying Jensen's inequality,

$$\mathbb{E}_{\mathsf{o}_h \sim \mathcal{D}_{\mathrm{exp},h}}\left[\Delta_{(\varepsilon^2)}\left(\pi^\star_{\mathrm{rep},\sigma,h}(\mathsf{o}_h), \hat{\pi}_{\sigma,h}(\mathsf{o}_h)\right)\right] \le \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}}\left[\Delta_{(\varepsilon^2)}\left(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)\right)\right].$$

Consequently, when the right hand side of the above inequality is small, the noised policy $\hat{\pi}_\sigma$ tracks the replica policy $\pi^\star_{\mathrm{rep},\sigma}$, which we have shown to be track $\pi^\star$ (and thus track $\mathcal{D}_{\mathrm{exp}}$). □

C.3 Merits and Drawbacks of the Synthesis Oracle

The role of a synthesis oracle satisfying Assumption 3.1 is to replace a strong assumption on the stability of an *expert demonstration* with an *algorithmic assumption* that allows post-hoc stabilizing of expert demonstrations. This approach presents a natural question: in what sense is this tradeoff a sensible one? To answer this, consider a paradigmatic case, where the demonstrations solve some complicated task in a smooth, nonlinear control system. Suppose further that the one-step dynamics of the system are known, but that the expert demonstrations come from some optimal control law which is computationally prohibitive to compute, or possibly even some mixture of different, mutually-incompatible trajectories. Assuming the Jacobian linearizations of the nonlinear system are stabilizable (see Appendix K for further details), one can implement a synthesis oracle for affine gains directly by solving a Ricatti recursion on the Jacobian-linearized dynamics around each expert trajectory. . Conceptually, this has the following interpretation: **Our framework reduces the problem of imitating a complex expert trajectory to (i) supervised generative modeling and (ii) solving strictly *local* control problems**.

That is, we offload complex behavior of the expert being imitated, and reduce the learner's burden to solving local control problems that are significantly simpler than global planning. For more general systems, Appendix E addresses possibly non-affine stabilizing gains, and discusses how these may arise from standard practices of using robotic position control or inverse dynamics. Appendix E.2 compares our hierarchical approach to stability to standard formulations that apply to the expert distribution, and we show how the latter rule out the possibility for complex behaviors such as bifurcated trajectories.

Limitations and Future Directions. Our above example required access to differentiable (indeed, smooth) system dynamics. Stabilizing systems with contact dynamics remains an outstanding challenge. More generally, an overtly hierarchical approach may be inefficient for many reasons, notably (1) the dimension of the primitive controller may be much higher than the dimension of raw control inputs; and (2) when the high-level and low-level controllers are parametrized by the neural networks, explicit heirarchy with separate models may preclude shared representation learning. 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. Nevertheless, we think that **our conceptual contribution of decoupling low-level stability and generative matching of demonstrated behavior will prove useful in future endeavors for reliable and performant behavior cloning.**

C.4 Formal Assumptions for Analysis of HINT

We now state the assumptions required for our theoretical guarantees on HINT. We require access to a class of score functions rich enough to represent the following deconvolution conditionals. To define these, we introduce the following distribution

Definition C.1. Recall the policy $\pi^\star_{\mathrm{dec},\sigma,h}$ from Definition 3.6. Given $\mathsf{o}_h \in \mathcal{O}$, let $\pi^\star_{\mathrm{dec},\sigma,h,[t]}(\mathsf{o}_h) \in \Delta(\mathcal{A})$ denote the law of $\mathsf{a}_{h,[t]} := e^{-t}\mathsf{a}_h^{(0)} + \sqrt{1-e^{-2t}}\boldsymbol{\gamma}$, where $\mathsf{a}_h \sim \pi^\star_{\mathrm{dec},\sigma,h}(\mathsf{o}_h)$, and $\boldsymbol{\gamma} \sim \mathcal{N}(0, \mathbf{I})$ is independent whte Gaussian noise. In words, $\mathsf{a}_{h,[t]}$ is generated from the Ornstein–Uhlenbeck process that interpolates a_h with white noise. To avoid notational clutter, we supress the dependence of $\pi^\star_{\mathrm{dec},\sigma,h,[t]}$ on σ via $\pi^\star_{\mathrm{dec},h,[t]} = \pi^\star_{\mathrm{dec},\sigma,h,[t]}$ when σ is clear from context.

Next, we need an assumption of bounded statistical complexity. We opt for the popular *Rademacher complexity* [11]. In defining this quantity, recall that scores are vector-valued, necessitating the vector analogue of Rademacher complexity [44, 24], studied for score matching in [13].

Definition C.2 (Function Class Θ and Rademacher Complexity)**.** Consider a class of score functions of the form $\Theta = \{(\mathbf{s}_{\theta,h})_{1\le h\le H} | \theta \in \Theta\}$, where $\mathbf{s}_{\theta,h}$ maps triples $(\mathsf{a}, \mathsf{o}_h, j)$ of composite actions $\mathsf{a} \in \mathcal{A}$, observation-chunks o_h, and DDPM-steps $j \in \mathbb{N}$ to vectors in $\mathbb{R}^{d_\mathcal{A}}$. For each chunk $h \in [H]$, DDPM-step $j \in \mathbb{N}$ and discretization size α, define the vector- Rademacher complexity of Θ as

$$\mathcal{R}_{n,h,j}(\Theta; \alpha) := \mathbb{E}\left[\sup_{\theta\in\Theta} \frac{1}{n}\sum_{k=1}^{n} \left\langle \boldsymbol{\varepsilon}^{(k)}, \mathsf{s}_{\theta,h}\left(\mathsf{a}^{(k)}_{h,[j\alpha]}, \tilde{\mathsf{o}}^{(k)}_h, j\right)\right\rangle\right],$$

where $\boldsymbol{\varepsilon}_k \in \mathbb{R}^d$ are i.i.d. random vectors with Rademacher coordinates, and where $(\tilde{\mathsf{o}}_h^{(k)}, \mathsf{a}^{(k)}_{h,[j\alpha]})$ are i.i.d. samples from $\mathcal{D}_{\sigma,h,[j\alpha]}$.

Assumption C.1. We suppose that, for any $\sigma > 0$, we are given a class of score functions $\Theta = \Theta(\tau_{\mathrm{chunk}}, \tau_{\mathrm{obs}}, \sigma, \alpha)$ of the form in Definition C.2 which satisfies the following conditions:

(a) *Realizability:* there exists a $\theta_\star$ such that, for all $h \in [H]$ and $j \in \mathbb{N}$, $\mathsf{s}_{\theta_\star,h}(\mathsf{a}, \mathsf{o}_h, j)$ is the score function of $\pi^\star_{\mathrm{dec},h,[j\alpha]}(\mathsf{o}_h)$ at $\mathsf{a} \in \mathcal{A}$.

(b) *The Rademacher complexity of Θ has polynomial decay in n and growth in α:*

$$\sup_{j\in\mathbb{N}} \max_{h\in[H]} \mathcal{R}_{n,h,j}(\Theta; \alpha) \le C_\Theta \alpha^{-1} n^{-\frac{1}{\nu}},$$

where $C_\Theta = C_\Theta(\sigma, \tau_{\mathrm{chunk}}, \tau_{\mathrm{obs}}) > 0$.

(c) The scores have *linear growth*; that is, there exists some $C_{\mathrm{grow}} = C_{\mathrm{grow}}(\sigma, \tau_{\mathrm{chunk}}, \tau_{\mathrm{obs}}) > 0$ sucht hat

$$\sup_{j\in\mathbb{N}, h\in[H]} \sup_{\theta\in\Theta} ||\mathsf{s}_\theta(\mathsf{a}, \tilde{\mathsf{o}}_h, j)|| \le C_{\mathrm{grow}} \alpha^{-1}(1 + ||\mathsf{a}|| + ||\tilde{\mathsf{o}}_h||),$$

As discussed in Appendix L, generalizing to polynomial growth is straightforward.

As justified in Appendix L, our decay condition on the Rademacher complexity is natural for statistical learning, and holds for most common function classes (often with $\nu \leq 2$ and even more benign dependence on J, α); our results can easily extend to approximate realizability as well. Our Rademacher bound depends implicitly on chunk and observation lengths $\tau_{\text{chunk}}, \tau_{\text{obs}} > 0$ and implicitly on dimension $d_{\mathcal{A}}$ via C_Θ. Realizability is motivated by the approximation power of deep neural networks [12]. Lastly, we do expect realizability to hold uniformly over $j \geq 0$ because, as $j \to 0$, the corresponding scores corresponds to a scaled identity function (i.e. the score of a standard Gaussian).

C.5 Theoretical Guarantee for HINT.

We now state our guarantee for HINT, invoking assumptions from the section above. Recall that $d_{\mathcal{A}}$ denotes the dimension of composite actions, and that $c_1, \ldots, c_5$ are as in Definition 3.2.

Theorem 3. Suppose Assumption 3.1 holds. Let $c_1, \ldots, c_5 > 0$, defined in Definition 3.2, and let $\Theta_{\text{Iss}}(x)$ denote a term which is upper and lower bounded by a x times a polynomial in those constants and their inverses. Let $\varepsilon \leq \Theta_{\text{Iss}}(1)$, if we choose $\sigma = \varepsilon/\Theta_{\text{Iss}}(\sqrt{d_x} + \log(1/\varepsilon))$ and let $\tau_{\text{chunk}} \leq c_3$ and $\tau_{\text{chunk}} - \tau_{\text{obs}} \geq \frac{1}{L_\beta} \log(c_1/\varepsilon)$. Consider running HINT for $\sigma > 0$ with parameters J, α polynomial in the parameters given in Assumption 3.1 specified in Appendix L. Then, Then, if

$$N_{\texttt{exp}} \geq \text{poly}\left(C_\Theta(\sigma, \tau_{\text{chunk}}, \tau_{\text{obs}}), 1/\varepsilon, R_{\mathbf{K}}, d_{\mathcal{A}}, \log(1/\delta)\right)^\nu > 0,$$

then with probability $1-\delta$, the policy $\hat{\pi}_\sigma$ returned by HINT satisfies

$$\mathcal{L}_{\text{marg},\varepsilon}(\hat{\pi}_\sigma) \leq \Theta_{\text{Iss}}\left(\varepsilon H \sqrt{\tau_{\text{obs}}} \cdot (\sqrt{d_x} + \log(1/\varepsilon)\right).$$

In addition, consider running HINT with $\sigma = 0$, and suppose $C_\Theta(\sigma, \tau_{\text{chunk}}, \tau_{\text{obs}})\big|_{\sigma=0}$ is finite. Then, for $N_{\texttt{exp}}$ satisfying the same bound as above, it holds that with probability $1-\delta$, the policy $\hat{\pi}$ produced by HINT satisfies the guarantees of Theorem 1 up to an additive factor of $H\varepsilon$ on the event that $\hat{\pi}$ happens to be γ-TVC.

Theorem 3 instantiates Theorem 2 by bounding the policy error terms $\Delta_{(\varepsilon^2)}(\pi^\star_{\text{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h))$ in that theorem when $\hat{\pi}$ is the policy learned by fitting the DDPM in Line 7. The formal bound on $N_{\texttt{exp}}$ is given in (J.6) in Appendix J. While making τ_{chunk} larger appears to improve the above bound, it generally increases the statistical and computational challenge of learning the DDPM itself. The guarantees for score matching are derived in Appendix L by applying [18, 41, 13]; these are applied to Theorem 2 in Appendix J.

D Analysis Overview

Our analysis abstracts away the vector-valued dynamics into a deterministic MDP – the ***composite MDP*** – with ***composite-states*** $\mathsf{s} \in \mathcal{S}$ and ***composite-actions*** $\mathsf{a} \in \mathcal{A}$, corresponding to trajectory-chunks and composite-actions in Section 2. We abstract away our dynamics as

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

A ***composite-policy*** π is a sequence of kernels $\pi_1, \pi_2, \ldots, \pi_H : \mathcal{S} \to \Delta(\mathcal{A})$. We let P_{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 $\mathsf{s}_1 \sim \mathsf{P}_{\text{init}}$, $\mathsf{a}_h \mid \mathsf{s}_{1:h}, \mathsf{a}_{1:h-1} \sim \pi_h(\mathsf{s}_h)$, and the composite-dynamics (D.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 $\mathsf{D}_{\pi^\star}$; ultimately, we shall take $\pi^\star$ to be the policy defined in Definition 3.4.

D.1 Structure of the proof.

We begin by explaining key objects, stability and continuity properties required in the composite MDP. Then, Appendix D.2 relates the composite MDP to our original setting by taking composite-states $\mathsf{s}_h = \mathsf{s}_h$ 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 2 from a generic guarantee for smoothed imitation learning in the composite MDP, Theorem 4, and from sampling guarantees in Appendix L.

We consider two pseudometrics on the space $\mathcal{S}$: $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\text{TVC}} : \mathcal{S}^2 \to \mathbb{R}_{\geq 0}$, 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 $\mathsf{d}_{\text{TVC}}(\cdot,\cdot)$ 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 ($\Gamma_{\text{joint},\varepsilon}$) (ii) marginal distribution of trajectories ($\Gamma_{\text{marg},\varepsilon}$) and (iii) one-step error in actions ($\mathsf{d}_{\text{os},\varepsilon}$). Formally:

Definition D.1 (Imitation Errors)**.** Given an error parameter $\varepsilon > 0$, define the ***joint-error*** $\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star) := \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\max_{h\in[H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}^\star_h, \hat{\mathsf{a}}_h)\} > \varepsilon\right]$, 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*** $\Gamma_{\text{marg},\varepsilon}(\hat{\pi} \parallel \pi^\star) := \max_{h\in[H]}\{\inf_{\mu_1} \mathbb{P}_{\mu_1}[\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1}, \hat{\mathsf{s}}_{h+1}) > \varepsilon], \inf_{\mu_1} \mathbb{P}_{\mu_1}[\mathsf{d}_{\mathcal{A}}(\mathsf{a}^\star_h, \hat{\mathsf{a}}_h) > \varepsilon]\}$ 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*** $\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}) \parallel \pi^\star_h(\mathsf{s})) := \inf_{\mu_2} \mathbb{P}_{\mu_2}\left[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) \leq \varepsilon\right]$, where the infimum is over $(\mathsf{a}^\star_h, \hat{\mathsf{a}}_h) \sim \mu_2 \in \mathscr{C}(\hat{\pi}_h(\mathsf{s}), \pi^\star_h(\mathsf{s}))$.

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 E.2 for further discussion).

Definition D.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 $\mathsf{d}_{\mathcal{S}}(\mathsf{s}'_{h+1}, \mathsf{s}_{h+1}) \vee \mathsf{d}_{\text{TVC}}(\mathsf{s}'_{h+1}, \mathsf{s}_{h+1}) \leq \max_{1\leq j\leq h} \mathsf{d}_{\mathcal{A}}\left(\mathsf{a}'_j, \mathsf{a}_j\right), \forall h \in [H]$. 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 d_{TVC}.

Definition D.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 $\mathsf{s}, \mathsf{s}' \in \mathcal{S}$, $\mathrm{TV}(\mathsf{W}(\mathsf{s}), \mathsf{W}(\mathsf{s}')) \leq \gamma(\mathsf{d}_{\text{TVC}}(\mathsf{s}, \mathsf{s}'))$. A policy π is γ***-TVC*** if $\pi_h : \mathcal{S} \to \Delta(\mathcal{A})$ is γ-TVC $\forall h \in [H]$.

Data Noising. In Appendix G, we show that under the strong condition that the learned policy $\hat{\pi}$ is γ-TVC, then HINT with no noise injection ($\sigma = 0$) 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 noising procedure; in HINT we let the kernel be a Gaussian, sending o_h to $\mathcal{N}(\mathsf{o}_h, \sigma^2\mathbf{I}) \in \Delta(\mathscr{P}_{\mathsf{o}_h})$. We will thus be able to replace TVC of $\hat{\pi}$ with TVC of W_σ. We now introduce a few key objects.

Definition D.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 ***smoothed distibution*** $\mathsf{P}^\star_{\text{aug},h}$ 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*** $\pi^\star_{\text{dec}}$ is defined by letting $\pi^\star_{\text{dec},h}(\mathsf{s})$ 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 $\mathsf{P}^\star_{\text{aug},h}$. Finally, the ***replica policy*** is $\pi^\star_{\text{rep}} = \pi^\star_{\text{dec}} \circ \mathsf{W}_\sigma$.

The operator $\pi \circ \mathsf{W}_\sigma$ composes π with the smoothing kernel. The deconvolution policy $\pi^\star_{\text{dec}}$ captures the distribution of actions under $\pi^\star$ given a smoothed state, and corresponds to $\pi^\star_{\text{dec}} = (\pi^\star_{\text{dec},h})_{h=1}^H$. We argue that if a policy $\hat{\pi}$ approximates $\pi^\star_{\text{dec}}$ at each step, then $\hat{\pi} \circ \mathsf{W}_\sigma$ imitates $\pi^\star_{\text{rep}} = \pi^\star_{\text{dec}} \circ \mathsf{W}_\sigma$. 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 $\mathsf{d}_{\text{TVC}}, \mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathcal{A}}$ as given.

Definition D.5. For a non-decreasing maps $\gamma_{\text{IPS,TVC}}, \gamma_{\text{IPS},\mathcal{S}} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ a pseudometric $\mathsf{d}_{\text{IPS}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ (possibly other than $\mathsf{d}_{\mathcal{S}}$ or d_{TVC}), and $r_{\text{IPS}} > 0$, we say a policy π is $(\gamma_{\text{IPS,TVC}}, \gamma_{\text{IPS},\mathcal{S}}, \mathsf{d}_{\text{IPS}}, r_{\text{IPS}})$*-input-&-process stable (IPS)* if the following holds for any $r \in [0, r_{\text{IPS}}]$. Consider any sequence of kernels $\mathsf{W}_1, \ldots, \mathsf{W}_H : \mathcal{S} \to \Delta(\mathcal{S})$ satisfying $\max_{h,\mathsf{s}\in\mathcal{S}} \mathbb{P}_{\tilde{\mathsf{s}}\sim\mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\text{IPS}}(\tilde{\mathsf{s}}, \mathsf{s}) \leq r] = 1$, and define a process $\mathsf{s}_1 \sim \mathsf{P}_{\text{init}}, \tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h), \mathsf{a}_h \sim \pi_h(\tilde{\mathsf{s}}_h)$, 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 (b) $\max_{h\in[H]} \mathsf{d}_{\text{TVC}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \leq \gamma_{\text{IPS,TVC}}(r)$ and (c) $\max_{h\in[H]} \mathsf{d}_{\mathcal{S}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \leq \gamma_{\text{IPS},\mathcal{S}}(r)$.

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* $\mathsf{W}^\star_{\text{rep},h}$ for which $\pi^\star_{\text{rep},h} = \pi^\star_h \circ \mathsf{W}^\star_{\text{rep},h}$. We show that the replica kernel inherits any concentration satisfied by W_σ, ensuring (via truncation) that $\mathbb{P}_{\tilde{\mathsf{s}}\sim\mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\text{IPS}}(\tilde{\mathsf{s}}, \mathsf{s})] \leq r$. Conditions (b & c) merely require that one-step dynamics are robust to small changes in state, measured in terms of both d_{TVC} and $\mathsf{d}_{\mathcal{S}}$.

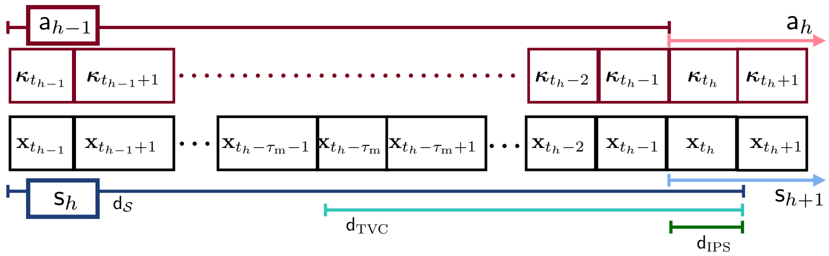


Figure 7: Schematic depicting the composite MDP. States $\mathbf{x}$ and stabilizing gains κ are chunked into composite states s and composite actions a (control inputs $\mathbf{u}$ not depicted). The distance labels correspond to the domain over which each distance is evaluated. Note that a_h begins at the same time that s_{h+1} does, an indexing convention that we adopt to make the notation in the proofs simpler.

D.2 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. Recall the definitions of $\mathsf{d}_{\max}$ and $\mathsf{d}_{\mathrm{traj}}$ from (3.1) and (3.2). In the former case, $\mathsf{s}_h = (\mathbf{x}_{t_{h-1}:t_h}, \mathbf{u}_{t_{h-1}:t_h-1}) \in \mathscr{P}_{\tau_{\mathrm{obs}}}$, and $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$ (as in Section 2). Importantly, a_h ***are primitive controllers, which allows us to meet the strong stability condition in Definition D.2***. Figure 7 provides a visual aid for the subtle indexing. For $\mathsf{s}_h = \mathsf{s}_h, \mathsf{s}'_h = \mathsf{s}'_h$, 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, $\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_{t \in [t_h - \tau_{\mathrm{obs}}:t_h]} \|\mathbf{x}_t - \mathbf{x}'_t\| \vee \max_{t \in [t_h-\tau_{\mathrm{obs}}:t_h-1]} \|\mathbf{u}_t - \mathbf{u}'_t\|$, which measures distance on the last τ_{obs} 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 $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{x}}_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{K}}_{1:\tau_{\mathrm{chunk}}})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{x}}'_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{K}}'_{1:\tau_{\mathrm{chunk}}})$, we recall from (3.1) $\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') := \max_{1 \le k \le \tau_{\mathrm{chunk}}} (\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|)$. We choose a $\mathsf{d}_{\mathcal{A}}$ that takes value ∞ when primitive controllers are too far apart

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') := c_1 \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \cdot \mathbf{I}_\infty\{\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le c_2\} \tag{D.2}$$

$\mathbf{I}_\infty$ is the indicator taking infinite value when the event $\mathcal{E}$ fails to hold and 1 otherwise, and c_1 and c_2 are constants depending polynomially on all problem parameters, given in Appendix K.

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 o_h under $\mathcal{D}_{\mathrm{exp}}$ (see Appendix J for a rigorous definition). As noted above, we take the smoothing kernel W_σ to map o_h to a $\mathcal{N}(\mathsf{o}_h, \sigma^2 \mathbf{I}) \in \Delta(\mathscr{P}_{\mathsf{o}_h})$, 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 C.1***.

Appendix K shows that Assumption 3.1 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 D.1. Let $c_3, c_4, c_5 > 0$ be as given in Appendix K (and polynomial in the terms in Assumption 3.1). Suppose $\tau_{\mathrm{chunk}} \ge c_3$, and let $r_{\mathrm{IPS}} = c_4$, $\gamma_{\mathrm{IPS,TVC}}(u) = c_5 u \exp(-L_\beta(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}}))$, $\gamma_{\mathrm{IPS},\mathcal{S}}(u) = c_5 u$. Then, for $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ as above, we have that $\pi^\star$ is $(\gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-IPS.

D.3 A Guarantee in the Composite MDP, and the derivation of Theorem 3

With the substitutions in Appendix D.2, 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 4. Suppose $\pi^\star$ is $(\gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-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 $\varepsilon' := \varepsilon + \gamma_{\mathrm{IPS},\mathcal{S}}(2r)$. Then, for any policy $\hat{\pi}$, both $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\mathrm{rep}})$ and $\Gamma_{\mathrm{marg},\varepsilon'}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ are upper bounded by

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

Deriving Theorem 2 from Theorem 4. A full proof is given in Appendix J. The key steps are using the stability guarantee of Proposition D.1, the aforementioned TVC-bound on W_σ, and Gaussian concentration to bound p_r with the bound in Theorem 4 to conclude. Moreover, we can show that, with $\tilde{\mathsf{s}}^\star_h = \tilde{\mathsf{o}}_h$,

$$\sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h)) = \sum_{h=1}^{H} \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}} \left[\Delta_{(\varepsilon^2)}\left(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)\right)\right].$$

Thus, Theorem 4 provides the desired guarantee for imitating the policy $\pi^\star$ constructed in Appendix D.2. We conclude with a subtle but powerful observation: that $\pi^\star$ as constructed has trajectories with the same marginals (but possibly different joint distributions) as $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$.

□

A simplified guarantee when $\hat{\pi}$ is TVC. Before we sketch the proof of Theorem 4, we present the simpler guarantee that underpins Theorem 1 in Section 3. This guarantee considers *no smoothing*, but where $\hat{\pi}$ is guaranteed to be γ-TVC. As discussed in Remark C.3, this situation is practically unrealistic when $\hat{\pi}$ is instantiated with a DDPM. Still, the following result is more transparent, and its proof sketch will inform the proof sketch of Theorem 4 following it.

Proposition D.2. Suppose for simplicity that $\mathsf{d}_{\mathcal{S}} = \mathsf{d}_{\textsc{tvc}}$. 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 $\varepsilon > 0$,

$$\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star) \le H\gamma(\varepsilon) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}^\star_h) \parallel \pi^\star(\mathsf{s}^\star_h)) \quad \text{(D.4)}$$

Proof Sketch of Proposition D.2. We couple a trajectory $(\mathsf{s}^\star_h, \mathsf{a}^\star_h)$ induced by $\pi^\star$ with a trajectory $(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h)$ induced by $\hat{\pi}$. To do so, consider an inductive event on which all past states and actions are close

$$\mathcal{E}_h = \{\forall j \le h,\ \mathsf{d}_{\textsc{tvc}}(\hat{\mathsf{s}}_j, \mathsf{s}^\star_j) \vee \mathsf{d}_{\mathcal{S}}(\hat{\mathsf{s}}_j, \mathsf{s}^\star_j) \le \varepsilon\} \cap \{\forall j \le h-1,\ \mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_j, \mathsf{a}^\star_j) \le \varepsilon\},$$

which can be made to hold with probability one for $h = 1$ by ensuring $\hat{\mathsf{s}}_1 = \mathsf{s}^\star_1 \sim \mathsf{P}_{\mathrm{init}}$. It suffices to find a μ coupling under which $\mathbb{P}_\mu[\mathcal{E}_{H+1}]$ is bounded by the righthand side of (D.4).

When $\mathcal{E}_h$ holds, then as $\hat{\pi}$ is γ-TVC, $\hat{\mathsf{a}}_h \sim \hat{\pi}_h(\hat{\mathsf{s}}_h)$ is $\gamma(\varepsilon)$-close in TV distance to an interpolating action $\hat{\mathsf{a}}^{\mathrm{inter}}_h \sim \hat{\pi}_h(\mathsf{s}^\star_h)$. Thus, there is a coupling under which $\mathbb{P}[\hat{\mathsf{a}}_h \ne \hat{\mathsf{a}}^{\mathrm{inter}}_h] \le \gamma(\varepsilon)$. Moreover, by definition of $\mathsf{d}_{\mathrm{os},\varepsilon}$, there exists a coupling under which $\mathbb{P}[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}^{\mathrm{inter}}_h, \mathsf{a}^\star_h) > \varepsilon] \le \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}^\star_h) \parallel \pi^\star(\mathsf{s}^\star_h))]$. Thus, by "gluing" the couplings together (an operation rigorously justified in Appendix F), we have

$$\mathbb{P}[\mathcal{E}_h \cap \{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) > \varepsilon\}] \le \gamma(\varepsilon) + \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}^\star_h) \parallel \pi^\star(\mathsf{s}^\star_h))]. \quad \text{(D.5)}$$

Invoking the defition of input stability (Definition D.2), $\mathcal{E}_h$ and $\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h, \mathsf{a}^\star_h) \le \varepsilon$ imply $\mathcal{E}_{h+1}$. Therefore, by selecting couplings appropriately, $\mathbb{P}[\mathcal{E}_h \cap \mathcal{E}^c_{h+1}]$ is also bounded by the righthand side of (D.5). As the events $(\mathcal{E}_h)$ are nested, we can finally telescope to bound $\mathbb{P}[\mathcal{E}_{H+1}]$ by summing up these terms for each h. A full proof is given in Appendix G. □

D.4 Proof Overview of Theorem 4

In this sketch, we focus on bounding $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\mathrm{rep}})$; we adress $\Gamma_{\mathrm{marg},\varepsilon'}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ at the end of the section. Our argument constructs a coupling between a *replica trajectory* over $(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{a}^{\mathrm{rep}}_h)$ sampled using the replica policy $\pi^\star_{\mathrm{rep}}$, and an *imitator trajectory* $(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h)$ sampled from $\hat{\pi}_\sigma$. The construction of this coupling draws inspiration from the notion of replica pairs in statistical physics (hence, the name replica) [45]. We refer the reader back to Figure 4 above the proof sketch in Section 3.3 for a visual depiction.

The replica kernel. A central object in our proof is the *replica kernel* $\mathsf{W}^\star_{\mathrm{rep},h} : \mathcal{S} \to \Delta(\mathcal{S})$, defined so that $\pi^\star_{\mathrm{rep},h} = \pi^\star_h \circ \mathsf{W}^\star_{\mathrm{rep},h}$, and constructed by analogy to the replica policy in Definition D.4. The key property of the replica kernel is that it preserves marginals: if $\mathsf{s}_h \sim \mathsf{P}^\star_h$ is drawn from the distribution of states under the expert demonstrations, then $\mathsf{s}'_h \sim \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s}_h)$ is also distributed as $\mathsf{P}^\star_h$; in other words, $\mathsf{P}^\star_h$ is a fixed point of $\mathsf{W}^\star_{\mathrm{rep},h}$:

Fact D.1 (Replica Property)**.** It holds that $\mathsf{P}^\star_h = \mathsf{W}^\star_{\mathrm{rep},h} \circ \mathsf{P}^\star_h$.

A second crucial property is that we can represent the replica kernel as a convolution between W_σ and a deconvolution kernel. Thus, data-processing implies that $\mathsf{W}^\star_{\mathrm{rep},h}$ inherits TVC from W_σ.

Fact D.2 (Replica Kernel is TVC)**.** If W_σ is $\gamma_{\textsc{tvc}}$-TVC, $\mathsf{W}^\star_{\mathrm{rep},h}$ is as well.

Constructing the replica and *teleporting* trajectories. Because the replica kernel satisfies $\pi^\star_{\mathrm{rep},h} = \mathsf{W}^\star_{\mathrm{rep},h} \circ \pi^\star$, we can realize the replica trajectory via

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

We then introduce a *teleporting* trajectory obeying the an almost identical generative process:

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

In words, $\mathsf{s}^{\mathrm{tel}}_h$ "teleports" to an independent and identically distributed copy conditional on the noise agumentation $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$, and draws an action from the expert policy evaluated on the new state. The replica and teleporting sequences differ only in the transitions: whereas $\mathsf{s}^{\mathrm{rep}}_{h+1} = F_h(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{a}^{\mathrm{rep}}_h)$ transitions from its *current* state $\mathsf{s}^{\mathrm{rep}}_h$, the telporting trajectory transition $\mathsf{s}^{\mathrm{tel}}_{h+1} = F_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h, \mathsf{a}^{\mathrm{tel}}_h)$ from the *replica-drawn*, "teleported" state $\tilde{\mathsf{s}}^{\mathrm{tel}}_h \sim \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s}^{\mathrm{tel}}_h)$. By iteratively applying Fact D.1, and the fact that $\mathsf{s}^{\mathrm{tel}}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)$, we can make the following claim:

Fact D.3. For each h, $\mathsf{s}^{\mathrm{tel}}_h$ and $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$ are distributed according to $\mathsf{P}^\star_h$, the marginal distribution of states under the expert policy. Hence, because $\mathsf{s}^{\mathrm{tel}}_h \sim \mathsf{P}^\star_h$, we know that $\hat{\pi}$ and $\pi^\star_h$ are close under the distribution of states induced teleporting sequence.

Constructing the coupling. We now describe how to use the teleporting trajectory to couple the replica and imitator trajectory. We begin by coupling the replica and imitator trajectories. The following diagram explains how we construct the coupling:

$$\underbrace{(\tilde{\mathsf{s}}^{\mathrm{rep}} \leftrightarrow \tilde{\mathsf{s}}^{\mathrm{tel}}), (\mathsf{a}^{\mathrm{rep}} \leftrightarrow \mathsf{a}^{\mathrm{tel}})}_{\gamma_{\textsc{tvc}},\ \gamma_{\textsc{ips},\textsc{tvc}},\ \text{and induction}} \to \underbrace{(\mathsf{a}^{\mathrm{tel}} \leftrightarrow \hat{\mathsf{a}}^{\mathrm{tel,inter}})}_{\text{learning and sampling}} \to \underbrace{(\hat{\mathsf{a}}^{\mathrm{tel,inter}} \leftrightarrow \hat{\mathsf{a}}^{\mathrm{rep,inter}})}_{\gamma_{\textsc{tvc}}\text{ and induction}} \to \underbrace{(\hat{\mathsf{a}}^{\mathrm{rep,inter}} \leftrightarrow \hat{\mathsf{a}})}_{\gamma_{\textsc{tvc}},\ \text{input-stability, and induction}}. \tag{D.6}$$

As the dynamics are deterministic, (D.6) determines the coupling of states $\mathsf{s}^{\mathrm{rep}}_h, \mathsf{s}^{\mathrm{tel}}_h, \hat{\mathsf{s}}_h$ as well.

(a) We begin by arguing that replica and teleporting trajectories are close to one another. The argument is inductive: suppose that $\mathsf{d}_{\textsc{tvc}}(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{s}^{\mathrm{tel}}_h)$ are close. By TVC of the replica kernel, $\tilde{\mathsf{s}}^{\mathrm{rep}}_h$ and $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$ are close in TV, and so there is a coupling under which

$$\mathbb{P}[(\mathsf{a}^{\mathrm{rep}}_h, \tilde{\mathsf{s}}^{\mathrm{rep}}_h) \neq (\mathsf{a}^{\mathrm{tel}}_h, \tilde{\mathsf{s}}^{\mathrm{tel}}_h)] \text{ is small.}$$

Next, recall $p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\textsc{ips}}(\mathsf{s}', \mathsf{s}) > r]$ as defined in Theorem 4, which describes the concentration behavior of W_σ. We use a Bayesian concentration argument (Lemma H.5) to ensure that with probability of failure at most $2p_r$, $\mathsf{d}_{\textsc{ips}}(\tilde{\mathsf{s}}^{\mathrm{rep}}_h, \mathsf{s}^{\mathrm{rep}}_h) \leq 2r$. We then use IPS (Definition D.5) to argue that, with the same failure probability,

$$\mathsf{s}^{\mathrm{rep}}_{h+1} = F_h(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{a}^{\mathrm{rep}}_h) \text{ is within } \mathcal{O}(r)\ \mathsf{d}_{\textsc{tvc}}\text{-distance of } F_h(\tilde{\mathsf{s}}^{\mathrm{rep}}_h, \mathsf{a}^{\mathrm{rep}}_h).$$

Thus, when $(\mathsf{a}^{\mathrm{rep}}_h, \tilde{\mathsf{s}}^{\mathrm{rep}}_h) = (\mathsf{a}^{\mathrm{tel}}_h, \tilde{\mathsf{s}}^{\mathrm{tel}}_h)$, we obtain that $\mathsf{s}^{\mathrm{rep}}_{h+1}$ is close to $F_h(\mathsf{a}^{\mathrm{tel}}_h, \tilde{\mathsf{s}}^{\mathrm{tel}}_h) = \mathsf{s}^{\mathrm{tel}}_{h+1}$ in $\mathsf{d}_{\textsc{tvc}}$ as well. For more detail, consult Figure 10 in the appendix.

(b) Because the marginals of $\mathsf{s}^{\mathrm{tel}}_h$ are distributed according to $\mathsf{P}^\star_h$, we can argue that a (fictitious) action $\hat{\mathsf{a}}^{\mathrm{tel,inter}}_h \sim (\hat{\pi}_h \circ \mathsf{W}_\sigma)(\mathsf{s}^{\mathrm{tel}}_h)$ is close to $\mathsf{a}^{\mathrm{tel}}_h$. Indeed, by the data processing inequality, it is bounded by the closeness of $\hat{\pi}_h$ and $\pi^\star_{\mathrm{dec},h}$ on $\tilde{\mathsf{s}}^{\mathrm{tel}}_h \sim \mathsf{W}_\sigma(\mathsf{s}^{\mathrm{tel}}_h)$, $\mathsf{s}^{\mathrm{tel}}_h \sim \mathsf{P}^\star_h$; this is controlled by the contribution of $\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}^\star_h))$ on the right-hand-side of (M.2).

(c) From part (a) of the coupling, we expect that $\mathsf{s}_h^{\mathrm{rep}}$ and $\mathsf{s}_h^{\mathrm{tel}}$ are close. As $\hat{\pi} \circ \mathsf{W}_\sigma$ is γ_{TVC} by the data-processing inequality, it follows that $\hat{\mathsf{a}}_h^{\mathrm{tel,inter}}$ is close in TV distance to another fictious action $\hat{\mathsf{a}}_h^{\mathrm{rep,inter}} \sim (\hat{\pi}_h \circ \mathsf{W}_\sigma)(\mathsf{s}_h^{\mathrm{rep}})$.

(d) Finally, we argue that that trajectory of actions induced by taking the replica-interpolating action $\hat{\mathsf{a}}_h^{\mathrm{rep,inter}}$ is close to TV to the imitation trajectories induced by taking $\hat{\mathsf{a}}_h$. The argument is similar in spirit to the proof of Proposition D.2 sketched above, and uses both TVC of the smoothed policy $\hat{\pi} \circ \mathsf{W}_\sigma$, and the form of input-stability guaranteed by Definition D.5.

Bounding the marginal gap. Because the teleporting sequence $\mathsf{s}_h^{\mathrm{tel}}$ has marginals $\mathsf{P}_h^\star$, bounding the marginal gap amounts to controlling the distance between $\mathsf{s}_h^{\mathrm{tel}}$ and $\hat{\mathsf{s}}_h$. This follows from more-or-less the same manipulations.[5]

Measure-theoretic considerations. 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 F.3. To conclude, we apply a measure theoretic result (Lemma F.2) to "glue" the pairwise couplings together and establish the main result. The full proof is given in Appendix H, relying on measure-theoretic details in Appendix F. $\square$

E Generalization to Generic Incrementally Stable Primitive Controllers

In this section, we consider a generalization of the theory to allow for general, nonlinear primitive controllers, as long as they obey the incremental stability considered in Pfrommer et al. [50]. We consider controllers of the form:

$$\kappa(\mathbf{x}) = \kappa(\mathbf{x}; \theta), \quad \theta \in \Theta.$$

We assume that $\Theta \subset \mathbb{R}^{d_\Theta}$ is a measurable subset of a finite dimensional space and (2) that $\kappa(\mathbf{x}; \theta)$ is jointly piecewise-Lipschitz with at most countably many pieces. We define composite actions just as for linear primitive controllers:

$$\mathsf{a} = (\kappa_1, \kappa_2, \dots, \kappa_{\tau_{\mathrm{chunk}}}).$$

We view $\mathcal{K} := \{\kappa(\cdot; \theta) | \theta \in \Theta\}$ and $\mathcal{A} = \mathcal{K}^{\tau_{\mathrm{chunk}}}$ as Polish spaces with the Euclidean metric on the controller parameters θ (resp. sequences of control parameters). Our definition of incremental stability from Definition 3.1 applies verbatim to these more general controllers.

Example E.1 (Approximate Inverse Dynamics & Position Control)**.** A natural example of the above is where θ corresponds to a sequence of position commands supplied to a robotic position controller, as in [19], or where θ is a of state-command given to an inverse dynamics model, as in [3]. In these settings, we can actually regard θ as the "control action," and envision the closed loop system of {system + position controller/inverse dynamics model} as itself being incrementally stable. However, our framework is considerably more general, and allows us, for example, to diffuse other parameters governing the performance of the low level controllers as well (e.g. joint spring constants in robotic position control).

In this section, we replace Assumption 3.1 with the more general assumption that allows arbitrary forms of the incremental stability moduli:

Assumption 3.1b. Let $\gamma(\cdot)$ be class K, $\beta(\cdot,\cdot)$ be class KL, and let $c_\xi, c_\beta, c_\gamma > 0$ be postive constants. We assume access to a synthesis oracle $\mathtt{synth} : \mathscr{P}_T \to \mathcal{A}^H$ such that, with probability 1 over $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T}) \sim \mathcal{D}_{\exp}$, $\mathsf{a}_{1:H} = \mathtt{synth}(\boldsymbol{\rho}_T)$ satisfies the following properties:

- $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$ is consistent with $\mathsf{s}_{h+1} = (\mathbf{x}_{t_h:t_{h+1}}, \mathbf{u}_{t_h:t_{h+1}-1})$; equivalently,

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \kappa_t(\mathbf{x}_t)), \quad t \in [T].$$

- a_h is locally $\mathtt{t}$-$\mathtt{ISS}$ at $\mathbf{x}_{t_h}$ with moduli $\gamma(\cdot), \beta(\cdot,\cdot)$ and constants $c_\gamma, c_\beta, c_\xi > 0$.

[5] A key difference is that we pick up an additive factor of $\gamma_{\mathrm{IPS},\mathcal{S}}$ (measuring sensitivity of $\mathsf{d}_\mathcal{S}$ to the smoothing from W_σ) in our tolerance ε'.

E.1 Main Results for General Primitive Controllers

To state our main results, we begin by defining distances on primitive controllers, as well as the induced imitation error of a policy $\hat{\pi}$. Throughout, $\mathcal{D}_{\sigma,h}$ denotes the noise-smoothed expert data distribution over $(\tilde{\mathsf{o}}_h, \mathsf{a}_h)$ as in Definition 3.6 (with a_h now representing sequence of general primitive controllers, not the linear ones considered in the main body.

Definition E.1. Define the local-distance between composite actions $\mathsf{a} = \kappa_{1:\tau_{\mathrm{chunk}}}, \mathsf{a}' = \kappa'_{1:\tau_{\mathrm{chunk}}} \in \mathcal{A}$ at state $\mathbf{x}$ and scale $\alpha > 0$ as

$$\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a}, \mathsf{a}' \mid \mathbf{x}) := \max_{1 \le i \le \tau_{\mathrm{chunk}}} \sup_{\delta\mathbf{x}: \|\delta\mathbf{x}\| \le \alpha} \|\kappa_i(\mathbf{x}_i + \delta\mathbf{x}) - \kappa'_i(\mathbf{x}_i + \delta\mathbf{x})\|,$$

where above $\mathbf{x}_1 = \mathbf{x}$, $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \kappa_t(\mathbf{x}))$, and $\mathsf{a} = \kappa_{1:\tau_{\mathrm{chunk}}}$. Finally, we define

$$\Delta_{\mathrm{Iss},\sigma,h}(\hat{\pi}; \varepsilon, \alpha) := \inf_{\mu \in \overline{\mathscr{C}}_{\sigma,h}(\hat{\pi})} \mathbb{P}_{(\mathsf{a}_h, \mathsf{a}'_h) \sim \mu} \left[\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a}, \mathsf{a}' \mid \mathbf{x}_{t_h})) > \varepsilon \right],$$

where $\overline{\mathscr{C}}_{\sigma,h}(\hat{\pi})$ denotes the set of couplings of $(\mathsf{o}_h, \tilde{\mathsf{o}}_h, \mathsf{a}_h, \mathsf{a}'_h)$, induced by drawing $(\mathsf{o}_h, \mathsf{a}_h) \sim \mathcal{D}_{\mathrm{exp},h}$, $\tilde{\mathsf{o}}_h \sim \mathcal{N}(\mathsf{o}_h, \sigma^2 \mathbf{I})$, and $\mathsf{a}'_h \sim \hat{\pi}_h(\tilde{\mathsf{o}}_h)$, and where above $\mathbf{x}_{t_h}$ is the last state in o_h.Note that the only degree of freedom for in selecting elements of $\overline{\mathscr{C}}_{\sigma,h}(\hat{\pi})$.

Remark E.1 (On the distance $\mathsf{d}_{\mathrm{loc},\alpha}$)**.** In words, $\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a}, \mathsf{a}' \mid \mathbf{x}_{1:\tau_{\mathrm{chunk}}})$ measures the supremal distance between the primitive controllers comprising a, a', along radius-α neighborhoods of a given sequence $\mathbf{x}_{1:\tau_{\mathrm{chunk}}}$. This supremal distance was studied in Pfrommer et al. [50] and motivates their proposed algorithm TaSIL. *Unlike affine primitive controllers,* the supremal distance between general primitive controllers may indeed dependence on the localizing sequence $\mathbf{x}_{1:\tau_{\mathrm{chunk}}}$.

Having defined our distance, $\Delta_{\mathrm{Iss},\sigma,h}(\hat{\pi}; \varepsilon, \alpha)$ measures the probability that this difference exceeds some threshold $\alpha > 0$, under appropriate couplings where $\tilde{\mathsf{o}}_h$ is induced by noising the expert distribution, a is the corresponding action, a' is from the policy $\hat{\pi}$, and the localizing sequence follows from rolling out a on the last state $\tilde{\mathbf{x}}_{t_h}$ of $\tilde{\mathsf{o}}_h$. $\Delta_{\mathrm{Iss},\sigma,h}$ is the natural analogue of the distances consider in Theorem 2. One subtlety of $\Delta_{\mathrm{Iss},\sigma,h}$ is that the couplings are constructed such that a_h is the action associated with o_h. This is conceptually correct because it specifies a localizing-state for $\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a}_h, \mathsf{a}'_h \mid \mathbf{x}_{t_h})$ (which is not an issue for affine primitive controllers). When a_h arises from a synthesis oracle, $\mathbf{x}_{t_h}$ lies on the trajectory from which a_h is synthesized.

Theorem 2 generalizes as follows.

Theorem 5 (Generalization of Theorem 2 to general primitive controllers)**.** Assume Assumption 3.1b, and let $\varepsilon > 0$ satisfy

$$\gamma^{-1}(\beta(2\gamma(\varepsilon), \tau_{\mathrm{chunk}}) \le \varepsilon \le \min\{c_\gamma, \gamma^{-1}(c_\xi/4)\} \tag{E.1}$$

Define

$$\omega = 2\sqrt{5d_x + 2\log\left(\frac{2\sigma}{\gamma(\varepsilon)}\right)}, \quad \varepsilon_1 = 2\beta(2\gamma(\varepsilon), 0) + 2\beta(2\sigma\omega, 0), \quad \varepsilon_2 = 2\beta(2\gamma(\varepsilon), 0).$$

Then, if $\sigma \le c_\xi/4\omega$ and $\gamma(\varepsilon) \le 2\sigma$, we have

$$\mathcal{L}_{\mathrm{marg},\varepsilon_1}(\hat{\pi}) \le \frac{3H\sqrt{2\tau_{\mathrm{obs}} - 1}}{2\sigma}\left(2\varepsilon_2 + \beta(2\sigma\omega, \tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})\right) + \sum_{h=1}^{H} \Delta_{\mathrm{Iss},\sigma,h}(\hat{\pi}; \varepsilon, \varepsilon_1).$$

Before continuing, let us remark on the parameters and the scaling. Here, ε parametrizes an error scale. ε_1 captures both the imitation error, as well as the radius in which $\Delta_{\mathrm{Iss},\sigma,h}$ is evaluated. ε_2 contributes to the upper bound on the $\mathcal{L}_{\mathrm{marg},\varepsilon_1}$ normalized by $1/\sigma$. We recall that $\mathcal{L}_{\mathrm{marg},\varepsilon_1}$ measures probabilities of deviating from the marginal by a magnitude of at most ε_1 under optimal couplings. To drive the upper bound on $\mathcal{L}_{\mathrm{marg},\varepsilon}$ to zero, we need (a) $\Delta_{\mathrm{Iss},\sigma,h}(\hat{\pi}; \varepsilon, \varepsilon_1) \to 0$, (b) $\varepsilon_2/\sigma \to 0$ and (c) $\frac{1}{\sigma}\beta(2\sigma\omega, \tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}}) \to 0$, which requires $\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}}$ to grow and β to decay in its second argument. To drive the tolerance ε_1 to zero, we require that $2\beta(2\gamma(\varepsilon), 0) + 2\beta(2\sigma\omega, 0)$ to tend to zero, which requires $\sigma \to 0$ as well. [6] The following remark examines the typical scalings of these terms and checks that (E.1) is generally easy to satisfy.

[6]Notice that $\varepsilon_1 \to \Delta_{\mathrm{Iss},\sigma,h}(\hat{\pi}; \varepsilon, \varepsilon_1)$ is non-increasing, so making ε_1 smaller does not increase this term (though making ε smaller does).

Remark E.2. Suppose that for some $c > 0$, $q_1, q_2 \in (0, 1]$, we have $\gamma(u) = cu^{q_1}$ and $\beta(u, \tau) = u^{q_2}\phi(\tau)$ for some decreasing function ϕ. This is the scaling studied in [50], and indeed for smooth systems with stabilizable systems, our analysis essentially shows that we can take $q_1 = q_2 = 1$ and $\phi(\tau)$ to decay exponentially in τ as shown in Appendix K, and which reflects in Theorem 2. For the more general power scalings, (E.1) reads

$$c^{q_2-1}\phi(\tau_{\text{chunk}})^{1/q_1}\varepsilon^{q_2} \le \varepsilon \le \text{constant}.$$

If $q_2 = 1$ (e.g. the stabilizable, smooth case), then this is satisfied whenever ε is sufficiently small and τ_{chunk} is sufficently large. Otherwise, it one has to take $\varepsilon \ge \frac{1}{c}\phi(\tau_{\text{chunk}})^{1/q_1(1-q_2)}$, whch becomes increasingly permissive as τ is enlarged. Morever, for these scalings, we have $\omega = \mathcal{O}(\log(1 + \sigma/\varepsilon^{q_1}))$, $\varepsilon_1 = \mathcal{O}(\varepsilon^{q_1 q_2} + \sigma\omega) = \varepsilon_1 = \widetilde{\mathcal{O}}(\varepsilon^{q_1 q_2} + \sigma)$, and $\varepsilon_2 = \mathcal{O}(\varepsilon^{q_1 q_2})$. In the regime where $q_1 = q_2$, this recovers the scaling observed in Theorem 2.

Similarly, we can generalize Theorem 1 to the general controller setting.

Theorem 6. Suppose Assumption 3.1b holds, and suppose that $\varepsilon > 0$ and $\tau_{\text{chunk}} \in \mathbb{N}$ satisfies (E.1). Then, for any non-decreasing non-negative $\gamma(\cdot)$ and γ-TVC chunking policy $\hat{\pi}$,

$$\mathcal{L}_{\text{marg},\varepsilon_1}(\hat{\pi}) \le H\gamma(\varepsilon) + \sum_{h=1}^{H} \Delta_{\text{ISS},\sigma,h}(\hat{\pi}; \varepsilon, \varepsilon_1), \quad \varepsilon_1 := 2\beta(2\gamma(\varepsilon), 0) \tag{E.2}$$

In addition, suppose the expert distribution $\mathcal{D}_{\text{exp}}$ has at most τ_{obs}-bounded memory (defined formally in Definition J.5). Then $\mathcal{L}_{\text{joint},\alpha(\varepsilon)}(\hat{\pi})$ satisfies the same upper bound (E.2), where $\mathcal{L}_{\text{joint},\alpha}(\hat{\pi})$, formally defined in Definition J.4, measures an optimal transport distance between the *joint distribution* of the expert trajectory and the one induced by $\hat{\pi}$.

The proofs of Theorems 5 and 6 are given in the Appendix M, generalizing the proofs of Theorems 1 and 2 in the main text, respectively. As with Theorem 1, Appendix N.1 shows that we can replace the condition on the chunk length τ_{chunk} on and on ε in (K.4) with the condition $\varepsilon \le c_\gamma$ and the vacuous condition $\tau_{\text{chunk}} \ge 1$, provided that the synthesis oracle produces entire primitive controllers for which the entire sequences $\kappa_{1:T}$ are incrementally stabilizing.

E.2 Comparison to prior notions of stability.

Prior theoretical work in imitation learning focuses either on constraining the learned policy to be stable [27, 68] or assumes the expert policy is suitably stable [50]. The principal notion of stability used in these prior works is *incremental-input-to-state* stability of the closed-loop system under a deterministic, but possibly sophisticated time-independent controller $\pi : \mathcal{X} \to \mathcal{U}$. Importantly, this work considers the imitation of a *joint distribution* over *sequences* of simple controlers κ we call the "primitive controllers". These approach necessitate subtle differences in our choice of definitions described below.

In what follows, we let γ be a class K function and β be a class KL function, as described above Definition 3.1.

Definition E.2 (Incremental Input-to-State Stability)**.** We say a policy $\pi : \mathcal{X} \to \mathcal{U}$ satisfies *Incremental Input-To-State Stability* (δ-ISS) with moduli γ and β if 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 $f_{\text{cl}}(\mathbf{x}_t, \Delta_t) = f(\mathbf{x}_t, \pi(\mathbf{u}_t) + \Delta_t)$ satisfies:

$$\|\mathbf{x}_t(\xi_1; \delta\mathbf{u}_{1:\tau}) - \mathbf{x}_t(\boldsymbol{\xi}_2; \mathbf{0}_{1:\tau})\| \le \beta(\|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|) + \gamma\left(\max_{0 \le s \le t-1} \|\delta\mathbf{u}_s\|\right),$$

where $\mathbf{x}_t(\boldsymbol{\xi}; \delta\mathbf{u}_{1:t-1})$ is the state at time t under f_{cl} with $x_0 = \boldsymbol{\xi}$ and input perturbations $\delta\mathbf{u}_{1:t-1}$.

We say that it satisfies π satisfies *local* δ-ISS with parameter c if the above holds for all of identical initial conditions $\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2$ (with $\beta(0) = 0$) and for $\delta\mathbf{u}_{1:t-1}$ satisfying $\max_{1 \le s \le t} \|\delta\mathbf{u}_s\| \le c$.

Notice that for $\boldsymbol{\xi}_1 \ne \boldsymbol{\xi}_2$, 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

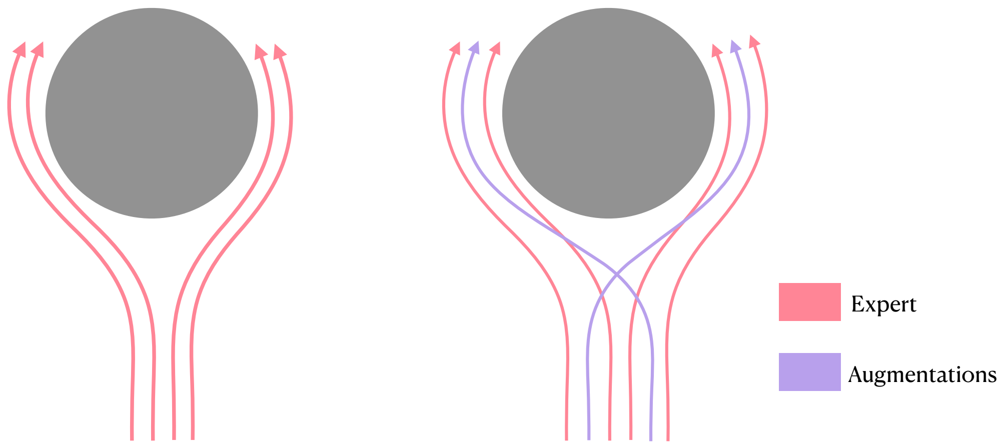


Figure 8: Instance of bifurcation, where augmentation is necessary for stability. The example on the left has an expert demonstrator bifurcating around a circular obstacle. The example on the right demonstrates the utility of augmentations, allowing for trajectories that navigate around the object in the direction farther from their starting point.

challenging. Tu et. al. [68] attempt to do so through regularization while Haven et. a. [27] use matrix inequalities to satisfy this stability property under linear dynamics. Pfrommer et. at. [50] avoid this difficulty only requiring local incremental stability. 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 $\leq c$, this only necessitates that the expert policy can correct from small input perturbations.

Comparing local δ-ISS and Definition 3.1. As stated above, past work consider imitation of a fixed, but possibly complex deterministic controller π. In contrast, we imititate joint distributions over *sequences* of primitive controllers $\mathsf{a} = (\kappa_{1:\tau})$. Moreover, our "primitive" controllers are intended to be much simpler than the policy π considered in past work; e.g. the affine controllers consided in the body in this work. Indeed, the real "policy" we try to imitate is potentially very complex expert distribution $\mathcal{D}_{\mathrm{exp}}$, and these primitive controllers serve to stablize to this distribution. To account for these differences, we modify the local stability considered by Pfrommer et al. [50] in three respects.

- Our notion of stability, Definition 3.1, is applied to fixed-length sequences of controllers $\mathsf{a} = (\kappa_1, \dots, \kappa_\tau)$; past notions of incremental stability are for time-varying controllers and are infinite horizon.
- Definition 3.1 only requires that our notion of incremental stability holds for initial conditions $\boldsymbol{\xi}$ in a radius c_ξ of a nomimal initial condition $\boldsymbol{\xi}_0$. The reason for this can be seen by considering jus tthe affine primitive controllers studied in the body: time-varying feedback that stabilizes the linearization of a smooth dynamical system is only stabilizing of the actual system in a tube around the nominal trajectory.
- Unlike the local notion of δ-ISS considered in Pfrommer et al. [50], we *do* require considering stability from different initial conditions $\boldsymbol{\xi}_1 \neq \boldsymbol{\xi}_2$. This is because we re-apply incremental stability at each chunk h, and must account for imitation error accumulated up to that point.

The power of a hierarchical approach to stability. Through the introduction of a synthesis oracle which can generate locally stabilizing primitive controllers, we decouple the stability properties of the expert's behavior 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 local δ-ISS since an expert policy cannot simultaneously stabilize to multiple branches of a bifurcation. For a concrete example, consider Figure 8. Indeed, continuity is the *sine*

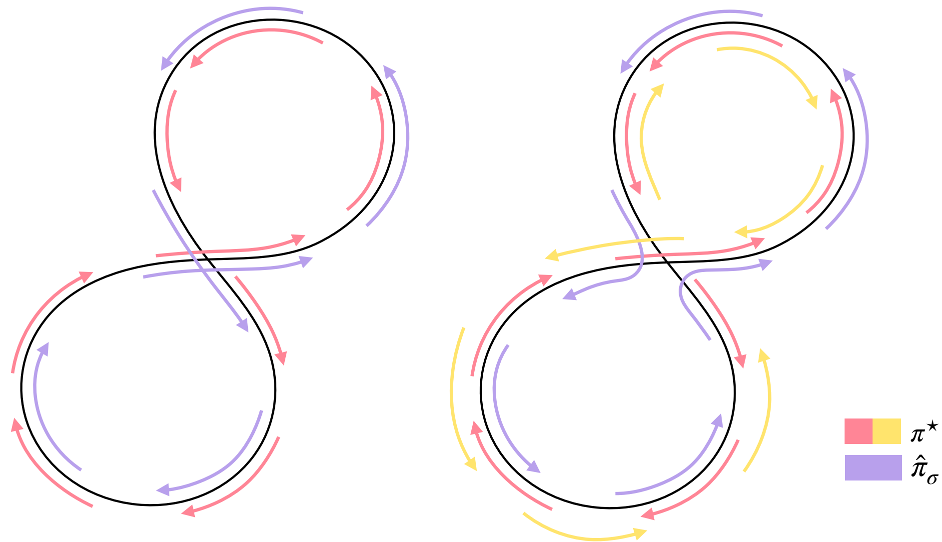


Figure 9: Instance where $\hat{\pi}_\sigma$ and $\pi^\star$ induce the same marginals and joint distributions (left), but in the presence of expert demonstration trajectories that traverse the figure eight both clockwise and counterclockwise directions, $\hat{\pi}_\sigma$ may switch with some probability between demonstrations where they overlap.

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 9, 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 II

Composite MDP

F Measure-Theoretic Background

In this section, we introduce the prerequisite notions from probability theory that we use to formally construct the couplings in Appendices G and H. We begin by introducing general preliminaries, followed by kernels, regular conditional probabilities and a "gluing" lemma in Appendix F.1. We then show that optimal transport costs commute in an appropriate sense with conditional probabilities (Proposition F.3 in Appendix F.2). We use the preliminaries in the previous sections to derive certain optimal-transport and data processing inequalities in Appendix F.3. We prove Proposition F.3 in Appendix F.4. Finally, we state a simple union bound lemma (Lemma F.11 in Appendix F.5) of use in later appendices.

General preliminaries. We rely extensively on the exposition in Durrett [22] 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 F.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 $Y \sim \mathsf{P}_Y$.[7] 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 F.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 [52].

Lemma F.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 F.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 F.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.

[7] 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 $A_2 \in \mathscr{B}(\mathcal{X})$, $\mu(\mathcal{X} \times A_2) = \mathsf{P}_2(A_2)$.

F.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 and Spinka [8] 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 F.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 F.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[8]

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

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 F.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 $\mu(A_1 \times A_2 \times B_2) = \mathbb{E}_{Y \sim \mathsf{P}_Y}\left[\mathsf{Q}_{X_1|Y}(A_1 \mid Y)\mathsf{Q}_{X_2|Y}(A_2)\mathbf{I}\{Y \in B\}\right]$.

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

Theorem 7 (Theorem 5.1.9, Durrett [22])**.** 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.

[8] 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 (F.1) defines a unique probability measure

Lemma F.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 $Q(\cdot \mid Y)$ be a regular conditional probability for Z given Y under μ_2 (who existence is ensured by Theorem 7).

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 F.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 F.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. □

F.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 F.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 F.4. One useful consequence is the following identity for the total variation distance.

Corollary F.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 F.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 F.3 with $\mathcal{X} = \mathcal{X}_1 = \mathcal{X}_2$, and $\phi(x_1, x_2) = \mathbf{I}\{x_1 \neq x_2\}$. □

F.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 F.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 [52, 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 F.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 F.6 and F.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 $\mu = \mathrm{law}(\mathsf{Q}; \mu_Y)$,

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

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 F.1 implies that

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

□

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

Lemma F.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 F.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 F.3. □

F.3.1 Deferred lemmas for the data processing inequalities

Lemma F.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. $\square$

Lemma F.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. $\square$

Lemma F.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, extends to a valid measure. $\square$

F.4 Proof of Proposition F.3

In the case that $\phi(\cdot, \cdot)$ is continuous, the result follows from Villani et al. [73, Corollary 5.22]. For general lower-semicontinuous ϕ, our argument adopts the strategy of "Step 3" of the proof of Villani [72, 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

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

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 F.9. $\psi(y)$ is measurable and satisfies $\psi_n(y) \uparrow \psi(y)$ pointwise.

Proof. We can write

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

Here, (i) follows from the "Step 3" in the proof of Villani [72, 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. $\square$

Claim F.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,

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

Thus, by the monotone convergence theorem,

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

Similarly, repeating some of the above steps,

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

Hence, $\mathbb{E}_Y\psi(Y)\leq\liminf_{m\geq1}\mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\phi(X_1,X_2)]$. By assumption, $\phi(X_1,X_2)$ is lower semicontinuous and bounded from below. Thus, the Portmanteau theorem [22] implies that, as $\mu_{\star,m}\overset{\text{weak}}{\to}\mu_\star$,

$$\liminf_{m\geq1}\mathbb{E}_{(X_1,X_2,Y)\sim\mu_{\star,m}}[\phi(X_1,X_2)]=\mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi(X_1,X_2)].$$

Hence, $\mathbb{E}_Y\psi(Y)\leq\mathbb{E}_{(X_1,X_2,Y)\sim\mu_\star}[\phi(X_1,X_2)]$, 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 7. 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 7, 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. □

F.5 A simple union-bound recursion.

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

Lemma F.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. □

G 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 H.

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

Assumption G.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

- $\mathsf{d}_\mathcal{S}, \mathsf{d}_{\text{TVC}}$ are pseudometrics and Borel measurable function from $\mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\ge 0}$
- For any $\varepsilon \ge 0$, 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 Appendix D.

Proof. The key to the proof is to construct an appropriate "interpolating sequence" of actions $\hat{\mathsf{a}}^{\text{inter}}_{1:H}$ to which we couple both $(\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H})$ 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 $(\mathsf{s}^\star_{1:h}, \mathsf{a}^\star_{1:h})$, $(\hat{\mathsf{s}}_{1:h}, \hat{\mathsf{a}}_{1:h})$, and $\hat{\mathsf{a}}^{\text{inter}}_{1:h}$, and let $\mathcal{F}_0$ denote the σ-algebra generated by $\mathsf{s}^\star_1, \hat{\mathsf{s}}_1$. 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:

$$\mathsf{s}_{h+1}^\star = F_h(\mathsf{s}_h^\star, \mathsf{a}_h^\star), \quad \hat{\mathsf{s}}_{h+1} = F_h(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h) \tag{G.1}$$

  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

$$\mathsf{a}_h^\star \mid \mathcal{F}_{h-1} \sim \pi_h^\star(\mathsf{s}_h^\star), \quad \hat{\mathsf{a}}_{h-1} \mid \mathcal{F}_{h-1} \sim \hat{\pi}_h(\hat{\mathsf{s}}_h), \quad \hat{\mathsf{a}}_h^{\mathrm{inter}} \mid \mathcal{F}_h \sim \hat{\pi}_h(\mathsf{s}_h^\star). \tag{G.2}$$

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 (G.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 (G.1)). Observe that, by the assumption of $\hat{\pi}_h$ being TVC, it holds that

$$\mathsf{TV}(\mathbb{P}_{\hat{\mathsf{a}}_h \mid Y_{h-1}}, \mathbb{P}_{\hat{\mathsf{a}}_h^{\mathrm{inter}} \mid Y_{h-1}}) \le \gamma(\mathsf{d}_{\mathrm{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star)).$$

Thus by Lemma F.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

$$\mathbb{P}[\hat{\mathsf{a}}_h \ne \hat{\mathsf{a}}_h^{\mathrm{inter}}] \le \mathbb{E}_{\mu^{(h-1)}}[\gamma(\mathsf{d}_{\mathrm{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star))].$$

Similarly by Proposition F.3, there is a coupling $\mu_2^{(h)}$ of $Y_{h-1}, \hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star$ such that

$$\mathbb{P}_{\mu_2^{(h)}}[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) > \varepsilon] \le \mathbb{E}_{\mathsf{s}_h^\star \sim \mu^{(h-1)}}[\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))].$$

By the gluing lemma Lemma F.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),

$$\begin{aligned}
&\mathbb{P}_{\mu^{(h)}}[\{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) > \varepsilon\} \cup \{\hat{\mathsf{a}}_h \ne \hat{\mathsf{a}}_h^{\mathrm{inter}}\} \mid \mathcal{F}_{h-1}] \\
&= \mathbb{P}_{\mu^{(h)}}[\{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) > \varepsilon\} \cup \{\hat{\mathsf{a}}_h \ne \hat{\mathsf{a}}_h^{\mathrm{inter}}\} \mid Y_{h-1}] \\
&\quad \le \gamma(\mathsf{d}_{\mathrm{TVC}}(\hat{\mathsf{s}}_h, \mathsf{s}_h^\star))] + \mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}_h^\star), \pi_h^\star(\mathsf{s}_h^\star))
\end{aligned} \tag{G.3}$$

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

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

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

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 $\hat{\mathsf{a}}_j^{\mathrm{inter}} = \hat{\mathsf{a}}_j$, $1 \le j \le h-1$, implies that

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

Thus, (G.3) implies that

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

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 (G.4) implies that

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

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

□

G.1 Relaxing the TVC Condition

In this section, we show our results hold with the following generalization of TVC,

Definition G.1 (Relaxed TVC)**.** We say $\hat{\pi}$ satisfies (γ, ε')-relaxed TVC if, for all h, and $\mathsf{s}, \mathsf{s}' \in \mathcal{S}$,

$$\inf_\mu \mathbb{P}_{(\mathsf{a},\mathsf{a}')}[\mathsf{d}_\mathcal{A}(\mathsf{a}, \mathsf{a}') > \varepsilon'] \le \gamma(\mathsf{d}_\mathcal{S}(\mathsf{s}, \mathsf{a}')),$$

where $\inf_\mu$ is the infimum over all couplings μ with $\mathsf{a} \sim \hat{\pi}_h(\mathsf{s})$ and $\mathsf{a}' \sim \hat{\pi}_h(\mathsf{s}')$.

The main result of this section is as follows.

Proposition G.1 (Generalization of Proposition D.2)**.** Let $\varepsilon > \varepsilon_1, \varepsilon_2 > 0$. Let $\pi^\star$ be input-stable w.r.t. $(\mathsf{d}_\mathcal{S}, \mathsf{d}_\mathcal{A})$ and let $\hat{\pi}$ be (γ, ε_1)-relaxed TVC. Further, suppose that $\mathsf{d}_\mathcal{A}$ (which need not satisfy the triangle inequality), satisfies

$$\{\mathsf{d}_\mathcal{A}(\mathsf{a}', \mathsf{a}) \le \varepsilon_2\} \text{ and } \{\mathsf{d}_\mathcal{A}(\mathsf{a}'', \mathsf{a}') \le \varepsilon_1\} \quad \text{implies} \quad \{\mathsf{d}_\mathcal{A}(, \mathsf{a}'', \mathsf{a}) \le \varepsilon\}, \tag{G.5}$$

for all $\mathsf{a}, \mathsf{a}', \mathsf{a}'' \in \mathcal{A}$ and all $\varepsilon_1, \varepsilon_2, \varepsilon$ given above. Then,

$$\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star) \le H\gamma(\varepsilon) + \sum_{h=1}^H \mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star} \mathsf{d}_{\mathrm{os},\varepsilon_2}(\hat{\pi}_h(\mathsf{s}_h^\star) \parallel \pi^\star(\mathsf{s}_h^\star)).$$

We remark that (G.5) holds the distance $\mathsf{d}_\mathcal{A}$ defined in Appendix D.2 whenever ε is sufficiently small, and $\varepsilon_1 + \varepsilon_2 = \varepsilon$.

Proof Sketch of Proposition G.1. The proof is nearly identical to the standard proof under (non-relaxed) TVC given above. The only difference is we replace the events $\{\mathsf{d}_\mathcal{A}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) > \varepsilon\}$ and $\{\hat{\mathsf{a}}_h \ne \hat{\mathsf{a}}_h^{\mathrm{inter}}\}$ with the events $\{\mathsf{d}_\mathcal{A}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) > \varepsilon_1\}$ and $\{\mathsf{d}_\mathcal{A}(\hat{\mathsf{a}}_h, \hat{\mathsf{a}}_h^{\mathrm{inter}}) > \varepsilon_2\}$. By (G.5), the intersection of the complement these events implies

$$\left\{\mathsf{d}_\mathcal{A}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \mathsf{a}_h^\star) \le \varepsilon_2\right\} \cap \left\{\mathsf{d}_\mathcal{A}(\hat{\mathsf{a}}_h, \hat{\mathsf{a}}_h^{\mathrm{inter}}) \le \varepsilon_1\right\} \quad \text{implies} \quad \left\{\mathsf{d}_\mathcal{A}(\hat{\mathsf{a}}_h, \mathsf{a}_h^\star) \le \varepsilon\right\},$$

which allows the same argument to be followed. □

Remark G.1 (Why Wasserstein continuity is not enough)**.** Given that relaxed TVC is enough, we may be tempted to believe that $\hat{\pi}$ need only satisfy a Wasserstein continuity condition with linear $\gamma(\cdot)$. However, this is not quick strong enough to imply our argument.

For simplicity, assume $\mathsf{d}_\mathcal{A}$ is a metric and uppose that we have $\hat{\pi}$ is continuous with respect to the p-Wasserstein distance with metric $\mathsf{d}_\mathcal{A}$. This means that, when $\mathsf{d}_\mathcal{S}(\mathsf{s}, \mathsf{s}') \le \varepsilon_0$, we can can hope by Markov's inequality that

$$\inf_\mu \mathbb{E}_\mu[\mathsf{d}_\mathcal{A}(\mathsf{a}, \mathsf{a}')^p]^{1/p} \le L\mathsf{d}_\mathcal{S}(\mathsf{s}, \mathsf{s}'),$$

where μ are the couplings of $\mathsf{a} \sim \hat{\pi}_h(\mathsf{s})$ and $\mathsf{a}' \sim \hat{\pi}(\mathsf{s}_h')$. Then, by Markov's inequality, the Most we have hope for is that

$$\inf_\mu \mathbb{P}_\mu[\mathsf{d}_\mathcal{A}(\mathsf{a}, \mathsf{a}') \ge tL\mathsf{d}_\mathcal{S}(\mathsf{s}, \mathsf{s}')] \le \frac{1}{t^p}.$$

In particular, using the argument from the proof of our proposition,

$$\inf_{\mu} \mathbb{P}_{\mu}[\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\mathrm{inter}}, \hat{\mathsf{a}}_h) \geq tL\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h^{\star}, \hat{\mathsf{s}}_h)] \leq \frac{1}{t^p}.$$

On the other hand, out stability assumption can only bound $\mathsf{d}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\star}, \hat{\mathsf{s}}_{h+1}) \leq \max_{1 \leq i \leq h} \mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_i^{\mathrm{inter}}, \hat{\mathsf{a}}_i)$. Thus, for $L > 1$, an any target ε, we can at most hope for bounds which scale as L^H in error. This is essentially the same issue tackled in Pfrommer et al. [50] by matching higher-order derivatives. What matching these derivatives in Wasserstein space would mean is still unclear.

H 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 4 that applies to vectorized notions of distance and helps tighten our bounds when instantiated in the control setting. In Appendix H.1, we introduce some notation and state our most general result, Theorem 8. We then proceed to show that Theorem 4 follows from Theorem 8 and in Appendix H.2, we provide a detailed and rigorous proof of the main result. In Appendix H.3, we show that the more general Theorem 8 impiles Theorem 4 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 o_h coordinates (observation chunk), not the full trajectory chunk s_h.

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

H.1 A generalization of Theorem 4

We now state a generalization of Theorem 4, 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 $\vec{\varepsilon} \in \mathbb{R}_{\geq 0}^K$ 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 G.1 as follows.

Assumption H.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 open and thus measurable.

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

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

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 D.5 in the body. We also restrict the kernels under consideration to those which produce distributions *absolutely continuous* (Definition F.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 $\mathcal{O}$ component of $\mathcal{S}$.

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

Consider any sequence of kernels $\mathsf{W}_h : \mathcal{S} \to \Delta(\mathcal{S})$, $1 \leq h \leq H$, satisfying
$$\forall h, \mathsf{s} \in \mathcal{S} : \quad \mathbb{P}_{\tilde{\mathsf{s}} \sim \mathsf{W}_h(\mathsf{s})}[\mathsf{d}_{\mathrm{IPS}}(\tilde{\mathsf{s}}, \mathsf{s}) \leq r] = 1, \quad \phi_{\mathrm{o}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathrm{o}} \circ \mathsf{P}^\star_h. \tag{H.1}$$

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 $1 - p_{\mathrm{IPS}}$,

(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 H.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 8. Suppose that there

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

(b) There is a direct decomposition of $\mathcal{S} = \mathcal{O} \oplus \mathcal{S}_{/\mathcal{O}}$, which associated projection maps ϕ_{o} and $\phi_{/\mathrm{o}}$, and which is compatible with the dynamics, and policies $\pi^\star$, $\hat{\pi}$, and smoothing kernel W_σ, and $\mathsf{d}_{\mathrm{IPS}}$.

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

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

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

Then,

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

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

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

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

We note that Theorem 4 is a special case of Theorem 8 and prove the former assuming the latter here at the end of the section.

H.2 Proof of Theorem 8

H.2.1 Proof Overview and Coupling Construction

We begin with an intuitive overview of the proof and partially construct the relevant intermediate trajectories used to define our coupling, after which we sketch the organization of the rest of Appendix H.2.

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 $\pi^\star_{\mathrm{rep}}$ by introducing several intermediate sequences of composite states and composite actions.

We partially specify this coupling below and formally construct it in Appendix H.2.4. 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 $\pi^\star_{\mathrm{rep}}$ and $\hat{\pi}$ are close in $\vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}$ 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 $(\hat{\mathsf{s}}_{1:H}, \mathsf{s}^{\mathrm{rep}}_{1:H}, \mathsf{s}^{\mathrm{tel}}_{1:H}, \tilde{\mathsf{s}}^{\mathrm{tel}}_{1:H}) \subset \mathcal{S}$, composite-actions $(\mathsf{a}^{\mathrm{rep}}_{1:H}, \hat{\mathsf{a}}_{1:h}, \mathsf{a}^{\mathrm{tel}}_{1:H}) \subset \mathcal{K}$ and interpolating composite-actions, $(\hat{\mathsf{a}}^{\mathrm{rep,inter}}_{1:H}, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_{1:H}) \subset \mathcal{A}$.

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

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

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

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

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

We remark that $\mathsf{W}^\star_{\mathrm{dec},h}$ and $\mathsf{W}^\star_{\mathrm{rep},h}$ are both kernels and by Theorem 7, 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 H.1. Note that the kernels $\mathsf{W}^\star_{\mathrm{dec},h}$ and $\mathsf{W}^\star_{\mathrm{rep},h}$ are compatible with the decomposition $\mathcal{S} = \mathcal{O} \oplus \mathcal{S}_{/\mathcal{O}}$ by construction. Moreover, note that if $\mathsf{s} = (\mathsf{o}, \mathsf{v})$, $\phi_{/\mathrm{o}} \circ \mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}) = \phi_{/\mathrm{o}} \circ \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s})$ is the dirac-delta distribution supported on v.

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

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

Proof. This follows imediately because $\pi^\star$ and W_σ are compatile with the direct decomposition, and by the definition of Definition D.4. $\square$

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 generatived by $\hat{\mathsf{s}}_1 = \mathsf{s}^{\mathrm{rep}}_1 = \mathsf{s}^{\mathrm{tel}}_1$. Let $\mathcal{F}_h$ denote the sigma-algebra generated by $(\hat{\mathsf{s}}_{1:h}, \mathsf{s}^{\mathrm{rep}}_{1:h}, \mathsf{s}^{\mathrm{tel}}_{1:h})$, $(\mathsf{a}^{\mathrm{rep}}_{1:h}, \tilde{\mathsf{s}}^{\mathrm{rep}}_{1:h}, \tilde{\mathsf{s}}^{\mathrm{tel}}_{1:h}, \mathsf{a}^{\mathrm{tel}}_{1:h}, \hat{\mathsf{a}}_{1:h})$, and $(\hat{\mathsf{a}}^{\mathrm{rep,inter}}_{1:h}, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_{1:h})$.

- The initial states are drawn as
$$\hat{\mathsf{s}}_1 = \mathsf{s}^{\mathrm{rep}}_1 = \mathsf{s}^{\mathrm{tel}}_1 \sim \mathsf{P}_{\mathrm{init}}.$$
- The dynamics satisfy
$$\hat{\mathsf{s}}_{h+1} = F_h(\hat{\mathsf{s}}_h, \hat{\mathsf{a}}_h), \quad \mathsf{s}^{\mathrm{rep}}_{h+1} = F_h(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{a}^{\mathrm{rep}}_h), \quad \mathsf{s}^{\mathrm{tel}}_{h+1} = F_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h, \mathsf{a}^{\mathrm{tel}}_h)$$
Note that determinism of the dynamics implies that $\mathsf{s}^{\mathrm{tel}}_{h+1}$, $\mathsf{s}^{\mathrm{rep}}_{h+1}$ and $\hat{\mathsf{s}}_{h+1}$ are $\mathcal{F}_h$-measurable.
- We generate
$$\begin{aligned}
&\tilde{\mathsf{s}}^{\mathrm{rep}}_h \mid \mathcal{F}_{h-1} \sim \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s}^{\mathrm{rep}}_h), \quad \mathsf{a}^{\mathrm{rep}}_h \mid \mathcal{F}_{h-1}, \tilde{\mathsf{s}}^{\mathrm{rep}}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\mathrm{rep}}_h),\\
&\tilde{\mathsf{s}}^{\mathrm{tel}}_h \mid \mathcal{F}_{h-1} \sim \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s}^{\mathrm{tel}}_h), \quad \mathsf{a}^{\mathrm{tel}}_h \mid \mathcal{F}_{h-1}, \tilde{\mathsf{s}}^{\mathrm{tel}}_h \sim \pi^\star_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h).\\
&\hat{\mathsf{a}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_{\sigma,h}(\hat{\mathsf{s}}_h)
\end{aligned}$$
Importantly, we note that, marginalizing over $\tilde{\mathsf{s}}^{\mathrm{tel}}_h$ and $\tilde{\mathsf{s}}^{\mathrm{rep}}_h$, respectively, $\mathsf{a}^{\mathrm{tel}}_h \mid \mathcal{F}_{h-1} \sim \pi^\star_{\mathrm{rep},h}(\mathsf{s}^{\mathrm{tel}})$ and $\mathsf{a}^{\mathrm{rep}}_h \mid \mathcal{F}_{h-1} \sim \pi^\star_{\mathrm{rep},h}(\mathsf{s}^{\mathrm{rep}}_h)$.
- Lastly, we select interpolating actions via
$$\hat{\mathsf{a}}^{\mathrm{rep,inter}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{rep}}_h), \qquad \hat{\mathsf{a}}^{\mathrm{tel,inter}}_h \mid \mathcal{F}_{h-1} \sim \hat{\pi}_{\sigma,h}(\mathsf{s}^{\mathrm{tel}}_h)$$

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 10. We now establish several key properties of the above constructions, separated into a subsection for the sake of clarity.

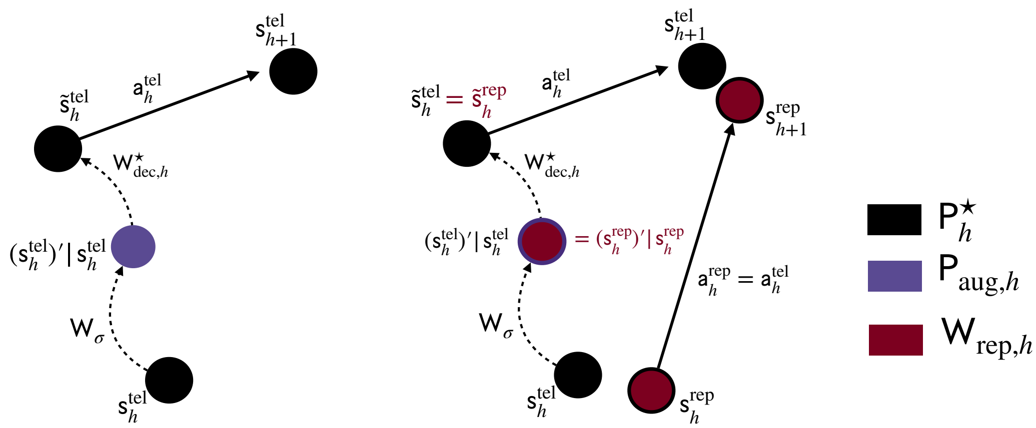


Figure 10: Graphical illustration of the coupling, in the special case where $\mathcal{O} = \mathcal{S}$ for simplicity. **On the left** is the teleporting sequence, with $\tilde{\mathsf{s}}^{\mathrm{tel}} \sim \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s}^{\mathrm{tel}}_h) = \mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{W}_\sigma(\mathsf{s}^{\mathrm{tel}}_h)$. 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 $\mathsf{s}^{\mathrm{rep}}_h$, which is close to $\mathsf{s}^{\mathrm{tel}}_h$ (a consequence of our proof). We then apply the replica kernel to achieve $\tilde{\mathsf{s}}^{\mathrm{rep}}_h$. Our argument uses that $\mathsf{W}^\star_{\mathrm{rep},h} = \mathsf{W}^\star_{\mathrm{dec},h} \circ \mathsf{W}_\sigma$ is TVC (a consequence of TVC of W_σ as shown in Lemma H.2). We depict this property pictorially: since W_σ is TVC and $\mathsf{s}^{\mathrm{tel}}_h$ and $\mathsf{s}^{\mathrm{rep}}_h$ 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 $(\mathsf{s}^{\mathrm{rep}}_h)' \sim \mathsf{W}_\sigma(\mathsf{s}^{\mathrm{rep}}_h)$ are equal. We then extend the coupling to that $\tilde{\mathsf{s}}^{\mathrm{rep}}_h = \tilde{\mathsf{s}}^{\mathrm{tel}}_h$ on the event $\{(\mathsf{s}^{\mathrm{tel}}_h)' = (\mathsf{s}^{\mathrm{rep}}_h)'\}$, both being drawn by applying $\mathsf{W}^\star_{\mathrm{dec},h}$ to both of $(\mathsf{s}^{\mathrm{tel}}_h)' = (\mathsf{s}^{\mathrm{rep}}_h)'$. We extend the coupling once more so that $\mathsf{a}^{\mathrm{tel}}_h \sim \pi^\star(\tilde{\mathsf{s}}^{\mathrm{tel}}_h)$ and $\mathsf{a}^{\mathrm{rep}}_h \sim \pi^\star(\tilde{\mathsf{s}}^{\mathrm{rep}}_h)$ are equal on this good probability event. Using our notion of stability, IPS, and the fact that $\mathsf{s}^{\mathrm{rep}}_h$ and $\mathsf{s}^{\mathrm{tel}}_h$ are close, the good probability event on which $\mathsf{a}^{\mathrm{tel}}_h$ and $\mathsf{a}^{\mathrm{rep}}_h$ are equal implies that $\mathsf{s}^{\mathrm{rep}}_{h+1}$ 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 $\mathsf{W}^\star_{\mathrm{rep},h}$ and the verification of its TVC property.

Organization of the remaining parts of Appendix H.2. In Appendix H.2.2, we prove several prerequisite properties of the construction given above, including concentration of the smoothing kernel, and key properties of the replica distribution. Next, Appendix H.2.3 shows that, due to these properties of the replica distribution, we can bound the marginal imitation gap by controlling the tracking of the teleporting sequence constructed above. Finally, in Appendix H.2.4 we formally construct the coupling and rigorously prove Theorem 8.

H.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 H.2. The following hold

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

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

Next, we show that the replica construction preserves marginals.

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

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

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

Proof. By Assumption G.1 and Theorem 7, 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 $\mathcal{O} = \mathcal{S}$. Fix a common measure $\mathbb{P}$ over which $\mathsf{s}^{\mathrm{tel}}_h, \tilde{\mathsf{s}}^{\mathrm{tel}}_h$, and s'_h are defined such that $\mathsf{s}^{\mathrm{tel}}_h \sim \mathsf{P}^\star_h$, $\mathsf{s}'_h \sim \mathsf{W}_\sigma(\mathsf{s}^{\mathrm{tel}}_h)$, and $\tilde{\mathsf{s}}^{\mathrm{tel}}_h \sim \mathsf{W}_{\mathrm{dec},h}(\mathsf{s}'_h)$. Then for any measurable sets A, B, we have

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

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 $\mathcal{O} = \mathcal{S}$ as in the proof of the first statement. Note that $\mathsf{s}^{\mathrm{tel}}_1 \sim \mathsf{P}^\star_1 = \mathsf{P}_{\mathrm{init}}$, and $\tilde{\mathsf{s}}^{\mathrm{tel}}_1 \sim \mathsf{W}^\star_{\mathrm{rep},1} \circ \mathsf{P}^\star_1$. Thus, from the first part of the lemma, $\phi_{\mathrm{o}}(\mathsf{s}^{\mathrm{tel}}_1) \sim \phi_{\mathrm{o}} \circ \mathsf{P}^\star_1$. Now, suppose the induction holds up to step h. Then, $\tilde{\mathsf{s}}^{\mathrm{tel}}_h \sim \mathsf{P}^\star_h$, as $\mathsf{a}^{\mathrm{tel}}_h \sim \pi^\star_h(\mathsf{a}^{\mathrm{tel}}_h)$, then $\mathsf{s}^{\mathrm{tel}}_{h+1} = F_h(\tilde{\mathsf{s}}^{\mathrm{tel}}_h, \mathsf{a}^{\mathrm{tel}}_h) \sim \mathsf{P}^\star_{h+1}$. Again $\tilde{\mathsf{s}}^{\mathrm{tel}}_{h+1} \sim \mathsf{W}^\star_{\mathrm{rep},h+1}(\mathsf{s}^{\mathrm{tel}}_{h+1})$, so that $\tilde{\mathsf{s}}^{\mathrm{tel}}_{h+1}$ has marginal $\mathsf{W}^\star_{\mathrm{rep},h+1} \circ \mathsf{P}^\star_{h+1} = \mathsf{P}^\star_{h+1}$, as needed. ∎

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

Lemma H.4. The kernel $\mathsf{W}_{\mathrm{rep},h}$ satisfies that $\phi_{\mathrm{o}} \circ \mathsf{W}_{\mathrm{rep},h} \ll \phi_{\mathrm{o}} \circ \mathsf{P}^\star_h$ as laws, validating the second condition in (H.1). It further holds that $\phi_{\mathrm{o}} \circ \mathsf{W}_{\mathrm{dec},h} \ll \phi_{\mathrm{o}} \circ \mathsf{P}^\star_h$.

Proof. The first statement follows immediately from Lemma H.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 H.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

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

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

Let $\mathbb{P}$ denote a distribution over $\mathsf{s}_h \sim \mathsf{P}^\star_h$, $\mathsf{s}'_h \sim \mathsf{W}_\sigma(\mathsf{s}_h)$, and $\tilde{\mathsf{s}}_h \sim \mathsf{W}^\star_{\mathrm{dec},h}(\mathsf{s}'_h)$. In this special case, we see that $\tilde{\mathsf{s}}_h \mid \mathsf{s}_h \sim \mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s}_h)$[9]. By a union bound,

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

where the equality follows from the first statment of Lemma H.3. ∎

[9]Notice that, for general $\mathcal{S} = \mathcal{O} \oplus \mathcal{S}_{/\mathcal{O}}$, this condition would become $\phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h) \mid \phi_{\mathrm{o}}(\mathsf{s}_h) \sim \phi_{\mathrm{o}} \circ \mathsf{W}^\star_{\mathrm{rep},h}(\phi_{\mathrm{o}}(\mathsf{s}_h), \cdot)$, where the $\cdot$ argument is irrelevant.

Remark H.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.

H.2.3 Bounding the marginal imitation gaps in terms of the teleporting sequence error

Before turning to the proof of Theorem 8, we verify that closeness to the *teleporting sequences* suffices to control error in marginal gap to $\pi^\star$. The key property here is that the teleporting sequence, as shown in Lemma H.3, has the same marginal distribution over states as does $\pi^\star$.

Lemma H.6. Let μ be any coupling obeying the construction of the couplings above. Then,

$$\vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star) \le \mathbb{P}_\mu\left[\exists h\in[H] : \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\mathrm{tel}},\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right\} \cup \left\{\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\mathrm{tel}},\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right\}\right]$$

Proof. We begin with a (reverse) union bound.

$$\begin{aligned}
&\mathbb{P}_\mu\left[\exists h\in[H] : \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\mathrm{tel}},\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right\} \cup \left\{\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\mathrm{tel}},\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right\}\right] \\
&\ge \max_h \max\left\{\mathbb{P}_\mu\left[\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\mathrm{tel}},\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right], \mathbb{P}_\mu\left[\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\mathrm{tel}},\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right]\right\}.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{P}_\mu\left[\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\mathrm{tel}},\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right] &\ge \inf_{\mu_1}\mathbb{P}\left[\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^\star,\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right] \\
\mathbb{P}_\mu\left[\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\mathrm{tel}},\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right] &\ge \inf_{\mu_1}\mathbb{P}\left[\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{a}_h^\star,\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right],
\end{aligned}$$

where the $\inf_{\mu_1}$ is, as in Definitions D.1 and H.2, the infinum over couplings between $\mathsf{D}_{\pi^\star}$ and $\mathsf{D}_{\hat{\pi}}$. Thus,

$$\begin{aligned}
&\mathbb{P}_\mu\left[\exists h\in[H] : \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\mathrm{tel}},\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right\} \cup \left\{\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\mathrm{tel}},\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right\}\right] \\
&\ge \max_h \max\left\{\inf \mathbb{P}_{\mu_1}\left[\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^\star,\hat{\mathsf{s}}_{h+1}) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right], \inf_\mu \mathbb{P}_{\mu_1}\left[\vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^\star,\hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\mathrm{marg}}\right]\right\} \\
&:= \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star).
\end{aligned}$$

□

H.2.4 Formal proof of Theorem 8

We now proceed to formally prove Theorem 8

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 H.6 (Coupling Events). Define the events

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

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

Definition H.7 (Inductive Event). Define the events

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

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

Definition H.8 (Stability Events). Define the events

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

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

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

Lemma H.7 (Stability Claim). By construction,

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

Proof. It suffices to show that on $\mathcal{Q}_{\text{all}} \cap \bar{\mathcal{C}}_{\text{all},h} \cap \bar{\mathcal{B}}_{\text{all},h}$, $\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\text{rep}}, \hat{\mathsf{s}}_{h+1}) \preceq \vec{\varepsilon}$ and $\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\text{rep}}, \mathsf{s}_{h+1}^{\text{tel}}) \preceq \vec{\gamma}_{\text{IPS}}(2r)$. 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

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

For the next point, note that the compatibility of the dynamics with the direct decomposition $\mathcal{S} = \mathcal{O} \oplus \mathcal{S}_{/\mathcal{O}}$ implies that there exists a dynamics map $F_{\text{o},h}$ for which

$$F_h(\mathsf{s}, \mathsf{a}) = F_{\text{o},h}(\phi_{\text{o}}(\mathsf{s}), \mathsf{a}).$$

Similarly, by applying $\mathcal{Q}_{\text{IPS}}$ and $\mathcal{Q}_{\text{close}}$ and the event $\{\phi_{\text{o}}(\tilde{\mathsf{s}}_h^{\text{rep}}) = \phi_{\text{o}}(\tilde{\mathsf{s}}_h^{\text{tel}}), \mathsf{a}_h^{\text{tel}} = \mathsf{a}_h^{\text{rep}}\}$ 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 $h \in [H]$,

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

□

From Lemma H.7, we decompose our error probability as follows:

Lemma H.8 (Key Error Decomposition)**.** Let μ respect the construction (in the sense of Appendix H.2.1). Then, for any coupling μ which respects the construction,

$$
\vec{\Gamma}_{\text{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\text{rep}}) \vee \vec{\Gamma}_{\text{marg},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star) \le \mathbb{P}_\mu[\mathcal{Q}_{\text{all}}^c] + \sum_{h=1}^{H} \mathbb{P}_\mu[\bar{\mathcal{B}}_{\text{all},h}^c \cap \bar{\mathcal{C}}_{\text{all},h} \cap \bar{\mathcal{B}}_{\text{all},h-1}] \tag{H.4}
$$

Proof. In what follows, we use $\vec{v} \vee \vec{w}$ to denote the entrywise maximum of two vectors of the same dimension. 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 $h \in [H]$

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

On $\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H$, we have that

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

Thus, by the triangle inequality and $\vec{\varepsilon}_{\text{marg}} = \vec{\varepsilon} + \vec{\gamma}_{\text{IPS}}(2r)$, on $\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H$,

$$
\max_h \vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_h^{\text{rep}}, \mathsf{s}_h^{\text{tel}}) \le \vec{\varepsilon}_{\text{marg}}, \quad \text{and} \quad \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\text{tel}}, \hat{\mathsf{a}}_h) = \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\text{rep}}, \hat{\mathsf{a}}_h) \le \vec{\varepsilon} \le \vec{\varepsilon}_{\text{marg}}.
$$

Thus,

$$
\begin{aligned}
&\mathbb{P}_\mu\left[\exists h \in [H] : \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\text{rep}}, \hat{\mathsf{s}}_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\text{rep}}, \hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}\right\} \cup \left\{\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\text{tel}}, \hat{\mathsf{s}}_{h+1}) \vee \vec{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}_h^{\text{tel}}, \hat{\mathsf{a}}_h) \not\preceq \vec{\varepsilon}_{\text{marg}}\right\}\right] \\
&\le \mathbb{P}_\mu[(\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H)^c]
\end{aligned} \tag{H.5}
$$

In particular, this shows that

$$
\vec{\Gamma}_{\text{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\text{rep}}) \le \mathbb{P}_\mu[(\mathcal{Q}_{\text{all}} \cap \mathcal{E}_H)^c],
$$

and similarly, by Lemma H.6,

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

As $(\mathsf{s}_{1:H+1}^{\text{rep}}, \mathsf{a}_{1:H}^{\text{rep}}) \sim \mathsf{D}_{\pi^\star_{\text{rep}}}$, (H.5) shows that

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

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

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

where the last step invokes Lemma H.7. □

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

Lemma H.9. For all μ which respect the construnction,

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

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

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

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

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

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

Lemma H.10. There exists a coupling μ which respects the construction and satisfies the following for any $h \in [H]$

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

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

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

Moreover, $\mathsf{s} \mapsto \vec{\mathsf{d}}_{\text{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}) \parallel \pi_{\text{rep},h}^\star(\mathsf{s}))$ 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 F.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:

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

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 H.2.5. □

Concluding the proof. Here, we finish the proof of Theorem 8. Recall that we wish to bound $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\mathrm{rep}}) \vee \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star)$. We begin by bounding $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi}_\sigma \parallel \pi^\star_{\mathrm{rep}}) \vee \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}_{\mathrm{marg}}}(\hat{\pi}_\sigma \parallel \pi^\star_{\mathrm{rep}})$. In light of Lemma H.8, 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 H.10. Applying Lemma H.9 and Lemma H.10,

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

To conclude, we note that for any μ which respects the construction, Lemma H.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

$$p_{\mathrm{IPS}} + H(2p_r + \hat{\gamma}(\vec{\varepsilon}_1) + (\hat{\gamma} + \gamma_\sigma) \circ \gamma_{\mathrm{IPS,TVC}}(2r)) + \sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \vec{\mathsf{d}}_{\mathrm{os},\vec{\varepsilon}}(\hat{\pi}_{\sigma,h}(\mathsf{s}^\star_h) \parallel \pi^\star_{\mathrm{rep},h}(\mathsf{s}^\star_h)) \tag{H.6}$$

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

To prove (H.3) for $\vec{\Gamma}_{\mathrm{joint},\vec{\varepsilon}}(\hat{\pi} \parallel \pi^\star_{\mathrm{rep}})$, 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 F.5)

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

for all $\mathsf{s}^\star_h$. Substituting this into (H.6), and setting $\hat{\gamma} = \gamma_\sigma$ (in view of Lemma H.2), finishes the argument.

H.2.5 Proof of Lemma H.10

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

Proof strategy. Our proof follows along similar lines as that of Proposition D.2, 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:

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

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 F.2 to glue the individual couplings together. We will then use Lemma F.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 $h \ge 1$, and suppose that we have constructed the coupling $\mu^{(1:h-1)}$ for steps $1, \dots, h-1$ which respects the construction. Recall that $\mathcal{F}_h$ denotes the sigma-algebra generated by $(\hat{\mathsf{s}}_{1:h}, \mathsf{s}^{\mathrm{rep}}_{1:h}, \mathsf{s}^{\mathrm{tel}}_{1:h})$, $(\mathsf{a}^{\mathrm{rep}}_{1:h}, \tilde{\mathsf{s}}^{\mathrm{rep}}_{1:h}, \tilde{\mathsf{s}}^{\mathrm{tel}}_{1:h}, \mathsf{a}^{\mathrm{tel}}_{1:h}, \hat{\mathsf{a}}_{1:h})$, and $(\hat{\mathsf{a}}^{\mathrm{rep,inter}}_{1:h}, \hat{\mathsf{a}}^{\mathrm{tel,inter}}_{1:h})$. Notice that $\mathsf{s}^{\mathrm{tel}}_{h+1}, \mathsf{s}^{\mathrm{rep}}_{h+1}, \hat{\mathsf{s}}_{h+1}$ are determined by $\mathcal{F}_h$ as well. Similarly, it can be seen from Definition H.5 that $\phi_{/\mathrm{o}}(\tilde{\mathsf{s}}^{\mathrm{tel}}_{h+1})$ and $\phi_{/\mathrm{o}}(\tilde{\mathsf{s}}^{\mathrm{rep}}_{h+1})$ are also determined by $\mathcal{F}_h$ (since the replica kernel preserves the

$\mathcal{S}_{/\mathcal{O}}$-components). We summarize all these aforementioned variables in a random variable Y_h. Let $\mathcal{F}_0$ denote the filtration generated by $\mathsf{s}_1^{\mathrm{rep}} = \mathsf{s}_1^{\mathrm{tel}} = \hat{\mathsf{s}}_1$. We let $Y_0 = (\mathsf{s}_1^{\mathrm{rep}}, \mathsf{s}_1^{\mathrm{tel}}, \hat{\mathsf{s}}_1)$.

Correspondingly, let Z_h denote the random variables $(\mathsf{a}_h^{\mathrm{rep}}, \phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h^{\mathrm{rep}}), \phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}), \mathsf{a}_h^{\mathrm{tel}}, \hat{\mathsf{a}}_h)$, and $(\hat{\mathsf{a}}_h^{\mathrm{rep,inter}}, \hat{\mathsf{a}}_h^{\mathrm{tel,inter}})$ 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 $\phi_{\mathrm{o}} \circ \mathsf{W}^\star_{\mathrm{rep},h}$ (Lemma H.2),

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

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

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

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

  $$\mathbb{E}_{\mu_{\mathrm{tel}}^{(h)}}[\mathcal{B}_{\mathrm{tel},h} \mid Y_{h-1}] = \mathbb{P}_{\mu_{\mathrm{tel}}^{(h)}}[(\phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h^{\mathrm{rep}}), \mathsf{a}_h^{\mathrm{rep}}) \ne (\phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}), \mathsf{a}_h^{\mathrm{tel}}) \mid Y_{h-1}] \le \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h^{\mathrm{rep}}, \mathsf{s}_h^{\mathrm{tel}})].$$

- In our construction, $\mathsf{a}_h^{\mathrm{tel}} \mid Y_{h-1} \sim \pi^\star_{\mathrm{rep},h}(\mathsf{s}_h^{\mathrm{tel}})$, and $\hat{\mathsf{a}}_h^{\mathrm{tel,inter}} \mid Y_{h-1} \sim \hat{\pi}_{\sigma,h}(\mathsf{s}_h^{\mathrm{tel}})$. 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 F.3 implies that we may find a coupling $\mu_{\mathrm{est}}^{(h)}$ of $(\mathsf{a}_h^{\mathrm{tel}}, \hat{\mathsf{a}}_h^{\mathrm{tel,inter}}, Y_{h-1})$ respecting the construction such that, almost surely,

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

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

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

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

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

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

- Similarly, since $\hat{\mathsf{a}}_h^{\mathrm{rep,inter}} \mid \mathcal{F}_{h-1} \sim \hat{\pi}_h(\mathsf{s}_h^{\mathrm{rep}})$ 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 F.1 implies that there is a coupling $\mu_{\hat{\mathsf{a}}}^{(h)}$ between $(\hat{\mathsf{a}}_h^{\mathrm{rep,inter}}, \hat{\mathsf{a}}_h, Y_{h-1})$ such that

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

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

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

with

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

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

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

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, $\vec{\mathsf{d}}_{\mathcal{S}}(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{s}^{\mathrm{tel}}_h) \preceq \vec{\gamma}_{\mathrm{IPS}}(2r)$ and $\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\mathrm{rep}}_h, \hat{\mathsf{s}}_h) \preceq \vec{\varepsilon}$. For our purposes, we use $\mathsf{d}_{\textsc{tvc}} = \mathsf{d}_{\mathcal{S},1}(\mathsf{s}^{\mathrm{rep}}_h, \mathsf{s}^{\mathrm{tel}}_h) \preceq \gamma_{\textsc{ips},\textsc{tvc}}(2r)$ and $\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\mathrm{rep}}_h, \hat{\mathsf{s}}_h) \preceq \vec{\varepsilon}_1$. Hence,

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

The result follows.

H.3 Proof of Theorem 4, and generalization to direct decompositions

In this subsection, we consider the special case dealt with in Theorem 4. Note that there always exists a trivial direct decomposition that is compatible with all policies and dynamics simply by letting $\mathcal{S}_{/\mathcal{O}} = \emptyset$ and $\mathcal{S} = \mathcal{O}$. 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 $\mathcal{O} = \{\mathsf{o}_h\}$ and $\mathcal{S} = \{\mathsf{s}_h\}$, i.e., projecting s_h onto the last τ_{obs} coordinates gives o_h. We further consider a restriction of IPS to consider kernels absolutely continuous with respect to $\mathsf{P}^\star_h$ in their $\mathcal{O}$ component.

Definition H.9 (Restricted IPS)**.** For a non-decreasing maps $\gamma_{\textsc{ips},\textsc{tvc}}, \gamma_{\textsc{ips},\mathcal{S}} : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ 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 $(\gamma_{\textsc{ips},\textsc{tvc}}, \gamma_{\textsc{ips},\mathcal{S}}, \mathsf{d}_{\textsc{ips}}, r_{\textsc{ips}})$-*restricted IPS* if the following holds for any $r \in [0, r_{\textsc{ips}}]$. Consider any sequence of kernels $\mathsf{W}_1, \dots, \mathsf{W}_H : \mathcal{S} \to \Delta(\mathcal{S})$ satisfying

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

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 (in the sense of Definition D.2) (b) $\max_{h\in[H]} \mathsf{d}_{\textsc{tvc}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \le \gamma_{\textsc{ips},\textsc{tvc}}(r)$ and (c) $\max_{h\in[H]} \mathsf{d}_{\mathcal{S}}(F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h), \mathsf{s}_{h+1}) \le \gamma_{\textsc{ips},\mathcal{S}}(r)$.

Note that the above is a slightly weaker condition than the one in Definition D.5 in the main text and consequently, the following theorem which uses it as an assumption implies Theorem 4 in the body.

Theorem 9. Suppose that

- $\mathcal{S} = \mathcal{O} \oplus \mathcal{S}_{/\mathcal{O}}$ is in Definition H.1 and projections $\phi_{\mathrm{o}}, \phi_{/\mathrm{o}}$, which is compatible with the dynamics and with given policies $\hat{\pi}, \pi^\star$, smoothing kernel W_σ, and pseudometric $\mathsf{d}_{\textsc{ips}}$.
- $\pi^\star$ satisfies $(\gamma_{\textsc{ips},\textsc{tvc}}, \gamma_{\textsc{ips},\mathcal{S}}, \mathsf{d}_{\textsc{ips}}, r_{\textsc{ips}})$-restricted IPS (Definition H.9) and $\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma$ is γ_σ-TVC.

Given $\varepsilon > 0$ and $r \in (0, \frac{1}{2} r_{\textsc{ips}}]$, define

$$p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\textsc{ips}}(\mathsf{s}', \mathsf{s}) > r], \quad \varepsilon' := \varepsilon + \gamma_{\textsc{ips},\mathcal{S}}(2r)$$

Then, for any policy $\hat{\pi}$, both $\Gamma_{\mathrm{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\mathrm{rep}})$ and $\Gamma_{\mathrm{marg},\varepsilon'}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ are upper bounded by

$$H\left(2p_r + 3\gamma_\sigma\left(\max\left\{\varepsilon, \gamma_{\mathrm{IPS},\mathrm{TVC}}(2r)\right\}\right)\right) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\varepsilon}\left(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h)\right). \tag{H.7}$$

Proof. Consider the special case $K = 2$ with $\mathsf{d}_{\mathcal{S},1} = \mathsf{d}_{\mathrm{TVC}}$, $\mathsf{d}_{\mathcal{S},2} = \mathsf{d}_{\mathcal{S}}$, $\mathsf{d}_{\mathcal{A},1} = \mathsf{d}_{\mathcal{A},2} = \mathsf{d}_{\mathcal{A}}$, $p_{\mathrm{IPS}} = 0$ and $\vec{\varepsilon} = (\varepsilon, \varepsilon)$. In this case, applying (H.3) in Theorem 8, we see that

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

We now observe that under this convention,

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

and similarly $\Gamma_{\mathrm{marg},\varepsilon'}(\hat{\pi}_\sigma \parallel \pi^\star) \le \vec{\Gamma}_{\mathrm{marg},\vec{\varepsilon}+\gamma_{\mathrm{IPS}}(2r)}(\hat{\pi}_\sigma \parallel \pi^\star)$. 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 $h \in [H]$,

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

Plugging in to (H.3) concludes the proof. □

I Lower Bounds

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

- In Appendix I.1 we show that Theorem 4 and Proposition D.2 are sharp in the regime where $\gamma_{\mathrm{IPS},\mathrm{TVC}} = \gamma_{\mathrm{IPS},\mathcal{S}} = 0$.
- In Appendix I.2, we show that the marginals of an expert policy $\pi^\star$ and replica policy $\pi^\star_{\mathrm{rep}}$ can coincide, but their joint distributions can be different. By considering $\hat{\pi} = \pi^\star_{\mathrm{dec}}$ in Theorem 4, this establishes the necessity of considering the marginal imitation gap with respect to $\pi^\star$.
- In Appendix I.3, we lower bound the distance between *marginal distributions* over states under $\pi^\star$ and $\pi^\star_{\mathrm{rep}}$ in the regime where $\gamma_{\mathrm{IPS},\mathcal{S}} \neq 0$. This example demonstrates that the dependence of $\gamma_{\mathrm{IPS},\mathcal{S}}$ in Theorem 4 is essentially sharp.
- In Appendix I.4, we show that for an expert policy $\pi^\star$ and smoothing kernel W_σ, the state distributions under $\pi^\star_{\mathrm{rep}}$ and $\pi^\star_{\mathrm{dec}}$ can have different marginals (and thus different joint distributions). By considering $\hat{\pi} = \pi^\star_{\mathrm{dec}}$ in Theorem 4, 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 Appendix D are sharp. We do not, however, establish necessity of $\gamma_{\mathrm{IPS},\mathrm{TVC}}$ in the interest of brevity; we believe this quantity is necessary. Still, the $\gamma_{\mathrm{IPS},\mathrm{TVC}}$ term contributes a factor exponentially small in τ_{chunk} in Theorem 3, so we deem lower bounds establishing its necessity of lesser importance.

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

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

which is the archetypal Polish space [22]. Throughout, we use δ_x to denote the dirac-delta distribution on $x \in \mathbb{R}$. We let $\mathsf{d}_{\mathcal{S}}(\mathsf{s}', \mathsf{s}) = \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}', \mathsf{s}) = |\mathsf{s}' - \mathsf{s}|$ and $\mathsf{d}_{\mathcal{A}}(\mathsf{a}', \mathsf{a}) = |\mathsf{a}' - \mathsf{a}|$ all be the Euclidean distance.

I.1 Sharpness of Proposition D.2 and Theorem 4

Here, we demonstrate that Proposition D.2 is tight up to constant factors, and that Theorem 4 is tight up to the terms $\gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}$ 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 $\gamma_{\mathrm{IPS,TVC}}(\cdot) = \gamma_{\mathrm{IPS},\mathcal{S}}(\cdot) \equiv 0$. Furthermore, let us assume policies do not depend on time index h. Let $\pi^\star : \mathsf{s} \to \delta_0$ be deterministic, and let $\mathsf{P}_{\mathrm{init}} = \delta_0$ 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 I.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) $\pi^\star$, $\pi^\star_{\mathrm{rep}}$ and $\pi^\star_{\mathrm{dec}}$ all map $\mathsf{s} \to \delta_0$, $\mathsf{P}^\star_h = \delta_0$, and thus for any $\tilde{\pi} \in \{\pi^\star, \pi^\star_{\mathrm{rep}}, \pi^\star_{\mathrm{dec}}\}$,

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

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

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

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

and the same holds with $\pi^\star$ replaced by $\pi^\star_{\mathrm{rep}}$ or $\pi^\star_{\mathrm{dec}}$.

In particular, the above proposition shows that

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

verifying the sharpness of Proposition D.2 (note that $\bar{\pi} = \hat{\pi}_{\varepsilon,p,\sigma} \circ \mathsf{W}_\sigma$ is γ_σ TVC). Similary, our above proposition shows that,

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

verying that Theorem 4 is sharp up to the additional stability terms $\gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}$.

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

$$\bar{\pi} = \hat{\pi}_{\varepsilon,p,\sigma} \circ \mathsf{W}_{\sigma,\varepsilon} = \begin{cases} (1-p)\updelta_\varepsilon + p\updelta_\sigma & \mathsf{s} \le \frac{\varepsilon}{2} \\ \updelta_\sigma & \mathsf{s} > \frac{\varepsilon}{2} \end{cases} \circ \begin{cases} \updelta_0 & \mathsf{s} \le 0 \\ (1-\frac{s}{\sigma})\updelta_0 + \frac{s}{\sigma}\updelta_\sigma & \mathsf{s} \in [0,\sigma] \\ \updelta_\sigma & \mathsf{s} \ge \sigma. \end{cases}$$

$$= \begin{cases} (1-p)\updelta_\varepsilon + p\updelta_\sigma & \mathsf{s} \le 0 \\ (1-\eta(s))\updelta_\varepsilon + \eta(s)\updelta_\sigma & 0 \le \mathsf{s} \le \sigma \\ \updelta_\sigma & \mathsf{s} > \sigma. \end{cases} \tag{I.1}$$

In particular,

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

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

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

Hence, as $\bar{\pi}(0) = \hat{\pi}_{\varepsilon,p,\sigma}(0) = (1-p)\updelta_\varepsilon + p\updelta_\sigma$, 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,

$$\mathsf{TV}(\mathsf{W}_{\varepsilon,\sigma}(\mathsf{s}), \mathsf{W}_{\varepsilon,\sigma}(\mathsf{s}')) = \mathsf{TV}(\updelta_0(1-\frac{\mathsf{s}}{\sigma}) + (\frac{\mathsf{s}}{\sigma})\updelta_\sigma, \updelta_0(1-\frac{\mathsf{s}'}{\sigma}) + (\frac{\mathsf{s}'}{\sigma})\updelta_\sigma = \frac{|\mathsf{s}'-\mathsf{s}|}{\sigma},$$

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: $\mathcal{A}^c_{1,h} = \mathcal{A}_{2,h}$. Moreover, from (I.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,

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

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

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

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

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

We give a simple example where $\pi^\star_{\mathrm{rep}}$ 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 9 in Appendix E.2.

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

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

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,

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

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

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

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

$$\mathbb{P}_{\mathsf{a}\sim\pi^\star_{\mathrm{rep},h}(\mathsf{s})}[\mathsf{a} = -\mathsf{s}] \ge \Omega(1). \tag{I.3}$$

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

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

where in the last step we used (I.3) and the the fact that the $\pi^\star_{\mathrm{rep}}$ 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 $\mu\in\mathscr{C}(\mathsf{D}_{\pi^\star},\mathsf{D}_{\pi^\star_{\mathrm{rep}}})$ over $(\mathsf{s}^\star_{1:H+1},\mathsf{a}^\star_{1:H})\sim\mathsf{D}_{\pi^\star}$ and $(\mathsf{s}^{\mathrm{rep}}_{1:H+1},\mathsf{a}^{\mathrm{rep}}_{1:H})\sim\mathsf{D}_{\pi^\star_{\mathrm{rep}}}$,

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

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

Remark I.1. The above result demonstrates the necessity of measuring the marginal error between $\hat{\pi} \circ \mathsf{W}_\sigma$ and $\pi^\star$ in Theorem 4: 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 $\hat{\pi} \circ \mathsf{W}_\sigma = \pi^\star_{\mathrm{rep}}$, and we know that $\Gamma_{\mathrm{joint},\varepsilon}(\pi^\star_{\mathrm{rep}}, \pi^\star) \geq \mathbb{P}_\mu[\exists 1 \leq h \leq H : \mathsf{d}(\mathsf{s}^\star_{h+1}, \mathsf{s}^{\mathrm{rep}}_{h+1}) > \varepsilon] \geq 1 - \exp(-\Omega(H))$. Thus, we cannot hope for smoothed policies to imitate expert demonstrations in joint state distributions without additional assumptions.

Remark I.2 (Importance of chunking)**.** Above we have shown that $\pi^\star_{\mathrm{rep}}$ 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 $\mathsf{s}[i]$,

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

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

I.3 $\pi^\star_{\mathrm{rep}}$ and $\pi^\star$ can have different marginals, implying necessity of $\gamma_{\mathrm{IPS},\mathcal{S}}$

Our construction lifts the construction in Appendix I.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

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

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

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

Let

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

Proposition I.2. In the above construction, we can take $\gamma_{\mathrm{IPS},\mathcal{S}}(u) \leq \nu \cdot u$ in Definition D.5, and p_r satisfies the conditions in Theorem 4 for $r = 2\sigma\sqrt{\log(1/p_r)}$. Moreover, for any $\varepsilon \leq \sigma$,

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

Remark I.3 (Sharpness of $\gamma_{\mathrm{IPS},\mathcal{S}}$)**.** 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 $\gamma_{\mathrm{IPS},\mathcal{S}}$ 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 I.2. To see $\gamma_{\mathrm{IPS},\mathcal{S}}(u) \leq \nu \cdot u$, we have $\|F_h(\mathsf{s}, \mathsf{a}) - F_h(\mathsf{s}', \mathsf{a})\| = \|(\mathsf{a}, \nu \cdot (\mathsf{s}[1] - \mathsf{a})) - (\mathsf{a}, \nu \cdot (\mathsf{s}'[1] - \mathsf{a}))\| = \nu|\mathsf{s}[1] - \mathsf{s}'[1]| \leq \nu \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}, \mathsf{s}')$. 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 (I.2) to find that, for any $b \in \mathbb{R}$,

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

This follows form the observation that $\mathsf{W}^\star_{\mathrm{rep},h}$ 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 $\mathsf{W}^\star_{\mathrm{rep},h}(\mathsf{s})$ is uninformative. For $\varepsilon \le \sigma$, we find that

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

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

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

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

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

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 $\mathsf{D}_{\pi^\star}, \mathsf{D}_{\pi^\star_{\mathrm{rep}}}$,

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

Thus,

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

□

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

Here, we show how $\pi^\star_{\mathrm{rep}}$ and $\pi^\star_{\mathrm{dec}}$ have different marginals even if the dynamics are memoryless. By considering $\hat\pi = \pi^\star_{\mathrm{dec}}$ in Theorem 4, the discussion below demonstrates why one needs to consider $\hat\pi_\sigma = \hat\pi \circ \mathsf{W}_\sigma$ 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

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

Let us consider an asymmetric initial state distribution

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

Note then that

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

We consider a smoothing kernel,

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

The salient part of our construction of W_σ is that

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

Denote the marginals of $\pi^\star_{\mathrm{rep}}$ 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

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

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

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

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

Claim I.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

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

which implies that the deconvolution policy $\pi^\star_{\mathrm{dec}}$ does approximate $\pi^\star_{\mathrm{rep}}$. From (I.6), it also follows that $\pi^\star_{\mathrm{rep}}$ 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 4, 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 $\hat\pi = \pi^\star_{\mathrm{dec}}$, $\hat\pi \circ \mathsf{W}_\sigma = \pi^\star_{\mathrm{rep}}$).

Proof of Claim I.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 (I.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 III

The Control Setting

J End-to-end Guarantees and the Proof of Theorem 3

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 the proof of Theorem 2 which provides the basis for the final proof of Theorem 3 in the body by uniting the analysis in the composite MDP from Appendix H, the control theory from Appendix K, and the sampling guarantees from Appendix L. We now summarize the organisation of the appendix:

- In Appendix J.1, we recall the association between the control setting and the composite MDP presented in Appendix D, as well as rigorously instantiating the direct decomposition and the expert policy.
- In Appendix J.2, we establish the correspondence between the imitation losses studied for the composite MDP in Appendix D with the disiderata in Section 2.
- In Appendix J.3, we provide the proof of Proposition D.2 and Theorems 2 and 3. We prove Theorem 2 as a consequence of a more granular guarantee, Theorem 10, which exposes the various tradeoffs in problem parameters.
- In Appendix J.4, we demonstrate that if the demonstrator policy is assumed to be TVC, then we can recover stronger guarantees than those provided in Theorem 3 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 J.5, 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 J.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 J.7, we collect a number of useful lemmata that we use throughout the appendix.

J.1 Preliminaries

Here, we state various preliminaries to the end-to-end theorems. Reall that $c_1, \dots, c_5$ are constants which are polynomial in the parameters in Assumption 3.1, and are spelled out explicitly in Appendix K. For simplicity, to avoid complications with the boundary effects at $h = 1$, we re-define $h = 1$-observation chunks o_1 as elements $\mathcal{O} = \mathscr{P}_{\tau_{\mathrm{obs}}-1}$ by prepending the necessary zeros – i.e. $\mathsf{o}_1 = (0, 0, \dots, 0, \mathbf{x}_1)$– and similarly modifying $\mathsf{s}_1 \in \mathcal{S} = \mathscr{P}_{\tau_{\mathrm{chunk}}}$ by prepending zeros. We first recall the definitions of the composite-states and -actions from Appendix D. 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 the distances

$$\mathsf{d}_{\mathrm{traj}}(\boldsymbol{\rho}, \boldsymbol{\rho}') := \max_{1 \le k \le \tau+1} \|\mathbf{x}_k - \mathbf{x}'_k\| \vee \max_{1 \le k \le \tau} \|\mathbf{u}_k - \mathbf{u}'_k\|$$

defined for trajectories of arbitrary length. Given $\mathsf{s}_h, \mathsf{s}'_h$ with observation-(sub)chunks $\mathsf{o}_h, \mathsf{o}'_h$, we define

$$\begin{aligned}
\mathsf{d}_{\mathcal{S}}(\mathsf{s}_h, \mathsf{s}'_h) &= \mathsf{d}_{\mathrm{traj}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_{t \in [t_{h-1}:t_h]} ||\mathbf{x}_t - \mathbf{x}'_t|| \vee \max_{t \in [t_{h-1}:t_h-1]} = ||\mathbf{u}_t - \mathbf{u}'_t|| \,, \\
\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}_h, \mathsf{s}'_h) &= \mathsf{d}_{\mathrm{traj}}(\mathsf{o}_h, \mathsf{o}'_h) = \max_{t \in [t_h - \tau_{\mathrm{obs}}:t_h]} ||\mathbf{x}_t - \mathbf{x}'_t|| \vee \max_{t \in [t_h - \tau_{\mathrm{obs}}:t_h-1]} ||\mathbf{u}_t - \mathbf{u}'_t|| \,, \\
\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) &= \left|\left|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\right|\right| .
\end{aligned}$$

Finally, for $\mathsf{a} = (\bar{\mathbf{u}}_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{x}}_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{K}}_{1:\tau_{\mathrm{chunk}}})$ and $\mathsf{a}' = (\bar{\mathbf{u}}'_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{x}}'_{1:\tau_{\mathrm{chunk}}}, \bar{\mathbf{K}}'_{1:\tau_{\mathrm{chunk}}})$, recall from (3.1) and (D.2) that

$$\begin{aligned}\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') &= \max_{1 \le k \le \tau_{\mathrm{chunk}}} \|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\| \\ \mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') &:= c_1 \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \cdot \mathbf{I}_\infty\{\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') > c_2\}\end{aligned}$$

We note the following fact.

Fact J.1. Suppose that $\varepsilon \le c_2$. Then $\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') \le \varepsilon$ whenever $\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le \varepsilon / c_1$.

J.1.1 Direct Decomposition and Smoothing Kernel.

This section will invoke the generalizations Theorem 4 which requires TVC only subspace of the state space. This invokes the direct decomposition explained in Appendix H.

Definition J.1 (Direct Decomposition and Smoothing Kernel)**.** We consider the decomposition of $\mathcal{S} = \mathcal{O} \oplus \mathcal{S}_{/\mathcal{O}}$, where $\mathcal{O} = \mathscr{P}_{\tau_{\mathrm{obs}}-1}$ are the coordinates of s_h corresponding to the observation chunk o_h, and $\mathcal{S}_{/\mathcal{O}}$ are all remaining coordinates We let $\phi_{\mathrm{o}} : \mathcal{S} \to \mathcal{O}$ denote the projection onto the coordinates in $\mathcal{O}$. We instantiate the smoothing kernel W_σ as follows: For $\mathsf{s} = \mathsf{s}_h \in \mathcal{S} = \mathscr{P}_{\tau_{\mathrm{chunk}}}$, we let

$$\mathsf{W}_\sigma(\mathsf{s}) = \mathcal{N}\left(\mathsf{s}_h, \begin{bmatrix} \sigma^2 \mathbf{I}_{\mathcal{O}} & 0 \\ 0 & 0 \end{bmatrix}\right), \tag{J.1}$$

where $\mathbf{I}_{\mathcal{O}}$ denotes the identity supported on the coordinates in $\mathcal{O}$ as described above.

We note that the above direct decomposition satisfies the requiste compatibility assumptions explained in Appendix H. Note also that d_{IPS} and W_σ are compatible with the above direct decomposition.

J.1.2 Chunking Policies.

We continue by centralizing a definition of chunking policies.

Definition J.2 (Policy and Initial-State Distributions)**.** Given an *chunking policy* $\pi = (\pi_h)_{h=1}^H$ with $\pi_h : \mathcal{O} = \mathscr{P}_{\tau_{\mathrm{obs}}-1} \to \Delta(\mathcal{A})$, we let $\mathcal{D}_\pi$ denote the distribution over $\boldsymbol{\rho}_T$ and $\mathsf{a}_{1:H}$ induced by selecting $\mathsf{a}_h \sim \pi_h(\mathsf{o}_h)$, and rolling out the dynamics as described in Section 2. We extend chunking policies to maps $\pi_h : \mathcal{S} = \mathscr{P}_{\tau_{\mathrm{chunk}}} \to \Delta(\mathcal{A})$ by expressing $\pi_h = \pi_h \circ \phi_{\mathrm{o}}$ (i.e., projection s_h onto its o_h-components). Further, we let $\mathsf{P}_{\mathrm{init}}$ denote the distribution of $\mathbf{x}_1$ under $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$.

Remark J.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 composite MDP. For composite MDPs instantiated as in Appendix D.2, the two exactly coincide.

Construction of $\pi^\star$ for composite MDP. We now recall the policies $\pi^\star$ and $\pi^\star_{\mathrm{dec},\sigma}$, defined in Definitions 3.4 and 3.6, respectively.

Definition J.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 $\mathsf{o}_{1:H}$ be the corresponding observation-chunks from $\boldsymbol{\rho}_T$. Let

- $\pi^\star_h(\cdot) : \mathcal{O} = \mathscr{P}_{\tau_{\mathrm{obs}}-1} \to \Delta(\mathcal{A})$ denote a regular conditional probability corresponding to the distribution over a_h given o_h in the above construction. In other words, the distribution of $\mathsf{a}_h \mid \mathsf{o}_h$ under $\mathcal{D}_{\sigma=0,h}$.
- Let $\pi^\star_{\mathrm{dec},\sigma,h}(\cdot) : \mathcal{O} = \mathscr{P}_{\tau_{\mathrm{obs}}-1} \to \Delta(\mathcal{A})$ denote a regular conditional probability corresponding to the distribution over a_h given an augmented $\tilde{\mathsf{o}}_h \sim \mathcal{N}(\mathsf{o}_h, \sigma^2 \mathbf{I})$. In other words, the distribution of $\mathsf{a}_h \mid \tilde{\mathsf{o}}_h$ under $\mathcal{D}_{\sigma,h}$.

When instantiating the composite MDP, $\pi^\star$ corresponds to its namesake, and $\pi^\star_{\mathrm{dec},\sigma,h}$ to $\pi^\star_{\mathrm{dec},h}$. Moreover, $\pi^\star$ as constructed above, $\mathsf{P}^\star_h$ denotes the distribution over s_h under $\mathcal{D}_{\pi^\star}$. By Lemma J.6, this is in fact equal to the distribution over s_h under $\mathcal{D}_{\mathrm{exp}}$. Notice further, therefore, that $\phi_{\mathrm{o}} \circ \mathsf{P}^\star_h$ is precisely the distribution of o_h under $\mathcal{D}_{\mathrm{exp}}$.

Remark J.2. We remark that by Theorem 7, $\pi^\star_h$ is unique up to a measure zero set of o_h as distributed as above, and $\pi^\star_{\mathrm{dec},h}$ is unique almost surely for $\tilde{\mathsf{o}}_h$ 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.

J.1.3 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 J.4 (Joint and Final 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. We also define

$$\mathcal{L}_{\mathrm{fin},\varepsilon}(\pi) := \inf_{\mu} \mathbb{P}_\mu \left[\|\mathbf{x}^{\mathrm{exp}}_{T+1} - \mathbf{x}^\pi_{T+1}\| > \varepsilon \right],$$

the loss restricted to the final states under each distribution.

Controlling $\mathcal{L}_{\mathrm{joint},\varepsilon}(\pi)$ requires various additional stronger assumptions (*which we do not require in Theorem 3*), one of which is that the demonstrator has bounded memory:

Definition J.5. 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 J.5 can be relaxed to a mixing time assumption (see Remark J.5). Moreover, we stress that we *do not* need the condition in Definition J.5 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 3.

J.2 Translating Control Imitation Losses to Composite-MDP Imitation Gaps

Lemma J.1. Recall the imitation losses Definitions 2.2 and J.4, and the compsite-MDP imitation gaps Definition D.1. Further consider, the substitutions defined in Appendix D.2, with $\pi^\star$ instantiated as in Definition J.3. Given policies $\pi = (\pi_h)$ with $\pi_h : \mathcal{O} = \mathscr{P}_{\tau_{\mathrm{obs}}-1} \to \mathcal{A}$, we can extend $\pi_h : \mathcal{S} = \mathscr{P}_{\tau_{\mathrm{chunk}}} \to \mathcal{A}$ by the natural embedding of $\mathscr{P}_{\tau_{\mathrm{obs}}-1}$ into $\mathscr{P}_{\tau_{\mathrm{chunk}}}$. 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 Appendix D.2, but set $\mathsf{d}_{\mathcal{S}}$ to equal $\mathsf{d}_{\mathrm{IPS}}$, which only measures distance in the final coordinate of each trajectory chunk s_h,

$$\mathcal{L}_{\mathrm{fin},\varepsilon}(\pi) \le \Gamma_{\mathrm{marg},\varepsilon}(\pi \parallel \pi^\star), \quad \mathsf{d}_{\mathcal{S}}(\cdot,\cdot) \leftarrow \mathsf{d}_{\mathrm{IPS}}(\cdot,\cdot) \tag{J.2}$$

Finally, if $\mathcal{D}_{\mathrm{exp}}$ has $\tau \le \tau_{\mathrm{obs}}$-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 $\mathcal{D}_{\mathrm{exp}}$ 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

$$\begin{aligned}
\mathcal{L}_{\mathrm{marg},\varepsilon}(\pi) &:= \max_{t\in[T]} \inf_{\mu} \left\{ \mathbb{P}_\mu \left[\|\mathbf{x}^{\mathrm{exp}}_{t+1} - \mathbf{x}^\pi_{t+1}\| > \varepsilon \right], \mathbb{P}_\mu \left[\|\mathbf{u}^{\mathrm{exp}}_t - \mathbf{u}^\pi_t\| > \varepsilon \right] \right\} \\
&:= \max_{t\in[T]} \inf_{\mu} \left\{ \mathbb{P}_\mu \left[\|\mathbf{x}^{\mathrm{exp}}_{t+1} - \mathbf{x}^\pi_{t+1}\| \vee \|\mathbf{u}^{\mathrm{exp}}_t - \mathbf{u}^\pi_t\| > \varepsilon \right] \right\} \\
&\overset{(a)}{\le} \max_{h\in[H]} \inf_{\mu} \left\{ \mathbb{P}_\mu \left[\max_{0\le i\le \tau_{\mathrm{chunk}}} \|\mathbf{x}^{\mathrm{exp}}_{t_{h+1}-i} - \mathbf{x}^\pi_{t_{h+1}-i}\| \vee \max_{1\le i\le \tau_{\mathrm{chunk}}} \|\mathbf{u}^{\mathrm{exp}}_{t_{h+1}-i} - \mathbf{u}^\pi_{t_{h+1}-i}\| \right] \right\} \\
&\le \max_{h\in[H]} \inf_{\mu} \left\{ \mathbb{P}_\mu \left[\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\mathrm{exp}}_{h+1}, \mathsf{s}^\pi_{h+1}) \right] \right\},
\end{aligned}$$

where step (a) uses that for any $t \in [H]$, we can find some h such that $t+1 \in t_{h+1} - \{0, 1, \ldots, \tau_{\text{chunk}}\}$ and $t \in t_{h+1} - \{1, 2, \ldots, \tau_{\text{chunk}}\}$.[10]

From Lemma J.6, $\mathsf{s}_h^{\text{exp}}$ has the same marginal distribution as $\mathsf{s}_h^{\pi^\star}$, the distribution induced by $\pi^\star$ in Definition J.3.[11] Still, letting $\inf_{\mu'}$ denote infimum over couplings between $\mathcal{D}_\pi$ and $\mathcal{D}_{\text{exp}}$, equality of thesemarginals suffices to ensure

$$\mathcal{L}_{\text{marg},\varepsilon}(\pi) \le \max_{h\in[H]} \inf_{\mu} \left\{ \mathbb{P}_{\mu'} \left[\mathsf{d}_{\mathcal{S}}(\mathsf{s}_{h+1}^{\pi^\star}, \mathsf{s}_{h+1}^{\pi}) \right] \right\},$$

As s_h corresponds to a composite state s_h in the composite MDP, the above is at most $\Gamma_{\text{marg},\varepsilon}(\pi \parallel \pi^\star)$ as in definition Definition D.1. For the final-state imitation loss,

$$\begin{aligned}\mathcal{L}_{\text{fin},\varepsilon}(\pi) &:= \inf_{\mu} \mathbb{P}_{\mu} \left[\|\mathbf{x}_{T+1}^{\text{exp}} - \mathbf{x}_{T+1}^{\pi}\| > \varepsilon \right] \\ &\le \max_{h\in[H]} \inf_{\mu} \left\{ \mathbb{P}_{\mu} \left[\mathsf{d}_{\text{IPS}}(\mathsf{s}_h^{\text{exp}}, \mathsf{s}_h^{\pi}) \right] \right\},\end{aligned}$$

where again d_{IPS} only measures error in the final state of s_h. The corresponding bound in (J.2) follows similarly.

Finally, we have

$$\mathcal{L}_{\text{joint},\varepsilon}(\pi) := \inf_{\mu} \mathbb{P}_{\mu} \left[\max_{t\in[T]} \max \left\{ \|\mathbf{x}_{t+1}^{\text{exp}} - \mathbf{x}_{t+1}^{\pi}\|, \|\mathbf{u}_t^{\text{exp}} - \mathbf{u}_t^{\pi}\| \right\} > \varepsilon \right],$$

When $\mathcal{D}_{\text{exp}}$ has $\tau \le \tau_{\text{obs}}$-bounded memory, then, the expert and $\pi^\star$-induced trajectories are identically distributed. Therefore, directly from this observation and Definition D.1,

$$\mathcal{L}_{\text{joint},\varepsilon}(\pi) = \inf_{\mu} \mathbb{P}_{\mu} \left[\max_{t\in[T]} \max \left\{ \|\mathbf{x}_{t+1}^{\pi^\star} - \mathbf{x}_{t+1}^{\pi}\|, \|\mathbf{u}_t^{\pi^\star} - \mathbf{u}_t^{\pi}\| \right\} > \varepsilon \right] \le \Gamma_{\text{joint},\varepsilon}(\pi \parallel \pi^\star).$$

□

J.3 Proofs of main results

J.3.1 Proof of Theorem 1

The result is a direct consequence of the following points. First, with our instantition of the composite MDP, we can bound $\mathcal{L}_{\text{marg},\varepsilon}(\hat{\pi}) \le \Gamma_{\text{marg},\varepsilon}(\hat{\pi} \parallel \pi^\star) \le \Gamma_{\text{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star)$ due to Lemma J.1; and moreover, we have $\mathcal{L}_{\text{joint},\varepsilon}(\hat{\pi}) \le \Gamma_{\text{joint},\varepsilon}(\hat{\pi} \parallel \pi^\star)$ when $\mathcal{D}_{\text{exp}}$ has $\tau \le \tau_{\text{obs}}$-bounded memory. The bound now follows from Proposition D.2, and the fact that Proposition D.1 verifies the input-stability property, and Fact J.1. □

J.3.2 Proof of Theorem 2 and a more precise statement.

In this section, we derive Theorem 2 from a more precise guarantee that exposes the various algorithmic knobs in a more explicit manner.

Theorem 10. Let Assumption 3.1 hold, and let $c_1, \ldots, c_5 > 0$ be as in Definition 3.2. Suppose that the $\varepsilon, \sigma, \tau_{\text{chunk}} > 0$ satisfy $\varepsilon < c_2$, and $\tau_{\text{chunk}} \ge c_3$, and $5d_x + \log\left(\frac{4\sigma}{c_1\varepsilon}\right) \le c_4^2/(16\sigma^2)$. Then the marginal imitation loss (Definition 2.2) and final-state imitation loss (Definition J.4) of the *smoothed* $\hat{\pi}_\sigma$ are bounded by

$$\begin{aligned}\max\left\{\mathcal{L}_{\text{marg},\varepsilon_1}(\hat{\pi}_\sigma), \mathcal{L}_{\text{fin},\varepsilon_2}(\hat{\pi}_\sigma)\right\} &\le H\sqrt{2\tau_{\text{obs}} - 1}\left(\frac{2\varepsilon}{\sigma} + \iota_\sigma(\varepsilon) e^{-L_\beta(\tau_{\text{chunk}} - \tau_{\text{obs}})}\right) \\ &+ \sum_{h=1}^{H} \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\text{exp},\sigma,h}} \Delta_{(\varepsilon/c_1)}\left(\pi^\star_{\text{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)\right).\end{aligned} \tag{J.3}$$

and where $\iota_\sigma(\varepsilon) = 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{c_1\varepsilon}\right)}$ is logarithmic in $1/\varepsilon$, and where

$$\varepsilon_1 = \varepsilon + \sigma\iota_\sigma(\varepsilon), \quad \varepsilon_2 = \varepsilon + \sigma e^{-L_\beta(\tau_{\text{chunk}} - \tau_{\text{obs}})}\iota(\varepsilon).$$

[10] Recall $t_h := (h-1)\tau_{\text{chunk}} + 1$.

[11] Note the subtlety that the joint distribution of these may defer because $\pi^\star$ has limited memory.

Remark J.3 (Parameter Dependencies)**.** Each term in (J.3) can be made small by decreasing the amount of noise σ in the smoothing, increasing the number of trajectories, and increasing the chunk length τ_{chunk}; indeed, these are the levers by means of which we derive Theorem 2 just below. Increasing τ_{chunk} comes at the (implicit) expense of increasing the length of composite actions, thereby inducing a more challenging conditional generative modeling problem. Decreasing σ increases sensitivity to the tolerance ε, and, as discuseed in Appendix L, may make the underlying generative modeling problem more challenging. Note that the contribution of the additive σ-term in ε_2, used for the final-state loss $\mathcal{L}_{\text{fin},\varepsilon}$, is exponentially-in-$\tau_{\text{chunk}}$ smaller than that in ε_1. Interestingly, our theory suggest no benefit to increasing τ_{obs} (corroborated empirically in [19]).

We now turn the proof of proof of Theorem 2.

Deduction. Theorem 2 from Theorem 10. Fix a desired ε_0 for which

$$\varepsilon_0 < \min\{1/2c_1, 1/\sqrt{c_1/c_2}, 3c_4c_5\}.$$

Then, taking $\varepsilon = c_1\varepsilon_0^2$, we have $\varepsilon < c_2$ and $\varepsilon \le \varepsilon_0/2$. Select $\sigma = \frac{1}{2}\varepsilon_0/\iota_\star(\varepsilon_0)$. For such an σ, we have that as $\iota_\star(\varepsilon_0) \ge 6c_5$, $\sigma \le \frac{1}{12c_5}\varepsilon_0$, and thus

$$\iota_\sigma(\varepsilon) = 6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{c_1\varepsilon}\right)} = 6c_5\sqrt{5d_x + 2\log\left(\frac{8\sigma}{c_1^2\varepsilon_0^2}\right)} \le 6c_5\sqrt{5d_x + 2\log\left(\frac{1}{c_5c_1^2\varepsilon_0}\right)} =: \iota_\star(\varepsilon_0)$$

and therefore, with our vari $\varepsilon_1 := \varepsilon + \sigma\iota_\sigma(\varepsilon) \le \varepsilon_0/2 + \varepsilon_0/2 \le \varepsilon_0$. Next, we verify the condition

$$c_4^2/(16\sigma^2) \ge 5d_x + \log\left(\frac{4c_1\sigma}{\varepsilon}\right) = (\iota_\sigma(\varepsilon)^2)/(6c_5)^2$$

Rearranging, we need $\frac{36c_5^2c_4^2}{16} \ge (\sigma\iota_\sigma(\varepsilon)^2$, and as $(\sigma\iota_\sigma(\varepsilon)^2 \le \varepsilon_0/2$, it then suffices that $\varepsilon_0 \le 3c_4c_5$, which holds. Therefore, Theorem 2 implies

$$\mathcal{L}_{\text{marg},\varepsilon_0}(\hat{\pi}_\sigma) \le H\iota_\star(\varepsilon_0)\sqrt{2\tau_{\text{obs}} - 1}\left(2c_1\varepsilon_0 + e^{-L_\beta(\tau_{\text{chunk}} - \tau_{\text{obs}})}\right) + \sum_{h=1}^{H}\mathbb{E}_{\tilde{\mathsf{o}}_h\sim\mathcal{D}_{\text{exp},\sigma,h}}\Delta_{(\varepsilon/c_1)}\left(\pi^\star_{\text{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)\right).$$

The first result follows by relabeling $\varepsilon \leftarrow \varepsilon_0$ and taking $\tau_{\text{chunk}} - \tau_{\text{obs}} \ge \frac{1}{L_\beta}\log(c_1/\varepsilon)$. The second result is a consequence of Markov's inequality, and the well behaved-ness of conditional distributions established throughout Appendix F. □

J.3.3 Proof of Theorem 10

Proof of Theorem 10. Lets begin by bounding $\mathcal{L}_{\text{marg},\varepsilon}(\pi)$. Recall the definitions of $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\text{TVC}}, \mathsf{d}_{\text{IPS}}$ in Appendix D, 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 Appendix D.2. We can view $\pi^\star$ and π (which depend only on observation chunks o_h) as policies in the composite MDP which are compatible with the decomposition Definition H.1. We make the following points:

- In light of Lemma J.1,

$$\mathcal{L}_{\text{marg},\varepsilon_1}(\pi \parallel \pi^\star) \le \Gamma_{\text{marg},\varepsilon_1}(\pi \parallel \pi^\star).$$

- By Lemma J.8, a consequence of Pinsker's inequality, it holds that the Gaussian kernel W_σ used in HINT is γ_σ-TVC (w.r.t. d_{TVC}) with

$$\gamma_\sigma(u) = \frac{u\sqrt{2\tau_{\text{obs}} - 1}}{2\sigma} \tag{J.4}$$

- Note that $\mathsf{d}_{\text{IPS}}(\mathsf{s}_h, \mathsf{s}'_h) = \|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\|$ 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 J.7 with $d = d_x$, that for $r = 2\sigma \cdot \sqrt{5d_x + 2\log\left(\frac{1}{p}\right)}$

$$p_r = \mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\text{IPS}}(\mathsf{s}, \mathsf{s}') > r] \le p. \tag{J.5}$$

- As (a) $\mathsf{s}_h^\star$ corresponds to s_h from $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$, (b) as $\hat{\pi}, \pi^\star_{\mathrm{dec}}$ are functions of o_h, and (c) by recalling the definition of $\mathsf{d}_{\mathrm{os},\varepsilon}$ in Definition D.1, $\varepsilon \le c_2$ ensures

$$
\begin{aligned}
&\mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star} \mathbb{E}_{\tilde{\mathsf{s}}_h^\star \sim \mathsf{W}_\sigma(\mathsf{s}_h^\star)} \mathsf{d}_{\mathrm{os},\varepsilon}(\pi_h(\tilde{\mathsf{s}}_h^\star) \parallel \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}_h^\star)) \\
&= \mathbb{E}_{\mathsf{o}_h \sim \mathcal{D}_{\mathrm{exp},h}} \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{N}(\mathsf{o}_h, \sigma^2 \mathbf{I})} \inf_{\mu \in \mathscr{C}(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h))} \mathbb{P}_{(\mathsf{a},\mathsf{a}') \sim \mu}[\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') \ge \varepsilon] \\
&\le \mathbb{E}_{\mathsf{o}_h \sim \mathcal{D}_{\mathrm{exp},h}} \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{N}(\mathsf{o}_h, \sigma^2 \mathbf{I})} \inf_{\mu \in \mathscr{C}(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h))} \mathbb{P}_{(\mathsf{a},\mathsf{a}') \sim \mu}[\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \ge \varepsilon / c_1] \\
&\hspace{30em} \text{(Fact J.1)} \\
&= \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}} \Delta_{(\varepsilon/c_1)}(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)).
\end{aligned}
$$

- Finally, Proposition K.1 (formalizing Proposition D.1) ensures that under our assumption $\tau_{\mathrm{chunk}} \ge c_3/$, and let $r_{\mathrm{IPS}} = c_4$, $\gamma_{\mathrm{IPS,TVC}}(u) = c_5 u \exp(-L_{\beta}(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}}))$, $\gamma_{\mathrm{IPS},\mathcal{S}}(u) = c_5 u$ for c_3, c_4, c_5 given in Appendix K. Then, for $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ as above, we have that $\pi^\star$ is $(\gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}, \mathsf{d}_{\mathrm{IPS}}, r_{\mathrm{IPS}})$-IPS.

Consequently, for $r = 2\sigma \cdot \sqrt{5d_x + 2\log\left(\frac{4\sigma}{c_1 \varepsilon}\right)} \in (0, \frac{1}{2} r_{\mathrm{IPS}})$, Theorem 9 (which, we recall, generalizes Theorem 4 to account for the direct decomposition structure) implies

$$
\begin{aligned}
\mathcal{L}_{\mathrm{marg},\varepsilon + 2rc_5}(\hat{\pi}_\sigma) &= \mathcal{L}_{\mathrm{marg},\varepsilon+2rc_5}(\hat{\pi}_\sigma \parallel \pi^\star) \le \Gamma_{\mathrm{marg},\varepsilon+2rc_5}(\hat{\pi}_\sigma \parallel \pi^\star) \\
&\le H\left(\frac{\varepsilon}{2\sigma} + \frac{3}{2\sigma}\sqrt{2\tau_{\mathrm{obs}} - 1}\left(\max\left\{\varepsilon,\, 2rc_5 e^{-L_{\beta}(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})}\right\}\right)\right) \\
&\quad + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}_h^\star \sim \mathsf{P}_h^\star} \mathbb{E}_{\tilde{\mathsf{s}}_h^\star \sim \mathsf{W}_\sigma(\mathsf{s}_h^\star)} \mathsf{d}_{\mathrm{os},\varepsilon}(\pi_h(\tilde{\mathsf{s}}_h^\star) \parallel \pi^\star_{\mathrm{dec}}(\tilde{\mathsf{s}}_h^\star)) \\
&\le H\sqrt{2\tau_{\mathrm{obs}} - 1}\left(\frac{2\varepsilon}{\sigma} + 6\sigma c_5 \sqrt{5d_x + 2\log\left(\frac{4\sigma}{c_1\varepsilon}\right)} e^{-L_{\beta}(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})}\right) \\
&\quad + \sum_{h=1}^{H} \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}} \Delta_{(\varepsilon/c_1)}(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)). \\
&\le H\sqrt{2\tau_{\mathrm{obs}} - 1}\left(\frac{2\varepsilon}{\sigma} + \sigma\iota(\varepsilon)\right) + \sum_{h=1}^{H} \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}} \Delta_{(\varepsilon/c_1)}(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)).
\end{aligned}
$$

Substituting in $\varepsilon_1 = \varepsilon + 2rc_5 = c_1\varepsilon + 4c_5\sigma \cdot \sqrt{5d + 2\log\left(\frac{4\sigma}{c_1\varepsilon}\right)} \le \varepsilon + \sigma\iota(\varepsilon)$, 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 Appendix D.2) 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 Proposition K.1, it follows we can take $\gamma_{\mathrm{IPS},\mathcal{S}}(u) = c_5 u e^{-L_{\beta}\tau_{\mathrm{chunk}}}$. Thus, we can replace ε_1 above with $\varepsilon_2 := \varepsilon + 4c_5 e^{-L_{\beta}\tau_{\mathrm{chunk}}}\sigma \cdot (5d_x + 2\log\left(\frac{1}{\varepsilon}\right))^{1/2}$. This concludes the proof that

$$
\mathcal{L}_{\mathrm{marg},\varepsilon_2}(\hat{\pi}_\sigma) \le H\sqrt{2\tau_{\mathrm{obs}} - 1}\left(6c_5\sqrt{5d_x + 2\log\left(\frac{4\sigma}{c_1\varepsilon}\right)} e^{-L_{\beta}(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})} + \frac{2\varepsilon}{\sigma}\right),
$$

which can be simplified as needed. □

J.3.4 Proof of Theorem 3

Adopt the shorthand $\Delta_h = \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}} \Delta_{(\varepsilon/c_1)}\left(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h)\right)$. From Theorem 2, it suffices to show that with probability at least $1 - \delta$, it holds that $\Delta_h \le \varepsilon^2$ for all $h \in [H]$. For $d_{\mathcal{A}} = \tau_{\mathrm{chunk}}(d_x + d_u + d_x d_u)$, we have that $\mathsf{a} \in \mathbb{R}^{d_{\mathcal{A}}}$. Note that by Assumption 3.1 it holds $\mathcal{D}_{\mathrm{exp}}$-almost surely that we can crudely bound $||\mathsf{a}_h|| \le \sqrt{d_{\mathcal{A}}}\max\{R_{\mathbf{K}}, R_{\mathrm{dyn}}\}$ and thus the condition on

q in Theorem 13 holds for

$$R = \sqrt{d_{\mathcal{A}}} \max\{R_{\mathbf{K}}, R_{\mathrm{dyn}}\}.$$

By Assumption C.1, the conditions on the score class $\mathbf{s}_\theta$ hold for us to apply Theorem 13. Note that by assumption,

$$N_{\mathtt{exp}} \geq c \left(\frac{C_\Theta d R (R \vee \sqrt{d_{\mathcal{A}}}) \log(dn)}{\varepsilon^8} \right)^{4\nu} \vee \left(\frac{d_{\mathcal{A}}^6 (R^4 \vee d_{\mathcal{A}}^2 \log^3\left(\frac{HndR\sigma}{\delta\varepsilon}\right))}{\varepsilon^{48}} d_{\mathcal{A}}^2 \right)^{4\nu}, \tag{J.6}$$

where we note that the right hand side is poly $(C_\Theta, 1/\varepsilon, R_{\mathbf{K}}, d_{\mathcal{A}}, \log(H/\delta))^\nu$, and J and α are set as in (L.1). Taking a union bound over $h \in [H]$ and applying Theorem 13 tells us that with probability at least $1-\delta$, for all $h \in [H]$, it holds that

$$\mathbb{E}_{\tilde{\mathsf{o}}_h \sim q_{\tilde{\mathsf{o}}_h}} \left[\inf_{\mu \in \mathscr{C}(\mathtt{DDPM}(\mathbf{s}_{\widehat{\theta}}, \tilde{\mathsf{o}}_h), q(\cdot|\tilde{\mathsf{o}}_h))} \mathbb{P}_{(\widehat{\mathsf{a}}, \mathsf{a}^*) \sim \mu} \left(||\widehat{\mathsf{a}} - \mathsf{a}^*|| \geq \varepsilon \right) \right] \leq \varepsilon.$$

Thus, it holds that with probability at least $1-\delta$,

$$\sum_{h=1}^H \Delta_h \leq H\varepsilon.$$

Plugging this in to Theorem 2 concludes the proof of the first statement. The proof of the second statement is analogous.

□

J.4 Imitation of the joint trajectory under total variation continuity of demonstrator policy

Here, we show that if the demonstrator policy has (a) bounded memory and (b) 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 J.4.

Theorem 11. Consider the setting Theorem 10, and define as shorthand

$$\Delta_{h,\varepsilon} := \mathbb{E}_{\tilde{\mathsf{o}}_h \sim \mathcal{D}_{\mathrm{exp},\sigma,h}} \Delta_{(\varepsilon/c_1)} \left(\pi^\star_{\mathrm{dec},\sigma,h}(\tilde{\mathsf{o}}_h), \hat{\pi}_h(\tilde{\mathsf{o}}_h) \right).$$

Suppose that, in addition, there is a strictly increasing function $\gamma(\cdot)$ such that for all $\mathsf{o}_h, \mathsf{o}'_h \in \mathcal{O}$,

$$\mathsf{TV}(\pi^\star(\mathsf{o}_h), \pi^\star(\mathsf{o}'_h)) \leq \gamma(\|\mathsf{o}_h - \mathsf{o}'_h\|),$$

where $\pi^\star$ is defined is the conditional in Definition J.3. Further, suppose that $\mathcal{D}_{\mathrm{exp}}$ has $\tau \leq \tau_{\mathrm{obs}}$ bounded memory (Definition J.5). Then, with $\varepsilon_1 := \varepsilon + \sigma\iota(\varepsilon)$ as in Theorem 2,

$$\begin{aligned}\mathcal{L}_{\mathrm{joint},\varepsilon_1}(\hat{\pi}_\sigma) &\leq H \cdot \textsc{ErrTvc}(\sigma, \gamma) \\ &+ H\sqrt{2\tau_{\mathrm{obs}} - 1} \left(\frac{2\varepsilon}{\sigma} + 6c_5 \sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)} e^{-L_\beta(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})} \right) + \sum_{h=1}^H \Delta_{h,\varepsilon}.\end{aligned}$$

where we define $d_0 = \tau_{\mathrm{obs}} d_x + (\tau_{\mathrm{obs}} - 1) d_u$ and $u_0 = \gamma(8\sigma\sqrt{d_0 \log(9)})$, and

$$\textsc{ErrTvc}(\sigma, \gamma) = \begin{cases} 2c\sigma\sqrt{d_0} & \text{linear } \gamma(u) = c \cdot u, c > 0 \\ u_0 + \int_{u_0}^\infty e^{-\frac{\gamma^{-1}(u)^2}{64\sigma^2}} \mathrm{d}u & \text{general } \gamma(\cdot) \end{cases}. \tag{J.7}$$

In particular, under Assumption C.1, if

$$N_{\mathtt{exp}} \geq c \left(\frac{C_\Theta d_{\mathcal{A}} R (R \vee \sqrt{d_{\mathcal{A}}}) \log(dn)}{(\varepsilon/\sigma)^4} \right)^{4\nu} \vee \left(\frac{d_{\mathcal{A}}^6 (R^4 \vee d_{\mathcal{A}}^2 \log^3\left(\frac{HndR\sigma}{\delta\varepsilon}\right))}{(\varepsilon/\sigma)^{24}} d_{\mathcal{A}}^2 \right)^{4\nu},$$

then with probability at least $1-\delta$, it holds that

$$\mathcal{L}_{\mathrm{joint},\varepsilon_1}(\hat{\pi}_\sigma) \leq H \cdot \textsc{ErrTvc}(\sigma, \gamma) + H\sqrt{2\tau_{\mathrm{obs}} - 1} \left(\frac{3\varepsilon}{\sigma} + 6c_5 \sqrt{5d_x + 2\log\left(\frac{4\sigma}{\varepsilon}\right)} e^{-L_\beta(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})} \right).$$

Remark J.4. The second term in our bound on $\mathcal{L}_{\text{joint},\varepsilon}(\pi)$ is identical to the bound in Theorem 2. 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 J.4.1; it mirrors that of Theorem 2, but replaces Theorem 4 with the following imitation guarantee in the composite MDP abstraction of Appendix D, which bounds the joint imitation gap relative to $\pi^\star$ if $\pi^\star$ is TVC.

Proposition J.2. Consider the set-up of Appendix D, and suppose that the assumptions of Theorem 9, but that, in addition, the expert policy $\pi^\star$ is $\tilde{\gamma}(\cdot)$-TVC with respect to the pseudometric $\mathsf{d}_{\textsc{tvc}}$, where $\tilde{\gamma}:\mathbb{R}_{\ge 0}\to\mathbb{R}_{\ge 0}$ is strictly increasing. Then, for all parameters as in Theorem 4, and any $\tilde{r}>0$,

$$\Gamma_{\text{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) \le H\int_0^\infty \max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\textsc{tvc}}(\mathsf{s},\mathsf{s}')>\tilde{\gamma}^{-1}(u)/2]\mathrm{d}u$$
$$+ H\left(2p_r+3\gamma_\sigma(\max\{\varepsilon,\gamma_{\textsc{ips},\textsc{tvc}}(2r)\})\right)+\sum_{h=1}^H\mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star}\mathbb{E}_{\tilde{\mathsf{s}}_h^\star\sim\mathsf{W}_\sigma(\mathsf{s}_h^\star)}\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}_h^\star)\parallel\pi^\star_{\text{dec}}(\tilde{\mathsf{s}}_h^\star)),$$

where the term in color on the first line is the only term that differs from the bound in Theorem 4.

Moreover, in the special case where all of the distributions of $\mathsf{d}_{\textsc{tvc}}(\mathsf{s},\mathsf{s}')\mid\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})$ 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

$$\Gamma_{\text{joint},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) \le 2\tilde{c}H\mathbb{E}[Z]$$
$$+ H\left(2p_r+3\gamma_\sigma(\max\{\varepsilon,\gamma_{\textsc{ips},\textsc{tvc}}(2r)\})\right)+\sum_{h=1}^H\mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star}\mathbb{E}_{\tilde{\mathsf{s}}_h^\star\sim\mathsf{W}_\sigma(\mathsf{s}_h^\star)}\mathsf{d}_{\text{os},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}_h^\star)\parallel\pi^\star_{\text{dec}}(\tilde{\mathsf{s}}_h^\star)).$$

Proof Sketch. Proposition J.2 is derived below in Appendix J.4.2. It is corollary of Theorem 4, combined with adjoining the coupling constructed therein to a TV distance coupling between $\pi^\star_{\text{rep}}$ (whose joints we *can always* imitate) and $\pi^\star$. Coupling trajectories induced by $\pi^\star_{\text{rep}}$ and $\pi^\star$ relies on the TVC of $\pi^\star$, as well as concentration of W_σ. □

Using the above proposition, we can derive the following consequences for imitation of the joint distribution.

J.4.1 Proof of Theorem 11

The proof is nearly identical to that of Theorem 2, with the modifications that we replace our use of Theorem 4 with Proposition J.2 taking $\tilde{\gamma}\leftarrow\gamma$. By Lemma J.1 and the assumpton that $\mathcal{D}_{\text{exp}}$ has $\tau\le\tau_{\text{obs}}$-bounded memory, it suffices to bound the joint-gap in the composite MDP:

$$\mathcal{L}_{\text{joint},\varepsilon}(\pi)\le\Gamma_{\text{joint},\varepsilon}(\pi\parallel\pi^\star).$$

We bound this directly from Proposition J.2. The final statement follows from Theorem 13 in the same way that it does in the proof of Theorem 3.

The only remaining modification, then, is to evaluate the additional additive terms colored in purple in Proposition J.2; we will show that $\textsc{ErrTvc}$ as defined in (J.7) suffices as an upper bound. We have two cases. In both, let $d_0=\tau_{\text{obs}}d_x+(\tau_{\text{obs}}-1)d_u$. As $\mathsf{d}_{\textsc{tvc}}$ measures the distance between the chunks $\mathsf{o}_h=\phi_{\mathrm{o}}(\mathsf{s}_h),\tilde{\mathsf{o}}_h=\phi_{\mathrm{o}}(\mathsf{s}'_h)$, which have dimension d_0, and since we $\phi_{\mathrm{o}}\circ\mathsf{W}_\sigma(\cdot)=\mathcal{N}(\cdot,\sigma^2\mathbf{I}_{d_0})$, we have

$$\mathsf{d}_{\textsc{tvc}}(\phi_{\mathrm{o}}\circ\mathsf{s},\phi_{\mathrm{o}}\circ\mathsf{s}')\mid\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})\overset{\text{dist}}{=}\|\boldsymbol{\gamma}\|,\quad\boldsymbol{\gamma}\sim\mathcal{N}(0,\sigma^2\mathbf{I}_{d_0})\tag{J.8}$$

General $\gamma(\cdot)$. Eq. (J.8) and Lemma J.7 imply that

$$\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\textsc{tvc}}(\mathsf{s},\mathsf{s}')]\le\exp(-r^2/16\sigma^2),\quad r\ge 4\sigma d_0\log(9).$$

Hence, if $u_0=\gamma(8\sigma d_0\log(9))$, then

$$\mathbb{P}[\mathsf{d}_{\textsc{tvc}}(\mathsf{s},\mathsf{s}')>\gamma^{-1}(u)/2]\le\exp(-\gamma^{-1}(u)^2/64\sigma^2),\quad u\ge u_0.$$

Thus, as probabilities are at most one,

$$\int_0^\infty\max_{\mathsf{s}}\mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\textsc{tvc}}(\mathsf{s},\mathsf{s}')>\gamma^{-1}(u)/2]\mathrm{d}u\le u_0+\int_{u_0}^\infty e^{-\frac{\gamma^{-1}(u)^2}{64\sigma^2}}\mathrm{d}u,$$

as needed.

Linear $\gamma(\cdot)$. In the special case where $\gamma(u) = c(u)$, Eq. (J.8) 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 J.2. The corresponding additive term is then $2Hc\mathbb{E}[\|\boldsymbol{\gamma}\|]$. By Jensen's inequality, $\mathbb{E}[\|\boldsymbol{\gamma}\|] \le \sqrt{\mathbb{E}[\|\boldsymbol{\gamma}\|^2]} = \sqrt{\sigma^2 d_0} = \sigma\sqrt{d_0}$, as needed. □

J.4.2 Proof of Proposition J.2

Define the shorthand

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

and recall that Theorem 4 ensures $\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\text{rep}}) \le B$. Further, recall from Definition D.1 that

$$\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\text{rep}}) = \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\max_{h\in[H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\text{rep}}_{h+1}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}_h^{\text{rep}}, \hat{\mathsf{a}}_h)\} > \varepsilon\right],$$

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 $(\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\text{rep}}}$ with $\mathbb{P}_{\mu_1}[\hat{\mathsf{s}}_1 = \mathsf{s}_1^{\text{rep}}] = 1$. 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 $(\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\text{rep}}}$ with $\mathbb{P}_{\mu_2}[\mathsf{s}_1^\star = \mathsf{s}_1^{\text{rep}}] = 1$. By the "gluing lemma" (Lemma F.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

$$\begin{aligned}\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star) &\le \mathbb{P}_{\tilde{\mu}_1}\left[\max_{h\in[H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}_h^\star, \hat{\mathsf{a}}_h)\} > \varepsilon\right] \\ &\le \mathbb{P}_{\mu_1}\left[\max_{h\in[H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^{\text{rep}}_{h+1}, \hat{\mathsf{s}}_{h+1}), \mathsf{d}_{\mathcal{A}}(\mathsf{a}_h^{\text{rep}}, \hat{\mathsf{a}}_h)\} > \varepsilon\right] \\ &\quad + \mathbb{P}_{\mu_2}\left[(\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H}) \ne (\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H})\right].\end{aligned}$$

Passing to the infinum over μ_1, μ_2,

$$\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star) \le \underbrace{\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\text{rep}})}_{\le B} + \inf_{\mu_2} \mathbb{P}_{\mu_2}\left[(\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H}) \ne (\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H})\right],$$

where again μ_2 quantify couplines of $(\mathsf{s}^\star_{1:H+1}, \mathsf{a}^\star_{1:H}) \sim \mathsf{D}_{\pi^\star}$ and $(\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H}) \sim \mathsf{D}_{\pi^\star_{\text{rep}}}$ with $\mathbb{P}_{\mu_2}[\mathsf{s}_1^\star = \mathsf{s}_1^{\text{rep}}] = 1$. Bounding the infinum over μ_2 with Proposition J.4, we have

$$\Gamma_{\text{joint},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star) \le B + \sum_{h=1}^H \mathbb{E}_{\mathsf{s}_h^\star} \mathsf{TV}(\pi_h^\star(\mathsf{s}_h^\star), \pi^\star_{\text{rep},h}(\mathsf{s}_h^\star))$$

To conclude, it suffices to show the following bound:

Claim J.3. For any $\mathsf{s} \in \mathcal{S}$, $h \in [H]$, and $\tilde{r} \ge 0$, $\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi^\star_{\text{rep},h}(\mathsf{s})) \le \int_0^\infty \max_{\mathsf{s}} \max_{\mathsf{s}} \mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\text{TVC}}(\mathsf{s}, \mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]$.

Proof. To show this claim, we note that we can represent (via the notation in Appendix H.3) $\pi^\star_{\text{rep},h}(\mathsf{s}) = \pi_h^\star \circ \mathsf{W}^\star_{\text{rep},h}(\mathsf{s})$, where $\mathsf{W}^\star_{\text{rep},h}$ is the replica-kernel defined in Definition H.5. Thus, we can construct a coupling of $\mathsf{a}^\star \sim \pi_h^\star(\mathsf{s})$ and $\mathsf{a}^{\text{rep}} \sim \pi^\star_{\text{rep},h}(\mathsf{s})$ by introducing an intermediate state $\mathsf{s}' \sim \mathsf{W}^\star_{\text{rep},h}(\mathsf{s})$ and $\mathsf{a}^{\text{rep}} \sim \pi^\star(\mathsf{s}')$. By Lemma F.4, the fact that TV distance is bounded by one, and the assumption that $\pi^\star$ is $\tilde{\gamma}$-TVC, we then have

$$\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi^\star_{\text{rep},h}(\mathsf{s})) \le \mathbb{E}_{\mathsf{s}'\sim\mathsf{W}^\star_{\text{rep},h}(\mathsf{s})} \mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi_h^\star(\mathsf{s}')).$$

Recall the well-known formula that, for a non-negative random variable X, $\mathbb{E}[X] = \int_0^\infty \mathbb{P}[X > u]\mathrm{d}u$ [22]. From this formula, we find

$$\begin{aligned}\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi^\star_{\text{rep},h}(\mathsf{s})) &\le \int_0^\infty \mathbb{P}[\mathsf{TV}(\pi_h^\star(\mathsf{s}), \pi_h^\star(\mathsf{s}')) > u]\mathrm{d}u \\ &\overset{(i)}{\le} \int_0^\infty \mathbb{P}[\mathsf{d}_{\text{TVC}}(\mathsf{s}, \mathsf{s}') > \tilde{\gamma}^{-1}(u)]\mathrm{d}u\end{aligned}$$

where in (i) we used that $\mathsf{TV}(\pi^\star_h(\mathsf{s}), \pi^\star_h(\mathsf{s}')) \le \tilde{\gamma}(\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}'))$ and that, as $\tilde{\gamma}(\cdot)$ is strictly increasing, we have the equality of events $\{\mathsf{TV}(\pi^\star_h(\mathsf{s}), \pi^\star_h(\mathsf{s}')) > u\} = \{\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \gamma^{-1}(u)\}$. Arguing as in the proof of Lemma H.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

$$\mathsf{TV}(\pi^\star_h(\mathsf{s}), \pi^\star_{\mathrm{rep},h}(\mathsf{s})) \le \int_0^\infty \max_{\mathsf{s}} \mathbb{P}_{\mathsf{s}'\sim\mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\mathrm{TVC}}(\mathsf{s},\mathsf{s}') > \tilde{\gamma}^{-1}(u)/2]\mathrm{d}u$$

which proves the first guarantee. □

With the above claim proven, we conclude the proof of the first statement of Proposition J.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. □

J.5 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 12. Define the error term

$$\Delta_{\mathsf{TV},h}(\hat{\pi}) := \sum_{h=1}^H \mathbb{E}_{\mathsf{o}_h\sim\mathcal{D}_{\mathrm{exp},h}} \Delta_{(\varepsilon)}\left(\pi^\star_h(\mathsf{o}_h), \hat{\pi}_h(\mathsf{o}_h)\right)\big|_{\varepsilon=0}$$

We have the characterization $\Delta_{\mathsf{TV},h}(\hat{\pi}) = \mathbb{E}_{\mathsf{o}_h}\mathsf{TV}(\pi^\star_h(\mathsf{o}_h), \hat{\pi}_h(\mathsf{o}_h))$ where o_h has the distribution induced by $\hat{\pi}$, and $\pi^\star(\mathsf{o}_h)$ denotes the distribution of $\mathsf{a}_h \mid \mathsf{o}_h$ under $\mathcal{D}_{\mathrm{exp}}$ as in Definition J.3. Then, under no additional assumption (not even those in Section 3),

$$\mathcal{L}_{\mathrm{fin},\varepsilon=0}(\hat{\pi}_\sigma) \le \mathcal{L}_{\mathrm{marg},\varepsilon=0}(\hat{\pi}_\sigma) \le \sum_{h=1}^H \Delta_{\mathsf{TV},h}(\hat{\pi})$$

In in addition $\pi^\star$ has τ-bounded memory(Definition J.5) for $\tau \le \tau_{\mathrm{obs}}$, then for $\mathcal{L}_{\mathrm{joint},\varepsilon}$ as in Definition J.4,

$$\mathcal{L}_{\mathrm{joint},\varepsilon=0}(\hat{\pi}) \le \sum_{h=1}^H \Delta_{\mathsf{TV},h}(\hat{\pi})$$

The first part of the above theorem is a direct consequence Proposition F.3. Then, the second part of the toerem follows by combining the proposition below in the composite MDP, together with the correct instantiations for control, and Lemma J.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 J.4. Consider the composite MDP setting of Appendix D. Then, there exists a coupling

$$\mathsf{TV}(\mathsf{D}_{\hat{\pi}}, \mathsf{D}_{\pi^\star}) \le \sum_{h=1}^H \mathbb{E}_{\mathsf{s}^\star_h\sim\mathsf{P}^\star_h}\mathsf{TV}(\pi^\star_h(\mathsf{s}^\star_h), \hat{\pi}_h(\mathsf{s}^\star_h))$$

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 $\mathbb{P}_\mu[\mathsf{s}^\star_1 = \hat{\mathsf{s}}_1]$.

Proof of Proposition J.4. This is a direct consequence of Lemma J.9, with $\mathsf{P}_1 \leftarrow \mathsf{P}_{\mathrm{init}}$, and Q_{h+1} corresponding to the kernel for sampling $\mathsf{a}^\star_h \sim \pi^\star(\mathsf{s}^\star_h)$ and incrementing the dynamics $\mathsf{s}^\star_{h+1} = F_h(\mathsf{s}^\star_h, \mathsf{a}^\star_h)$, and Q'_h the same for $\hat{\mathsf{a}}_h \sim \hat{\pi}_h(\hat{\mathsf{s}}_h)$, and similar incrementing of the dynamics. □

J.6 Consequence for expected costs

Finally, we prove Proposition J.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 J.5. Recall the marginal and final imitation losses in Definition 2.2, and also the joint imitation loss in Definition J.4. 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[12] (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 J.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,

$$\begin{aligned}|\mathbb{E}[\mathfrak{J}(\boldsymbol{\rho}_T)] - \mathbb{E}[\mathfrak{J}(\boldsymbol{\rho}'_T)]| &= |\sum_{t=1}^T (\mathbb{E}[\ell_{t,1}(\mathbf{x}_{t+1})] - \mathbb{E}[\ell_{t,1}(\mathbf{x}'_{t+1})) + (\mathbb{E}[\ell_{t,2}(\mathbf{u}_t)] - \mathbb{E}[\ell_{t,2}(\mathbf{u}'_1))| \\ &\le \sum_{t=1}^T |\mathbb{E}[\ell_{t,1}(\mathbf{x}_{t+1})] - \mathbb{E}[\ell_{t,1}(\mathbf{x}'_{t+1})| + |\mathbb{E}[\ell_{t,2}(\mathbf{u}_t)] - \mathbb{E}[\ell_{t,2}(\mathbf{u}'_1)|\end{aligned}$$

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). □

[12] Of course, Lipschitznes in other norms can be derived, albeit with different T dependence

J.7 Useful Lemmata

J.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 J.3, which induces distribution $\mathcal{D}_{\pi^\star}$.

Lemma J.6. There exists a sequence of probability kernels $\pi_h^\star$ mapping $\mathsf{o}_h \to \Delta(\mathcal{A})$ such that the chunking policy $\pi^\star = (\pi_h^\star)_{1 \le h \le H}$ satisfies the following:

(a) $\pi_h^\star(\mathsf{o}_h)$ is equal to the almost-sure conditional probability of a_h conditioned on o_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 $\mathsf{s}_h \sim \mathsf{P}_h^\star$ (as defined in Definition J.3) is the same as the marginals of each s_h under $\boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}$, and hence, $(\mathsf{s}_h, \mathsf{a}_h) \sim \mathcal{D}_{\mathrm{exp}}$ has the same distribution as $(\mathsf{s}_h, \mathsf{a}_h)$ where $\mathsf{s}_h \sim \mathsf{P}_h^\star$, $\mathsf{a}_h \mid \mathsf{s}_h \sim \pi_h^\star(\mathsf{s}_h)$.

(c) If $\mathcal{D}_{\mathrm{exp}}$ has τ-bounded memory (Definition J.5) and if $\tau \le \tau_{\mathrm{obs}}$, then the joint distribution of $\boldsymbol{\rho}_T$ induced by $\pi^\star$ is equal to the joint distribution over $\boldsymbol{\rho}_T$ under $\mathcal{D}_{\mathrm{exp}}$.

(d) Again, let $\mathsf{P}_h^\star$ be as defined in Definition J.3. Consider any sequence of kernels $\mathsf{W}_{1:H}$ satisfying the constraint that $\phi_{\mathrm{o}} \circ \mathsf{W}_h(\mathsf{s}) \ll \phi_{\mathrm{o}} \circ \mathsf{P}_h^\star$ forall s, h. Consider a sequence of states $\mathsf{s}_{1:H+1}$ drawn as in Definition H.9 by $\mathsf{s}_1 \sim \mathsf{P}_{\mathrm{init}}$, $\tilde{\mathsf{s}}_h \sim \mathsf{W}_h(\mathsf{s}_h)$, and $\mathsf{s}_a \sim \pi_h^\star(\tilde{\mathsf{s}}_h)$, $\mathsf{s}_{s+1} = F_h(\mathsf{s}_h, \mathsf{a}_h)$. Then, let $\tilde{\mathsf{o}}_h = \phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h)$ and $\mathsf{a}_h \sim \pi_h^\star(\mathsf{a}_h)$. Then the distribution of $(\tilde{\mathsf{o}}_h, \mathsf{a}_h)$ is absolutely continuous with respect to the distribution of $(\tilde{\mathsf{o}}_h, \mathsf{a}_h) \sim \mathcal{D}_{\mathrm{exp}}$.

Remark J.5 (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

$$\mathsf{Q}_h(\mathsf{o}_h) = \mathbb{P}_{\mathsf{a}_{1:h-1}, \boldsymbol{\rho}_{1:t_h - \tau_{\mathrm{obs}}-1}, \mathsf{a}_{h:H}, \boldsymbol{\rho}_{t_h:T+1} \mid \mathsf{o}_h}$$
$$\mathsf{Q}_h^{\otimes}(\mathsf{o}_h) = \mathbb{P}_{\mathsf{a}_{1:h-1}, \boldsymbol{\rho}_{1:t_h - \tau_{\mathrm{obs}}-1} \mid \mathsf{o}_h} \otimes \mathbb{P}_{\mathsf{a}_{h:H}, \boldsymbol{\rho}_{t_h:T+1} \mid \mathsf{o}_h}.$$

which describes the conditional distribution of the whole trajectory without o_h and the product-distribution of the conditional distributions of the before-o_h part of the trajectory, and after o_h-part. Under the condition

$$\mathbb{E}_{\mathsf{o}_h \text{ from } \boldsymbol{\rho}_T \sim \mathcal{D}_{\mathrm{exp}}} \mathsf{TV}\left(\mathsf{Q}_h(\mathsf{o}_h), \mathsf{Q}_h^{\otimes}(\mathsf{o}_h)\right) \le \varepsilon_{\mathrm{mix}}(\tau_{\mathrm{obs}}),$$

which measures how close the before- and after-o_h parts of the trajectory are to being conditionally independent, one can leverage Lemma J.9 to show that

$$\mathsf{TV}(\mathcal{D}_{\pi^\star}, \mathcal{D}_{\mathrm{exp}}) \le H\varepsilon_{\mathrm{mix}}(\tau_{\mathrm{obs}})$$

Lemma J.6 corresponds to the special when when $\varepsilon_{\mathrm{mix}} = 0$.

Proof of Lemma J.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 7.

Part (b). This follows by marginalization and Markovianity of the dynamics. Specifically, let $(\boldsymbol{\rho}_T^\star, \mathsf{a}_{1:H}^\star$ 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 $\mathsf{o}_h^\star$ denote observation chunks of $\boldsymbol{\rho}_T^\star$, and let o_h observation chunks of $\boldsymbol{\rho}_T$ (length $\tau_{\mathrm{obs}} - 1$); similarly, denote by $\mathsf{s}_h^\star$ and s_h the respective trajectory chunks (length $\tau_{\mathrm{chunk}} \ge \tau_{\mathrm{obs}}$).

We argue inductively that the trajectory chunks $\mathsf{s}_h^\star$ and s_h are identically distribued for each h. For $h = 1$, $\mathsf{s}_1^\star$ and s_1 are identically distributed according to $\mathcal{D}_{\mathbf{x}_1}$. Now assume we have show that $\mathsf{s}_h^\star$ and s_h are identically distributed. As observation chunks are sub-chunks of trajectory chunks, this means that $\mathsf{o}_h^\star$ and o_h are identically distributed. By part (a), it follows that $(\mathsf{o}_h^\star, \mathsf{a}_h^\star)$ and $(\mathsf{o}_h, \mathsf{a}_h)$ are identically distributed. In particular, $(\mathbf{x}_{t_h}^\star, \mathsf{a}_h^\star)$ and $(\mathbf{x}_{t_h}, \mathsf{a}_h)$ are identically distributed, where $\mathbf{x}_{t_h}^\star$ (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, $\mathsf{s}_{h+1}^\star$ and s_{h+1} are functions of $(\mathbf{x}_{t_h}^\star, \mathsf{a}_h^\star)$ and $(\mathbf{x}_{t_h}, \mathsf{a}_h)$, respectively, $\mathsf{s}_{h+1}^\star$ and s_{h+1} are identically distributed, as needed.

Part (c). When $\mathcal{D}_{\mathrm{exp}}$ has τ-bounded memory and $\tau \le \tau_{\mathrm{obs}}$, then we have the almost-sure equality

$$\mathbb{P}_{\mathcal{D}_{\mathrm{exp}}}[\mathsf{a}_h \in \cdot \mid \mathbf{x}_{1:t_h}, \mathbf{u}_{1:t_h}] = \mathbb{P}_{\mathcal{D}_{\mathrm{exp}}}[\mathsf{a}_h \in \cdot \mid \mathsf{o}_h] = \pi_h^\star(\mathsf{o}_h)[\mathsf{a}_h \in \cdot].$$

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.

Part (d) That the distributions of $\tilde{\mathsf{o}}_h = \phi_{\mathrm{o}}(\tilde{\mathsf{s}}_h)$ under the construction in part (d) is absolutely continuous with respect to $\tilde{\mathsf{o}}_h$ under $\mathcal{D}_{\mathrm{exp}}$ coincide is immediate from the condition of W_h. The second part follows from part (a). □

J.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 J.7. Suppose that $\gamma \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ 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

$$||\gamma|| \le 2\sigma \cdot \sqrt{2d\log(9) + 2\log\left(\frac{1}{p}\right)} \le 2\sigma \cdot \sqrt{5d + 2\log\left(\frac{1}{p}\right)}$$

Moreover, for $r \ge 4\sigma\sqrt{d\log(9)}$, $\mathbb{P}[||\gamma|| \ge r] \le \exp(-r^2/16\sigma^2)$.

Proof. We apply the standard covering based argument as in, e.g., Vershynin [71, Section 4.2]. Note that

$$||\gamma|| = \sup_{\mathbf{w} \in \mathcal{S}^{d-1}} \langle \gamma, \mathbf{w} \rangle,$$

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

$$\sup_{\mathbf{w} \in \mathcal{S}^{d-1}} \langle \gamma, \mathbf{w} \rangle \le 2 \cdot \max_{\mathbf{w} \in \mathcal{U}} \langle \mathbf{w}, \gamma \rangle.$$

A classical volume argument (see for example, Vershynin [71, 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 $r > 0$,

$$\mathbb{P}(\langle \mathbf{w}, \gamma \rangle > r) \le e^{-\frac{r^2}{2\sigma^2}}.$$

Thus by a union bound, we have

$$\mathbb{P}(||\gamma|| > r) \le |\mathcal{U}| \cdot \max_{\mathbf{w} \in \mathcal{U}} \mathbb{P}\left(||\gamma|| > \frac{r}{2}\right) \le 9^d \cdot e^{-\frac{r^2}{8\sigma^2}}.$$

Inverting concludes the proof.

□

The second lemma shows that the relevant smoothing kernel is TVC:

Lemma J.8. For any $\sigma > 0$, let ϕ_{o} and W_σ be as in Definition J.1 kernel, then W_σ is γ_{TVC}-TVC for with respect to $\mathsf{d}_{\mathrm{TVC}}$ (as defined in Appendix D.2)

$$\gamma_{\mathrm{TVC}}(u) = \frac{u\sqrt{2\tau_{\mathrm{obs}} - 1}}{2\sigma}.$$

Proof. Recall that ϕ_{o} denotes projection onto the $\mathcal{O}$-component of the direct decomposition in Definition H.1, i.e. projects onto the observation chunk o_h. We apply Pinsker's inequality [52]: Then, for for $\mathsf{s}, \mathsf{s}' \in \mathbb{R}^p$, we have

$$\mathrm{TV}(\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}), \phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}')) \le \sqrt{\frac{1}{2} \cdot \mathrm{D}_{\mathrm{KL}}(\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}) \parallel \phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}'))}.$$

Note that for $\mathsf{s} = \mathsf{s}_h$ with corresponding observation chunk o_h $\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}) \sim \mathcal{N}(\mathsf{o}_h, \sigma^2 \mathbf{I})$. Similarly, for o'_h corresponding to s', $\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}') \sim \mathcal{N}(\mathsf{o}'_h, \sigma^2 \mathbf{I})$. Hence,

$$\mathrm{D}_{\mathrm{KL}}\left(\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}) \parallel \phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}')\right) \leq \frac{||\mathsf{o}_h - \mathsf{o}'_h||^2}{2\sigma^2}.$$

Thus, we conclude $\mathrm{TV}\left(\phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}), \phi_{\mathrm{o}} \circ \mathsf{W}_\sigma(\mathsf{s}')\right) \leq \frac{||\mathsf{o}_h - \mathsf{o}'_h||}{2\sigma}$. Finally, we upper bound the Euclidean norm $||\mathsf{o}_h - \mathsf{o}'_h||$ of vectors consistening of $2\tau_{\mathrm{obs}} - 1$ sub-vectors via $\mathsf{d}_{\mathrm{TVC}}$ (which is the maximum Euclidean norm of these subvectors) via $||\mathsf{o}_h - \mathsf{o}'_h|| \leq \sqrt{2\tau_{\mathrm{obs}} - 1}\mathsf{d}_{\mathrm{TVC}}(\mathsf{s}, \mathsf{s}')$. $\square$

J.7.3 Total Variation Telescoping

Lemma J.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,

$$\mathsf{TV}(\mathsf{P}_{H+1}, \mathsf{P}'_{H+1}) \leq \sum_{h=1}^{H} \mathbb{E}_{Y_{1:h} \sim \mathsf{P}_h} \mathsf{TV}(\mathsf{Q}_{h+1}(\cdot \mid Y_{1:h}), \mathsf{Q}'_{h+1}(\cdot \mid Y_{1:h}))$$

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

$$\mathbb{P}_\mu[Y_1 = Y'_1] = 1, \quad \mathbb{P}_\mu[Y_{1:H+1} \neq Y'_{1:H+1}] \leq \sum_{h=1}^{H} \mathbb{E}_{Y_{1:h} \sim \mathsf{P}_h} \mathsf{TV}(\mathsf{Q}_{h+1}(\cdot \mid Y_{1:h}), \mathsf{Q}'_{h+1}(\cdot \mid Y_{1:h})).$$

Proof. To prove the first part of the lemma, define $\mathsf{Q}'_{i,j}$ for $2 \leq i \leq j \leq H+1$ 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 convection $\mathrm{law}(Q'_{H+2,H+1}; \mathsf{P}_{H+1}) = \mathsf{P}_{H+1}$. 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,

$$\mathsf{TV}(\mathsf{P}_{H+1}, \mathsf{P}'_{H+1}) \leq \sum_{h=1}^{H} \mathsf{TV}(\mathsf{P}^{(i)}, \mathsf{P}^{(i+1)})$$

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,

$$\begin{aligned}
\mathsf{TV}(\mathsf{P}^{(i)}, \mathsf{P}^{(i+1)}) &= \mathsf{TV}(\mathrm{law}(Q'_{i+2,H+1}; \mathrm{law}(\mathsf{Q}'_{i+1}; \mathsf{P}_i)), \mathrm{law}(Q'_{i+2,H+1}; \mathrm{law}(\mathsf{Q}_{i+1}; \mathsf{P}_i)) \\
& \qquad \text{(Lemma F.4)} \\
&= \mathsf{TV}(\mathrm{law}(\mathsf{Q}'_{i+1}; \mathsf{P}_i), \mathrm{law}(\mathsf{Q}_{i+1}; \mathsf{P}_i) \\
&= \mathbb{E}_{Y_{1:i} \sim \mathsf{P}_i} \mathsf{TV}(\mathsf{Q}'_i(Y_{1:i}), \mathsf{Q}_i(Y_{1:i})). \qquad \text{(Corollary F.1)}
\end{aligned}$$

This completes the first part of the demonstration (noting symmetry of TV). The second part follows from Corollary F.1, by letting $Y \leftarrow Y_1$, and $X \leftarrow Y_{2:H+1}$ in that lemma. $\square$

K Stability in the Control System

This section proves our various stability conditions. Precisely, we establish the following guarantee:

Proposition K.1. Let $c_\gamma, c_\xi, \bar{c}_\beta, \bar{c}_\gamma, L_\beta$ be the constants defined in Assumption 3.1. In terms of these, define

$$\begin{aligned}
\alpha &= \bar{c}_\beta(4\bar{c}_\gamma \min\{c_\gamma, c_\xi/4\bar{c}_\gamma\} + c_\xi) \\
c_1 &:= 4\bar{c}_\gamma\bar{c}_\beta(2 + \alpha L_{\mathrm{stab}} + 2R_{\mathrm{dyn}}) \\
c_2 &:= \max\{1, c_1\}^{-1} \min\{c_\gamma, c_\xi/2\bar{c}_\gamma\} \\
c_3 &:= \frac{1}{L_\beta}\log(2e\bar{c}_\beta) \\
c_4 &:= c_\xi/2 \\
c_5 &:= 2\bar{c}_\beta
\end{aligned}$$

For actions $\mathsf{a} = (\kappa_k)_{1\le k\le\tau_{\mathrm{chunk}}}$ where $\kappa_k(\mathbf{x}) = \bar{\mathbf{u}}_k + \bar{\mathbf{K}}_k(\mathbf{x} - \bar{\mathbf{x}}_k)$ are affine primitive controllers, define $\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') := \max_{1\le k\le\tau_{\mathrm{chunk}}}(\|\bar{\mathbf{u}}_k - \bar{\mathbf{u}}'_k\| + \|\bar{\mathbf{x}}_k - \bar{\mathbf{x}}'_k\| + \|\bar{\mathbf{K}}_k - \bar{\mathbf{K}}'_k\|)$, and let

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') := c_1\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \cdot \mathbf{I}_\infty\{\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le c_2\}$$

Then, if $\tau_{\mathrm{chunk}} \ge c_3/\eta$, the policy $\pi^\star$ as defined in Definition J.3 satisfies $(r_{\textsc{ips}}, \gamma_{\textsc{ips},\textsc{Tvc}}, \gamma_{\textsc{ips},\mathcal{S}}, \mathsf{d}_{\textsc{ips}})$-restricted IPS (Definition H.9) with $\mathsf{d}_{\mathcal{A}}$ as above, and with

$$r_{\textsc{ips}} = c_4, \quad \gamma_{\textsc{ips},\textsc{Tvc}}(u) = c_5 u \exp(-\eta(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})/L_{\mathrm{stab}}), \quad \gamma_{\textsc{ips},\mathcal{S}}(u) = c_5 u.$$

Appendix K.1 proves Proposition K.1, based on a lemma whose proof is given in Appendix K.2.

In what follows, we justify our assumption of a stabilizing synthesis oracle, Assumption 3.1. First, Appendix K.4 shows that if the system dynamics are *smooth*, than time-varying affine controllers whose gains stabilize the Jacobian linearization of the given system satisfy Definition 3.1. This result is stated in Appendix K.3, along with the requisite assumptions, and proven in Appendix K.4 , based on a lemma whose proof is given in Appendix K.5.

Finally, Appendix K.6 shows how a synthesis oracle can produce gains which stabilize the linearized dynamics can obtained by solving the Riccati equation, assuming sufficient dynamical regularity. Finally, Appendix K.7 gives the solutions to various scalar recursions used in the proofs throughout.

K.1 Proof of Proposition K.1

We now translate the incremental stability guarantee about into the IPS guarantee needed by Proposition K.1. The core technical ingredient is the following lemma, whose proof we defer to Appendix K.2.

Lemma K.2 (Trajectory-tracking via t-ISS)**.** Consider a given sequence $(\tilde{\mathbf{x}}_{t_h})$, and suppose that $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$ is local-t-ISS at $\tilde{\mathbf{x}}_{t_h}$ for each $1 \le h \le H$ (with parameters as in Definition 3.1. Consider consistent trajectories $(\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T})$, $(\mathbf{x}'_{1:T+1}, \mathbf{u}_{1:T})$ satisfying

$$\mathbf{u}_t = \kappa_t(\mathbf{x}_t), \quad \mathbf{u}'_{t+1} = \kappa'_t(\mathbf{x}'_t), \quad \mathbf{x}_1 = \mathbf{x}'_1, \quad \max_h \|\tilde{\mathbf{x}}_{t_h} - \mathbf{x}_{t_h}\| \le r \le c_\xi/2$$

Further, define the sequence $(\tilde{\mathbf{x}}_t)$ by setting, for each h,

$$\hat{\mathbf{x}}_{t_h} := \tilde{\mathbf{x}}_{t_h}, \quad \hat{\mathbf{x}}_{t_h+i} := f(\hat{\mathbf{x}}_{t_h+i}, \kappa_t(\hat{\mathbf{x}}_{t_h+i-1})), \quad i \in \{1, 2, \dots, \tau_{\mathrm{chunk}} - 1\} \tag{K.1}$$

Then, the following guarantees hold

(a) $\|\mathbf{x}_{t_h+i} - \hat{\mathbf{x}}_{t_h+i}\| \le \beta(r, i)$ for $i \in \{0, 1, 2, \dots, \tau_{\mathrm{chunk}} - 1\}$ and $h \in [H]$.

(b) Suppose that $\varepsilon > 0$ satisfies

$$\gamma^{-1}(\beta(2\gamma(\varepsilon), \tau_{\mathrm{chunk}}) \le \varepsilon \le \min\{c_\gamma, \gamma^{-1}(c_\xi/4)\}$$

and that one of the following hold

$$\max_{1\le t\le T} \sup_{\|\delta\mathbf{x}\|\le\alpha(\varepsilon)} \|\kappa_t(\mathbf{x}_t + \delta\mathbf{x}) - \kappa'_t(\mathbf{x}_t + \delta\mathbf{x})\| \le \varepsilon, \quad \alpha(\varepsilon) := 2\beta(2\gamma(\varepsilon), 0), \tag{K.2}$$

$$\max_{1\le t\le T} \sup_{\|\delta\mathbf{x}\|\le\alpha(\varepsilon)+\beta(r,0)} \|\kappa_t(\hat{\mathbf{x}}_t + \delta\mathbf{x}) - \kappa'_t(\hat{\mathbf{x}}_t + \delta\mathbf{x})\| \le \varepsilon, \tag{K.3}$$

Then for all $h \in [H]$, $i \in \{0, 1, \dots, \tau_{\mathrm{chunk}}\}$, and $t \in [T]$,

$$\|\mathbf{u}_t - \mathbf{u}'_t\| \le \varepsilon \le \alpha \quad \|\mathbf{x}_{t_h+i} - \mathbf{x}'_{t_h+i}\| \le \beta(2\gamma(\varepsilon), i) + \gamma(\varepsilon) \le \alpha$$

As a consequence, we derive the following reduction from IPS and input-stability in the composite MDP to $\mathtt{t\text{-}ISS}$.

Definition K.1 (Instantiation of the composite MDP for general primitive controllers)**.** In this section, we summarize the instantiation of the MDP in Appendix J:

- States $\mathsf{s}_h = \mathsf{s}_h$ and $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}, \mathsf{d}_{\mathrm{IPS}}$ are just as in Appendix D. Moreover, $\mathsf{o}_h = \phi_{\mathrm{o}} \circ \mathsf{s}_h$.
- The kernel $\mathsf{W}_\sigma(\cdot)$ is the same as (J.1) in Appendix J, applying $\mathcal{N}(0, \sigma^2 I)$ noise in the coordinates in o_h.
- Actions a_h are sequences of affine primitive controllers $\kappa_{1:\tau_{\mathrm{chunk}}}$.
- $\pi^\star = (\pi_h^\star)$ be the policy induced by the conditional distribution of $\mathsf{a} \mid \mathsf{o}_h$ as constructed in Definition J.3 in Appendix J.

Lemma K.3. Instantiate the composite MDP as in Definition K.1, with $\pi^\star$ as in Definition J.3. Furthermore, suppose that under $(\boldsymbol{\rho}_T, \mathsf{a}_{1:H}) \sim \mathcal{D}_{\mathrm{exp}}$ with $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_T)$, the following both hold with probability one:

- Each action a_h satisfies our notion of incremental stability (Definition 3.1) with moduli $\gamma(\cdot), \beta(\cdot, \cdot)$, constants c_γ, c_ξ
- $\mathbf{x}_t \in \mathcal{X}_0$ for some set $\mathcal{X}_0 \subset \mathbb{R}^{d_x}$, and $\kappa_t \in \mathcal{K}_0$ for some set of primitive controllers $\mathcal{K}_0 \subset \mathcal{K}$.[13]

Finally, let $\varepsilon_0 > 0$ satisfy (E.1), that is:

$$\gamma^{-1}(\beta(2\gamma(\varepsilon_0), \tau_{\mathrm{chunk}}) \le \varepsilon_0 \le \min\{c_\gamma, \gamma^{-1}(c_\xi/4)\}, \tag{K.4}$$

For given $\alpha > 0$, let $\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}')$ be a function which, for all composite actions $\mathsf{a} = \kappa_{1:\tau_{\mathrm{chunk}}}$ satisfying $\kappa_i \in \mathcal{K}_0$ all arbitrary composite actions $\mathsf{a}' = \kappa_{1:\tau_{\mathrm{chunk}}} \in \mathcal{K}^{\tau_{\mathrm{chunk}}}$, satisfies

$$\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}') \ge \sup_{\mathbf{x} \in \mathcal{X}_0} \sup_{\delta\mathbf{x} : \|\delta\mathbf{x}\| \le \alpha} \max_{1 \le i \le \tau_{\mathrm{chunk}}} \|\kappa_i(\mathbf{x}_i + \delta\mathbf{x}) - \kappa_i'(\mathbf{x}_i + \delta\mathbf{x})\|.$$

and let

$$\bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}; \alpha) := \psi(\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}')) \cdot \mathcal{I}_\infty\left\{\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}') \le \varepsilon_0\right\}, \quad \psi(u) := 2\beta(2\gamma(u), 0).$$

Then, the following hold:

(a) $\pi^\star$ is input-stable with respect to $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}$ as defined in Appendix D

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') = \bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}'; \psi(\varepsilon)),$$

(b) For any $r_{\mathrm{IPS}} \le c_\xi/2$, $\pi^\star$ is $(r_{\mathrm{IPS}}, \gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}, \mathsf{d}_{\mathrm{IPS}})$- restricted-IPS (Definition H.9) with

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') = \bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}'; \psi(\varepsilon) + \beta(r_{\mathrm{IPS}}, 0)), \quad \gamma_{\mathrm{IPS,TVC}}(r) = \beta(r, \tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}}), \quad \gamma_{\mathrm{IPS},\mathcal{S}}(r) = \beta(r, 0),$$

Note that the above lemma holds for general forms of incremental stability. Let us now instantiate it for the form of incremental stability of the form established in Proposition K.6.

Corollary K.1. Suppose that $\gamma(\varepsilon) = \bar{c}_\gamma \cdot \varepsilon$ and $\beta(\varepsilon, k) = \bar{c}_\gamma \phi(k) \cdot \varepsilon$. Then, as long as we take

$$2\phi(\tau_{\mathrm{chunk}})\bar{c}_\beta \le 1, \quad \varepsilon_0 := \min\{c_\gamma, c_\xi/4\bar{c}_\gamma\},$$

and setting $\psi(\varepsilon) = \varepsilon$, we have that

(a) $\pi^\star$ is input-stable with $\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') = \bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}'; 4\bar{c}_\gamma\bar{c}_\beta\varepsilon_0)$.

(b) For any $r_{\mathrm{IPS}} = c_\xi$, $\pi^\star$ is $(r_{\mathrm{IPS}}, \gamma_{\mathrm{IPS,TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}, \mathsf{d}_{\mathrm{IPS}})$- restricted-IPS (Definition H.9) with

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') = 4\bar{c}_\beta\bar{c}_\gamma\bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}'; \bar{c}_\beta(4\bar{c}_\gamma\varepsilon_0 + r)), \quad \gamma_{\mathrm{IPS,TVC}}(r) = \bar{c}_\beta\phi(\tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})r, \quad \gamma_{\mathrm{IPS},\mathcal{S}}(r) = \bar{c}_\beta r$$

We are now ready to prove the main result of this appendix.

[13] This can be directly generalized to a constraint on the composite states s_h and composite actions a_h.

Proof of Proposition K.1. Note that, by assumption, we are in the regime of Corollary K.1, with $\phi(k) = e^{-L_\beta(k-1)}$ and $\varepsilon_0 := \min\{c_\gamma, c_\xi/4\bar{c}_\gamma\}$. We note that, under our assumption $L_\beta \le 1$,

$$\phi(k) = e^{L_\beta} e^{-L_\beta(k-1)} \le e \cdot e^{-L_\beta(k-1)}. \tag{K.5}$$

Hence, $2\phi(\tau_{\text{chunk}})\bar{c}_\beta \le 1$ for $\tau_{\text{chunk}} \ge c_3 = \log(2e\bar{c}_\beta)/L_\beta$.

Next, we develop D_α. Express the primitive controllers $\kappa_i = (\bar{\mathbf{u}}_i, \bar{\mathbf{x}}_i, \bar{\mathbf{K}}_i)$ and $\kappa_i' = (\bar{\mathbf{u}}_i', \bar{\mathbf{x}}_i', \bar{\mathbf{K}}_i')$. Recall

$$\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') = \max_{1 \le i \le \tau_{\text{chunk}}} \max\{\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_i'\| + \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_i'\| + \|\bar{\mathbf{K}}_i - \bar{\mathbf{K}}_i'\|\}.$$

By assumption, the expert distribution $\mathcal{D}_{\text{exp}}$ ensures that $\|\mathbf{x}_t\| \le R_{\text{dyn}}$ and that $\|\bar{\mathbf{K}}_t\| \le R_{\mathbf{K}}$. Moreover, it also ensures $\|\bar{\mathbf{x}}_t\| \le R_{\text{dyn}}$, since under the expert distribution, $\bar{\mathbf{x}}_t = \mathbf{x}_t$. Thus, to find an upper upper bound on the distance $\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}')$, it suffices to take $\mathcal{X}_0 = \{\mathbf{x} : \|\mathbf{x}\| \le R_{\text{dyn}}\}$ and bound the following quantity for all $\mathsf{a} = \kappa_{1:\tau_{\text{chunk}}}$ and $\mathsf{a}' = \kappa'_{1:\tau_{\text{chunk}}}$ for which $\kappa_i = (\bar{\mathbf{u}}_i, \bar{\mathbf{x}}_i, \bar{\mathbf{K}}_i)$ satisfies $\|\bar{\mathbf{x}}_i\| \le R_{\text{dyn}}$ and $\|\bar{\mathbf{K}}_i\| \le R_{\mathbf{K}}$:

$$\begin{aligned}
&\sup_{\mathbf{x}:\|\mathbf{x}\| \le R_{\text{dyn}}} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\| \le \alpha} \max_{1 \le t \le \tau_{\text{chunk}}} \|\kappa_t(\mathbf{x}_t + \delta\mathbf{x}) - \kappa_t'(\mathbf{x}_t + \delta\mathbf{x})\| \\
&= \sup_{\mathbf{x}:\|\mathbf{x}\| \le R_{\text{dyn}}} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\| \le \alpha} \max_{1 \le t \le \tau_{\text{chunk}}} \|\bar{\mathbf{u}}_t - \bar{\mathbf{u}}_t' + (\bar{\mathbf{K}}_t - \bar{\mathbf{K}}_t')(\mathbf{x}_t + \delta\mathbf{x}) + \bar{\mathbf{K}}_t\bar{\mathbf{x}}_t - \bar{\mathbf{K}}_t'\bar{\mathbf{x}}_t'\| \\
&\le \max_{1 \le t \le \tau_{\text{chunk}}} \|\bar{\mathbf{u}}_t - \bar{\mathbf{u}}_t'\| + \|\bar{\mathbf{K}}_t - \bar{\mathbf{K}}_t)\|(\alpha + \sup_{\mathbf{x}:\|\mathbf{x}\| \le R_{\text{dyn}}} \|\mathbf{x}\|) + \|\bar{\mathbf{K}}_t(\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_t')\|) + \|(\bar{\mathbf{K}}_t - \bar{\mathbf{K}}_t)\bar{\mathbf{x}}_t'\| \\
&\max_{1 \le t \le \tau_{\text{chunk}}} \|\bar{\mathbf{u}}_t - \bar{\mathbf{u}}_t'\| + \|\bar{\mathbf{K}}_t - \bar{\mathbf{K}}_t)\|(\alpha + R_{\text{dyn}}) + R_{\mathbf{K}}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_t'\|) + \|\bar{\mathbf{K}}_t - \bar{\mathbf{K}}_t\|\|\bar{\mathbf{x}}_t'\| \\
&\hfill (\|\bar{\mathbf{K}}_t\| \le R_{\mathbf{K}}) \\
&\le \max_{1 \le t \le \tau_{\text{chunk}}} \|\bar{\mathbf{u}}_t - \bar{\mathbf{u}}_t'\| + \|\bar{\mathbf{K}}_t - \bar{\mathbf{K}}_t)\|(\alpha + 2R_{\text{dyn}} + \|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_t'\|) + R_{\mathbf{K}}\|\bar{\mathbf{x}}_t - \bar{\mathbf{x}}_t'\| \\
&\hfill (\|\bar{\mathbf{x}}_t\| \le R_{\text{dyn}}) \\
&\le \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')(1 + R_{\mathbf{K}}\alpha + 2R_{\text{dyn}} + \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}'));
\end{aligned}$$

that is, we can take

$$\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}') = \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')(1 + R_{\mathbf{K}}\alpha + 2R_{\text{dyn}} + \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')).$$

Now, set $\alpha = 4\bar{c}_\beta\bar{c}_\gamma\varepsilon_0 + \bar{c}_\beta r_{\text{IPS}} = \bar{c}_\beta(4\bar{c}_\gamma \min\{c_\gamma, c_\xi/4\bar{c}_\gamma\} + c_\xi)$. For $c_1 = 4\bar{c}_\gamma\bar{c}_\beta(2 + \alpha R_{\mathbf{K}} + 2R_{\text{dyn}})$ and $c_2 = \max\{1, c_1\}^{-1} \min\{c_\gamma, c_\xi/4\bar{c}_\gamma\}$. Then if $\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le c_2$, then,

$$\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}') \le \mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')(2 + R_{\mathbf{K}}\alpha + 2R_{\text{dyn}}) \le \min\{c_\gamma, c_\xi/4\bar{c}_\gamma\}.$$

and, in particular,

$$\bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}' \mid \alpha) \le 4\bar{c}_\beta\bar{c}_\gamma((2 + R_{\mathbf{K}}\alpha + 2R_{\text{dyn}})\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')) = c_1\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')$$

Hence, unconditionally,

$$\bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}' \mid \alpha) \le c_1\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')\mathbf{I}_\infty\{\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le c_2\}$$

Thus, $\pi^\star$ satisfies $(r_{\text{IPS}}, \gamma_{\text{IPS,TVC}}, \gamma_{\text{IPS},\mathcal{S}}, \mathsf{d}_{\text{IPS}})$- restricted-IPS (Definition H.9) with $r_{\text{IPS}} = c_\xi/2 = c_4$

$$\begin{gathered}
\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') = c_1\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')\mathbf{I}_\infty\{(\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le c_2\} \\
\gamma_{\text{IPS,TVC}}(r) = \bar{c}_\beta \cdot \phi(k), \quad \gamma_{\text{IPS},\mathcal{S}}(r) = \bar{c}_\beta r
\end{gathered}$$

Using (K.5) and recalling $c_5 = e\bar{c}_\beta$, we conclude $(r_{\text{IPS}}, \gamma_{\text{IPS,TVC}}, \gamma_{\text{IPS},\mathcal{S}}, \mathsf{d}_{\text{IPS}})$- restricted-IPS (Definition H.9) with $r_{\text{IPS}} = c_\xi/2 = c_4$ and

$$\begin{gathered}
\mathsf{d}_{\mathcal{A}}(\mathsf{a}, \mathsf{a}') = c_1\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}')\mathbf{I}_\infty\{\mathsf{d}_{\max}(\mathsf{a}, \mathsf{a}') \le c_2\} \\
\gamma_{\text{IPS,TVC}}(r) = c_5 e^{-L_\beta(\tau_{\text{chunk}} - \tau_{\text{obs}})}, \quad \gamma_{\text{IPS},\mathcal{S}}(r) = c_5 r.
\end{gathered}$$

This concludes the proof. $\square$

K.1.1 Proof of Lemma K.3

Let's prove Lemma K.3(a). Let $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ be drawn from the distribution induces by $\pi^\star$, and let $\mathsf{a}'_{1:H}$ be some other sequences of actions. The primitive controllers and states under the instantiation of the composite MDP for $\mathsf{a}_{1:H}, \mathsf{a}_{1:H}$ $\mathsf{s}_{1:H+1}$ respectively be $\kappa_{1:T}, \kappa'_{1:T}$ and $\mathbf{x}_{1:T+1}$. Note that, by Lemma J.6(b), each $\mathbf{x}_t$ has the same marginals as under the expert distribution $\mathcal{D}_{\rm exp}$ and similarly so does a_h, so by the assumption of the lemma, $\mathbf{x}_t \in \mathcal{X}_0$ and $\kappa_t \in \mathcal{K}_0$ with probability one. Thus,

$$\sup_{\mathbf{x}\in\mathcal{X}_0} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\|\le\alpha} \max_{t_h \le t \le t_{h+1}-1} \|\kappa_t(\mathbf{x}_t + \delta\mathbf{x}) - \kappa'_i(\mathbf{x}_t + \delta\mathbf{x})\| \le \mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}').$$

In particular if $\varepsilon \le \varepsilon_0$ and $\mathsf{D}_\alpha(\mathsf{a}, \mathsf{a}') \le \varepsilon$, then

$$\sup_{\mathbf{x}\in\mathcal{X}_0} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\|\le\alpha} \max_{t_h \le t \le t_{h+1}-1} \|\kappa_t(\mathbf{x}_t + \delta\mathbf{x}) - \kappa'_i(\mathbf{x}_t + \delta\mathbf{x})\| \le \varepsilon \le \varepsilon_0,$$

By Lemma K.2, and the fact that $\beta(\varepsilon, i)$ is non-increasing in i, we find $\max_h \mathsf{d}_{\mathcal{S}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_t\{\|\mathbf{x}_t - \mathbf{x}'_t\|, \|\mathbf{u}_t - \mathbf{u}'_t\|\} \le \psi(\varepsilon)$, as needed.

To prove Lemma K.3(b), let $(\mathsf{s}_{1:H+1}, \tilde{\mathsf{s}}_{1:H+1}, \mathsf{a}_{1:H})$ be as in the definition of restricted IPS (Definition H.9), let $\mathsf{a}'_{1:H}$ be an alternative sequence of composite actions, and unpack these into $(\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T})$, $(\tilde{\mathbf{x}}_{1:T+1}, \tilde{\mathbf{u}}_{1:T})$, $\kappa_{1:T}$ and $\kappa_{1:T}$ as above. We let $\tilde{\mathsf{o}}_h = (\tilde{\mathbf{x}}_{t_h-\tau_{\rm obs}+1:t_h}, \tilde{\mathbf{u}}_{t_h-\tau_{\rm obs}+1:t_h}) = \phi_{\rm o} \circ \tilde{\mathsf{s}}_h$ denote the observation-chunk associated with $\tilde{\mathsf{s}}_h$. It follows from Lemma J.6(d) and the construction in (Definition H.9) that the distribution $(\tilde{\mathsf{o}}_h, \mathsf{a}_h)$ under this construction is absolutely continuous w.r.t. the distribution of $(\mathsf{o}_h, \mathsf{a}_h)$ under $\mathcal{D}_{\rm exp}$. In particular, this implies that $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$ satisfies the incremental stability condition on $\tilde{\mathbf{x}}_{t_h}$, as well as the following property: let $\hat{\mathsf{s}}_{h+1} = F_h(\tilde{\mathsf{s}}_h, \mathsf{a}_h)$, which concretely are states $(\hat{\mathbf{x}}_{t_h:t_{h+1}}, \hat{\mathbf{u}}_{t_h:t_{h+1}-1})$ corresponding to the dynamics induced by rolling out $\mathsf{a}_h = \kappa_{t_h:t_{h+1}-1}$ from $\hat{\mathbf{x}}_{t_h}$, depicted in (K.1). Then, absolute continuity of $(\tilde{\mathsf{o}}_h, \mathsf{a}_h)$ with respect to its analogues under $\mathcal{D}_{\rm exp}$ implies that $\hat{\mathbf{x}}_{t_h:t_{h+1}}$ is absolutely continuous w.r.t. the distribution of $\mathbf{x}_{t_h:t_{h+1}}$ under $\mathcal{D}_{\rm exp}$. Hence, $\hat{\mathbf{x}}_t \in \mathcal{X}_0$ for $t_h \le t \le t_{h+1}$. By a similarly argument, we also have $\kappa_t \in \mathcal{K}_0$ with probability one.Thus, we have

$$\sup_{\mathbf{x}\in\mathcal{X}_0} \sup_{\|\delta\mathbf{x}\|\le\alpha} \max_{t_h \le t \le t_{h+1}-1} \|\kappa_t(\mathbf{x} + \delta\mathbf{x}) - \kappa'_t(\mathbf{x} + \delta\mathbf{x})\| \le \mathsf{D}_\alpha(\mathsf{a}_h, \mathsf{a}'_h),$$

Hence, whenever $\mathsf{D}_\alpha(\mathsf{a}_h, \mathsf{a}'_h) \le \varepsilon$ for $\alpha = \psi(\varepsilon) + \beta(r, 0)$, then

$$\max_{1\le t\le T} \sup_{\mathbf{x}\in\mathcal{X}_0} \sup_{\|\delta\mathbf{x}\|\le\psi(\varepsilon)+\beta(r,0)} \|\kappa_t(\mathbf{x} + \delta\mathbf{x}) - \kappa'_t(\mathbf{x} + \delta\mathbf{x})\| \le \varepsilon \le \varepsilon_0, \tag{K.6}$$

then we find (using $\hat{\mathbf{x}}_t \in \mathcal{X}_0$)

$$\max_{1\le t\le T} \sup_{\|\delta\mathbf{x}\|\le\psi(\varepsilon)+\beta(r,0)} \|\kappa_t(\hat{\mathbf{x}}_t + \delta\mathbf{x}) - \kappa'_t(\hat{\mathbf{x}}_t + \delta\mathbf{x})\| \le \varepsilon \le \varepsilon_0$$

Thus when (K.6) is true for all h, Lemma K.2 again implies $\max_h \mathsf{d}_{\mathcal{S}}(\mathsf{s}_h, \mathsf{s}'_h) = \max_t\{\|\mathbf{x}_t - \mathbf{x}'_t\|, \|\mathbf{u}_t - \mathbf{u}'_t\|\} \le \psi(\varepsilon)$ (again, using $\beta(\cdot, i)$ being non-increasing in i). This concludes the proof. $\square$

K.2 Proof of Lemma K.2

We begin with the following simplifying observation, which follows from considering the definition of local t-ISS with $\delta\mathbf{u}_t \equiv 0$ at time $t = 0$:

Observation K.4. $\beta(m, u) \ge u$ for any $u \in [0, c_\xi)$.

The inequality $\|\mathbf{x}_{t_h+i} - \hat{\mathbf{x}}'_{t_h,i}\| \le \beta(r, i)$ is an imediate consequence of local-t-ISS of a_h at $\hat{\mathbf{x}}_{t_h,0}$. Note further that this means that

$$\|\mathbf{x}_{t_h+i} - \hat{\mathbf{x}}'_{t_h,i}\| \le \beta(r, i) \le \beta(r, 0) \le r \le c_\xi/2. \tag{K.7}$$

Let us prove $\|\mathbf{x}_{t_h+i} - \mathbf{x}'_{t_h+i}\| \le \beta(2\gamma(\varepsilon), i) + \gamma(\varepsilon)$. Next, define $\delta\mathbf{u}_t = \kappa'_t(\mathbf{x}'_t) - \kappa_t(\mathbf{x}'_t)$ and $\delta\mathbf{x}_t = \mathbf{x}'_t - \mathbf{x}_t$. We begin by fixing a chunk h and arguing along the lines of Pfrommer et al. [50, Proposition 3.1].In what follows, we assume either (K.2) or (K.3), restated here for convenience:

$$\max_{1\le t\le T} \sup_{\|\delta\mathbf{x}\|\le\alpha(\varepsilon)} \|\kappa_t(\mathbf{x}_t + \delta\mathbf{x}) - \kappa'_t(\mathbf{x}_t + \delta\mathbf{x})\| \le \varepsilon, \quad \alpha(\varepsilon) := 2\beta(2\gamma(\varepsilon), 0), \quad \text{or} \tag{K.8}$$

$$\max_{1\le t\le T} \sup_{\|\delta\mathbf{x}\|\le\alpha(\varepsilon)+\beta(r,0)} \|\kappa_t(\hat{\mathbf{x}}_t + \delta\mathbf{x}) - \kappa'_t(\hat{\mathbf{x}}_t + \delta\mathbf{x})\| \le \varepsilon, \tag{K.9}$$

Claim K.5. Fix $c_0 > 0$. Suppose that, at a given step h, $\|\delta\mathbf{x}_{t_h}\| \le c_0 \le c_\xi/2$, and that $\beta(c_0, 0) + \gamma(\varepsilon) \le \alpha$. Then, for all $0 \le i \le \tau_{\text{chunk}} - 1$, $\|\delta\mathbf{u}_{t_h+i}\| \le \varepsilon \le \alpha$ and

$$\forall 0 \le i \le \tau_{\text{chunk}}, \quad \|\delta\mathbf{x}_{t_h+i}\| \le \beta(c_0, i) + \gamma(\varepsilon) \le \alpha$$

Proof. We perform induction over $t \ge t_h$. Assume inductively that $\|\delta\mathbf{x}_t\| \le \beta(c_0, t - t_h) + \gamma(\varepsilon) \le \alpha$ and $\max_{1\le s\le t-1} \|\delta\mathbf{u}_s\| \le \varepsilon$; note that this base case $t = t_h$ holds as $\beta(c_0, 0) \le \alpha$ by Observation K.4 and our assumption on c_0. From the inductive hypothesis and the condition (K.8),

$$\|\delta\mathbf{u}_t\| = \|\kappa'_t(\mathbf{x}'_t) - \kappa_t(\mathbf{x}_t)\| \le \max_{t_h\le t\le t_{h+1}-1} \sup_{\|\delta\mathbf{x}\|\le\alpha} \|\kappa_t(\mathbf{x}_t + \delta\mathbf{x}) - \kappa'_t(\mathbf{x}_t + \delta\mathbf{x})\| \le \varepsilon.$$

Note that by (K.7), (K.9) also suffices for the above to hold. Hence, in either case $\max_{1\le s\le t} \|\delta\mathbf{u}_s\| \le \Delta_h$. As $\|\mathbf{x}_{t_h} - \mathbf{x}'_{t_h}\| \le c_\xi/2$, the triangle inequality and (K.7) imply $\|\mathbf{x}'_{t_h} - \hat{\mathbf{x}}_{t_h}\| \le c_\xi$. This, and the fact that $\varepsilon \le c_\gamma$, allows us to invoke our definition of incremental stabiity in Definition 3.1, implying

$$\|\delta\mathbf{x}_{t+1}\| \le \beta(c_0, t + 1 - t_h) + \gamma(\varepsilon),$$

as needed. □

To conclude, we argue inductively on h that we can take $c_0 = 2\gamma(\varepsilon)$ in the above claim. First note that $2\gamma(\varepsilon) \le c_\xi/2$ by assumption. Thus, from Observation K.4, $\gamma(\varepsilon) = \frac{1}{2}\cdot 2\gamma(\varepsilon) \le \frac{1}{2}\beta(2\gamma(\varepsilon), 0)$. Hence, for $c_0 = 2\gamma(\varepsilon)$ $\beta(c_0, 0) + \gamma(\varepsilon) \le \frac{3}{2}\beta(2\gamma(\varepsilon), 0) \le \alpha$. Moreover, by assumption $\delta\mathbf{x}_1 = 0$, the bound $\|\delta\mathbf{x}_{t_h}\| \le 2\gamma(\varepsilon)$ holds trivially for step $h = 1$. Assuming it holds for h, Claim K.5 yields

$$\forall 0 \le i \le \tau_{\text{chunk}}, \quad \|\delta\mathbf{x}_{t_h+i}\| \le \beta(c_0, i) + \gamma(\varepsilon) \le \beta(c_0, 0) + \gamma(\varepsilon) \le \alpha,$$

where the final inequality follows from the computation above. Moreover, by taking $i = \tau_{\text{chunk}}$,

$$\|\delta\mathbf{x}_{t_{h+1}}\| = \|\delta\mathbf{x}_{t_h+\tau_{\text{chunk}}}\| \le \beta(2\gamma(\varepsilon), \tau_{\text{chunk}}) + \gamma(\varepsilon) \le 2\gamma(\varepsilon),$$

where the last inequality is by the assumption of the lemma. □

K.3 Synthesized Linear Controllers are Incrementally Stabilizing

In this section, we give a sufficient condition for incremental stability of affine primitive controllers. Recal our notation of a length-K *control trajectory* is denoted $\rho = (x_{1:K+1}, u_{1:K}) \in \mathscr{P}_K = (\mathbb{R}^{d_x})^{K+1} \times (\mathbb{R}^{d_u})^K$. Given such a trajectory, the *Jacobian linearizations* are denoted

$$\mathbf{A}_k(\rho) := \frac{\partial}{\partial x} f_\eta(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{B}_k(\rho) := \frac{\partial}{\partial u} f_\eta(\mathbf{x}_k, \mathbf{u}_k)$$

for $k \in [K]$. Recalling our dynamics map $f(\cdot,\cdot)$, and step size $\eta > 0$, we say that ρ 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 now introduce a nother of *regularity* on the dynamics, which essentially enforces boundedness and smoothness.

Definition K.2 (Trajectory Regularity)**.** A control trajectory $\rho = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K})$ is $(R_{\text{dyn}}, L_{\text{dyn}}, M_{\text{dyn}})$-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 $\|\mathbf{x}'_k - \mathbf{x}_k\| \vee \|\mathbf{u}_k - \mathbf{u}'_k\| \le R_{\text{dyn}}$,[14]

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

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.

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

Definition K.3 (Jacobian Stability). Consider $R_{\mathbf{K}}, L_{\mathrm{stab}}, B_{\mathrm{stab}} \geq 1$. 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 $(R_{\mathbf{K}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian Stable if $\max_k \|\mathbf{K}_k\|_{\mathrm{op}} \leq B_{\mathrm{stab}}$, 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

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

The following proposition is proven in Appendix K.4, establishing incremental stability of affine gains.

Proposition K.6 (Incremental Stability of Affine Primitive Controller)**.** Suppose that $\bar{\boldsymbol{\rho}} = (\bar{\mathbf{x}}_{1:K+1}, \bar{\mathbf{u}}_{1:K})$ is $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$ regular, and suppose $(\bar{\boldsymbol{\rho}}, \bar{\mathbf{K}}_{1:K})$ is $(R_{\mathbf{K}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$ stable. Suppose that $\eta \leq L_{\mathrm{stab}}/2$, that $R_{\mathbf{K}} \geq 1$, define the constants

$$\begin{aligned}
c_{\xi,1} &= \frac{1}{4R_{\mathbf{K}}B_{\mathrm{stab}}} \min\left\{1, \frac{1}{4L_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}B_{\mathrm{stab}}}\right\} \\
c_{\xi,2} &= \min\left\{\frac{1}{96B_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2}, \frac{R_{\mathrm{dyn}}}{32R_{\mathbf{K}}}\right\} \\
c_{\xi} &= \min\{c_{\xi,1}, c_{\xi,2}/2\} \\
c_{\gamma} &= \min\left\{\frac{1}{48B_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2}, \frac{R_{\mathrm{dyn}}}{16L_{\mathrm{stab}}R_{\mathbf{K}}}\right\} \\
\bar{c}_{\beta} &:= 16B_{\mathrm{stab}} \\
\bar{c}_{\gamma} &:= 8L_{\mathrm{stab}}B_{\mathrm{stab}}L_{\mathrm{dyn}}
\end{aligned}$$

and set

$$\beta(u,k) = \bar{c}_{\beta}\left(1 - \frac{\eta}{L_{\mathrm{stab}}}\right)^{k-1} \cdot u, \quad \gamma(u) := \bar{c}_{\gamma} \cdot u$$

Then, the controllers $\kappa_k(\mathbf{x}) = \bar{\mathbf{K}}_k(\mathbf{x} - \bar{\mathbf{x}}_k) + \bar{\mathbf{u}}_k$ are incrementally stabilizing in the sense of Definition 3.1 with moduli $\gamma(\cdot)$ and $\beta(\cdot,\cdot)$ and constants c_{ξ}, c_{γ} as above.

K.4 Proof of Proposition K.6 (incremental stability of affine gains)

We require the following lemma, proven in the section below.

Lemma K.7 (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 $(R_{\mathbf{K}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable. Assume that $R_{\mathbf{K}} \geq 1$, $L_{\mathrm{stab}} \geq 2\eta$. 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

$$\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \leq c_{\xi,1} := \frac{1}{4R_{\mathbf{K}}B_{\mathrm{stab}}} \min\left\{1, \frac{1}{4L_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}B_{\mathrm{stab}}}\right\},$$

then

- $\|\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}\| \leq 2B_{\mathrm{stab}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\|\beta_{\mathrm{stab}}^k$.
- $(\boldsymbol{\rho}, \mathbf{K}_{1:K})$ is $(R_{\mathbf{K}}, 2B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stable.
- $\|\mathbf{B}_k(\boldsymbol{\rho})\| \leq L_{\mathrm{dyn}}$, and in addition, the trajectory $\boldsymbol{\rho}$ is $(R_{\mathrm{dyn}}/2, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular.

This lemma is proven in Appendix K.5 just below. Now, set $\beta_{\mathrm{stab}} = (1 - \eta/L_{\mathrm{stab}})$, and define $\delta\mathbf{x}_k = \mathbf{x}'_k - \mathbf{x}_k$. Let $\mathbf{A}_t = \frac{\partial}{\partial x} f_{\eta}(x,u)\big|_{(x,u)=(\mathbf{x}_k,\mathbf{u}_k)}$

$$\mathbf{x}'_{t+1} = \mathbf{x}_k + \eta f_{\eta}(\mathbf{x}'_k, \kappa_k(\mathbf{x}'_k) + \delta\mathbf{u}_k), \quad \mathbf{x}_{t+1} = \mathbf{x}_k + \eta f_{\eta}(\mathbf{x}_k, \kappa_k(\mathbf{x}_k))$$

This means that

$$\delta\mathbf{x}_{t+1} = \underbrace{(\mathbf{I} + \eta(\mathbf{A}_k + \mathbf{B}_k\bar{\mathbf{K}}_k))}_{=\mathbf{A}_{\mathrm{cl},k}}\delta\mathbf{x}_k + \eta\mathbf{B}_k\delta\mathbf{u}_k + \eta\mathrm{rem}_k,$$

where $\mathrm{rem}_k = f_\eta(\mathbf{x}_k', \kappa_k(\mathbf{x}_k') + \delta\mathbf{u}_k) - f_\eta(\mathbf{x}_k, \kappa_k(\mathbf{x}_k)) - (\mathbf{A}_k + \mathbf{B}_k\bar{\mathbf{K}}_k)\delta\mathbf{x}_k - \mathbf{B}_k\delta\mathbf{u}_k$. Defining $\mathbf{\Phi}_{\mathrm{cl},k,j} := \mathbf{A}_{\mathrm{cl},t-1}\mathbf{A}_{\mathrm{cl},t-2}\dots\mathbf{A}_{\mathrm{cl},s}$ and unfolding the recursion,

$$\delta\mathbf{x}_{t+1} = \eta\sum_{s=1}^{t}\mathbf{\Phi}_{\mathrm{cl},k+1,j+1}(\mathbf{B}_s\delta\mathbf{u}_k + \mathrm{rem}_k) + \mathbf{\Phi}_{\mathrm{cl},k+1,1}\delta\mathbf{x}_1$$

Define $\varepsilon_k = \|\delta\mathbf{x}_k\|$ and $\varepsilon_{\mathbf{u}} := \max_{1\le t\le T}\|\delta\mathbf{u}_k\|$. Then, we have

$$\begin{aligned}
\varepsilon_{t+1} &\le \eta\sum_{j=1}^{k}\|\mathbf{\Phi}_{\mathrm{cl},k+1,j+1}\|(L_{\mathrm{dyn}}\varepsilon_{\mathbf{u}} + \|\mathrm{rem}_k\|) + \|\mathbf{\Phi}_{\mathrm{cl},k+1,1}\|\varepsilon_1 \\
&\overset{(i)}{\le} 2B_{\mathrm{stab}}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}(L_{\mathrm{dyn}}\varepsilon_{\mathbf{u}} + \|\mathrm{rem}_k\|) + \beta_{\mathrm{stab}}^{k}\varepsilon_1\right) \\
&\overset{(ii)}{\le} 2B_{\mathrm{stab}}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}(L_{\mathrm{dyn}}\varepsilon_{\mathbf{u}} + M_{\mathrm{dyn}}((1+2R_{\mathbf{K}}^2)\varepsilon_k^2 + 2\varepsilon_{\mathbf{u}}^2)) + \beta_{\mathrm{stab}}^{t}\varepsilon_1\right) \\
&\overset{(iii)}{\le} 2B_{\mathrm{stab}}\left(\eta\sum_{j=1}^{k}\beta_{\mathrm{stab}}^{k-j}(2L_{\mathrm{dyn}}\varepsilon_{\mathbf{u}} + M_{\mathrm{dyn}}(1+2R_{\mathbf{K}}^2)\varepsilon_k^2) + \beta_{\mathrm{stab}}^{t}\varepsilon_1\right)
\end{aligned}$$

where we in (i) $\|\mathbf{\Phi}_{\mathrm{cl},k,j}\| \le 2B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{t-s}$, and (ii) follows by Claim K.8, stated and proven below, and the following inductive hypothesis

$$\max_{1\le j\le k}\varepsilon_k \le C' = \frac{R_{\mathrm{dyn}}}{4R_{\mathbf{K}}}, \qquad \text{(Inductive Hypothesis)}$$

and (ii) uses the assumption $\varepsilon_{\mathbf{u}} \le \frac{L_{\mathrm{dyn}}}{2M_{\mathrm{dyn}}}$. Setting $\Delta_1 = 2B_{\mathrm{stab}}\varepsilon_1, \Delta_2 = 4B_{\mathrm{stab}}L_{\mathrm{dyn}}\varepsilon_{\mathbf{u}}$ and $C = 2B_{\mathrm{stab}}M_{\mathrm{dyn}}(1+2R_{\mathbf{K}}^2) \le 6B_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2$, Lemma K.17 implies

$$\varepsilon_k \le 4\Delta_1\beta_{\mathrm{stab}}^{k-1} + 2L_{\mathrm{stab}}\Delta_2 = \underbrace{8B_{\mathrm{stab}}\varepsilon_1\beta_{\mathrm{stab}}^{k-1}}_{\beta(\varepsilon_1,k)} + \underbrace{8L_{\mathrm{stab}}B_{\mathrm{stab}}L_{\mathrm{dyn}}\varepsilon_{\mathbf{u}}}_{\gamma(\varepsilon_{\mathbf{u}})}$$

provided that that $\Delta_2 \le \min\left\{\frac{1}{8CL}, \frac{C'}{4L}\right\}$ and $\Delta_1 \le \min\left\{\frac{1}{16CL}, \frac{C'}{8}\right\}$ for $L = L_{\mathrm{stab}}$. Subsituting in relevant quantities and keeping the shorthand $L = L_{\mathrm{stab}}$, it suffices that

$$\min\left\{\frac{1}{16CL}, \frac{C'}{8}\right\} \ge \underbrace{\min\left\{\frac{1}{96B_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2}, \frac{R_{\mathrm{dyn}}}{32R_{\mathbf{K}}}\right\}}_{=c_{\xi,2}/2} \ge \varepsilon_1$$

$$\min\left\{\frac{1}{8CL}, \frac{C'}{4L}\right\} \ge \underbrace{\min\left\{\frac{1}{48B_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2}, \frac{R_{\mathrm{dyn}}}{16L_{\mathrm{stab}}R_{\mathbf{K}}}\right\}}_{=c_\gamma} \ge \varepsilon_{\mathbf{u}}.$$

□

Claim K.8. Suppose that $\varepsilon_{\mathbf{u}} \le \frac{R_{\mathrm{dyn}}}{4}$ and $\varepsilon_k \le \frac{R_{\mathrm{dyn}}}{4R_{\mathbf{K}}}$. Then, $\|\mathrm{rem}_k\| \le M_{\mathrm{dyn}}((1+2R_{\mathbf{K}}^2)\varepsilon_k^2 + 2\varepsilon_{\mathbf{u}}^2)$.

Proof. Define $\mathbf{u}_k = \kappa_k(\mathbf{x}_k)$ and $\delta\mathbf{u}_k' = \kappa_k(\mathbf{x}_k')\delta\mathbf{u}_k - \mathbf{u}_k$. We have that $\delta\mathbf{u}_k' = \delta\mathbf{u}_k + \kappa_k(\mathbf{x}_k') - \kappa(\mathbf{x}_k) = \delta\mathbf{u}_k + \bar{\mathbf{K}}_k(\mathbf{x}_k' - \mathbf{x}_k)$. We bound $\|\delta\mathbf{u}_k'\| \le \|\delta\mathbf{u}_k\| + \|\bar{\mathbf{K}}_k(\mathbf{x}_k' - \mathbf{x}_k)\| \le \|\delta\mathbf{u}_k\| + R_{\mathbf{K}}\|\delta\mathbf{x}_k\|$, where we recall $\|\bar{\mathbf{K}}_k\| \le R_{\mathbf{K}}$ and $\delta\mathbf{x}_k = \mathbf{x}_k' - \mathbf{x}$. By assumption and definition, $\|\delta\mathbf{u}_k\| \le \varepsilon_{\mathbf{u}}$ and by definition of ε_k we conclude that

$$\|\delta\mathbf{x}_k\| \le \varepsilon_k, \quad \|\delta\mathbf{u}_k'\| \le (1+R_{\mathbf{K}})\varepsilon_k + \varepsilon_{\mathbf{u}} \qquad \text{(K.10)}$$

Consider a curve $\mathbf{x}_k(s) = \mathbf{x}_k + s\delta\mathbf{x}_k$ and $\mathbf{u}_k(s) = \delta\mathbf{u}'_k + \mathbf{u}_k$. With these definition

$$
\begin{aligned}
\mathrm{rem}_k &= f_\eta(\mathbf{x}_k(1), \mathbf{u}_k(1)) - f_\eta(\mathbf{x}_k(0), \mathbf{u}_k(0)) - (\mathbf{A}_k + \mathbf{B}_k\bar{\mathbf{K}}_k)\delta\mathbf{x}_k - \mathbf{B}_k\delta\mathbf{u}_k \\
&= f_\eta(\mathbf{x}_k(1), \mathbf{u}_k(1)) - f_\eta(\mathbf{x}_k(0), \mathbf{u}_k(0)) - \mathbf{A}_k\mathbf{x}_k + \mathbf{B}_k\delta\mathbf{u}'_k \\
&= \underbrace{\frac{\partial}{\partial s}(f_\eta(\mathbf{x}_k(s), \mathbf{u}_k(s))) - \mathbf{A}_k\delta\mathbf{x}_k + \mathbf{B}_k\delta\mathbf{u}'_k}_{=0} \\
&\quad + \int_0^s (1-s)^2 \frac{\partial^2}{\partial s^2}(f_\eta(\mathbf{x}_k(s), \mathbf{u}_k(s)))^\top\big|_{s=0}(\delta\mathbf{x}_k, \delta\mathbf{u}'_k)\mathrm{d}s
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|\mathrm{rem}_k\| &\le \|\frac{1}{2}\sup_{s\in[0,1]}\frac{\partial^2}{\partial s^2}(f_\eta(\mathbf{x}_k(s), \mathbf{u}_k(s)))\| \le M_{\mathrm{dyn}}\sup_s\|\frac{\partial}{\partial s}(\mathbf{x}_k(s), \mathbf{u}_k(s)\|^2 \\
&\overset{(i)}{\le} M_{\mathrm{dyn}}(\|\delta\mathbf{x}_k\|^2 + \|\delta\mathbf{u}'_k\|^2) \\
&\le M_{\mathrm{dyn}}((1+2R_{\mathbf{K}}^2)\varepsilon_k^2 + 2\varepsilon_{\mathbf{u}}^2), && \text{((K.10) and Am-GM)}
\end{aligned}
$$

To justify inequality (i), we observe that $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K})$ is $(R_{\mathrm{dyn}}/2, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$ regular. Note tat $\sup_s\|\mathbf{x}_k(s) - \mathbf{x}_k\| = \|\delta\mathbf{x}_t\|$ and $\sup_s\|\mathbf{u}_k(s) - \mathbf{u}_k\| = \|\delta\mathbf{u}'_t\|$. Hence, by the definition of trajectory regularity (Definition K.2), (i) holds as long as we check that $\|\delta\mathbf{x}_k\| \vee \|\delta\mathbf{u}'_k\| \le R_{\mathrm{dyn}}/4$. As $\|\delta\mathbf{x}_k\| \vee \|\delta\mathbf{u}_k\| \le \max\{R_{\mathbf{K}}\varepsilon_k + \varepsilon_{\mathbf{u}}, \varepsilon_k\}$ and as we take $R_{\mathbf{K}} \ge 1$, it suffices that $\varepsilon_{\mathbf{u}} \le \frac{R_{\mathrm{dyn}}}{4}$ and $\varepsilon_k \le \frac{R_{\mathrm{dyn}}}{4R_{\mathbf{K}}}$, which is ensured by the claim. □

K.5 Proof of Lemma K.7 (state perturbation)

Define $\bar{\Delta}_{\mathbf{x},k} = \mathbf{x}_k - \bar{\mathbf{x}}_k$. Then

$$
\begin{aligned}
\bar{\Delta}_{\mathbf{x},k+1} &= \bar{\Delta}_{\mathbf{x},k} + \eta\left(f_\eta(\mathbf{x}_k, \bar{\mathbf{u}}_k + \bar{\mathbf{K}}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k) - f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\right) \\
&= \bar{\Delta}_{\mathbf{x},k} + \eta(\mathbf{A}_k(\bar{\boldsymbol{\rho}}) + \mathbf{B}_k(\bar{\boldsymbol{\rho}})\mathbf{K}_k)\Delta_{\mathbf{x},k} + \mathrm{rem}_k, \qquad \text{(K.11)}
\end{aligned}
$$

where

$$
\mathrm{rem}_k = f_\eta(\mathbf{x}_k, \bar{\mathbf{u}}_k + \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k)) - f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k) - (\mathbf{A}_k(\bar{\boldsymbol{\rho}}) + \mathbf{B}_k(\bar{\boldsymbol{\rho}})\mathbf{K}_k)\bar{\Delta}_{\mathbf{x},k}.
$$

Claim K.9. Take $R_{\mathbf{K}} \ge 1$, and suppose $\|\bar{\Delta}_{\mathbf{x},k}\| \le R_{\mathrm{dyn}}/2R_{\mathbf{K}}$. Then,

$$
\|\bar{\mathbf{x}}_k - \mathbf{x}_k\| \vee \|\bar{\mathbf{u}}_k - \mathbf{u}_k\| \le R_{\mathrm{dyn}}/2, \qquad \text{(K.12)}
$$

and $\|\mathrm{rem}_k\| \le M_{\mathrm{dyn}}R_{\mathbf{K}}^2\|\bar{\Delta}_{\mathbf{x},k}\|^2$.

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 $\|\mathbf{u}_k - \bar{\mathbf{u}}_k\| \vee \|\mathbf{x}_k \vee \bar{\mathbf{x}}_k\| \le R_{\mathrm{dyn}}/2$. From Taylor's theorem and the fact that $\bar{\boldsymbol{\rho}}$ is $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regular imply that

$$
\begin{aligned}
\|f_\eta(\mathbf{x}_k, \mathbf{u}_k) - f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| &\le \frac{1}{2}M_{\mathrm{dyn}}(\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^2 + \|\mathbf{u}_k - \bar{\mathbf{u}}_k\|) \\
&\le \frac{1}{2}(1 + R_{\mathbf{K}}^2)M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|^2 \le R_{\mathbf{K}}^2 M_{\mathrm{dyn}}\|\bar{\Delta}_{\mathbf{x},k}\|^2,
\end{aligned}
$$

where again use $R_{\mathbf{K}} \ge 1$ above. □

Solving the recursion from (K.11), 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 $M := \frac{\eta}{\beta_{\mathrm{stab}}^{-1} - 1} = L_{\mathrm{stab}}$. By assumption, $\|\mathbf{\Phi}_{\mathrm{cl},k,j}\| \le B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k-j}$, so using Claim K.9 implies that, if $\max_{j\in[k]} \|\bar{\Delta}_{\mathbf{x},j}\| \le R_{\mathrm{dyn}}/2R_{\mathbf{K}}$ for all $j \in [k]$,

$$\|\bar{\Delta}_{\mathbf{x},k+1}\| \le \eta \sum_{j=1}^{k} B_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathbf{K}}^2 \beta_{\mathrm{stab}}^{k-j} \|\bar{\Delta}_{\mathbf{x},j}\|^2 + B_{\mathrm{stab}}\beta_{\mathrm{stab}}^{k}\|\bar{\Delta}_{\mathbf{x},1}\|.$$

Appling Lemma K.15 with $\alpha = 0$, $C_1 = B_{\mathrm{stab}} M_{\mathrm{dyn}} R_{\mathbf{K}}^2$, and $C_2 = B_{\mathrm{stab}} \ge 1$ and $M = L_{\mathrm{stab}}$ (noting $\beta_{\mathrm{stab}} \ge 1/2$), it holds that for $\|\bar{\Delta}_{\mathbf{x},1}\| = \varepsilon_1 \le 1/4MC_1C_3 = 1/4L_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2 B_{\mathrm{stab}}^2$,

$$\|\bar{\Delta}_{\mathbf{x},k+1}\| \le 2B_{\mathrm{stab}}\|\bar{\Delta}_{\mathbf{x},1}\|(1 - \frac{\eta}{L_{\mathrm{stab}}})^k.$$

To ensure the inductive hypothesis that $\max_{j\in[k]} \|\bar{\Delta}_{\mathbf{x},j}\| \le R_{\mathrm{dyn}} R_{\mathbf{K}}$, it suffices to ensure that $2B_{\mathrm{stab}}\|\bar{\Delta}_{\mathbf{x},1} \le R_{\mathrm{dyn}}/2R_{\mathbf{K}}$, which is assumed by the lemma. Thus, we have shown that, if

$$\|\bar{\Delta}_{\mathbf{x},1}\| \le \min\left\{\frac{R_{\mathrm{dyn}}}{2R_{\mathbf{K}}B_{\mathrm{stab}}}, \frac{1}{8L_{\mathrm{stab}}M_{\mathrm{dyn}}R_{\mathbf{K}}^2 B_{\mathrm{stab}}^2}\right\},$$

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 (K.12),

$$\begin{aligned}
&\|\mathbf{A}_k(\boldsymbol{\rho}) + \mathbf{B}_k(\boldsymbol{\rho})\mathbf{K}_k - \mathbf{A}_k(\bar{\boldsymbol{\rho}}) + \mathbf{B}_k(\bar{\boldsymbol{\rho}})\mathbf{K}_k\| \\
&= \left\|\begin{bmatrix}\mathbf{A}_k(\boldsymbol{\rho}) - \mathbf{A}_k(\bar{\boldsymbol{\rho}}) & \hat{\mathbf{B}}_k(\boldsymbol{\rho}) - \mathbf{B}_k(\bar{\boldsymbol{\rho}})\end{bmatrix}\begin{bmatrix}\mathbf{I}\\ \mathbf{K}_k\end{bmatrix}\right\| \\
&= \left\|(\nabla f_\eta(\hat{\mathbf{x}}_k, \mathbf{u}_k) - \nabla f_\eta(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k))\begin{bmatrix}\mathbf{I}\\ \mathbf{K}_k\end{bmatrix}\right\| \\
&\le M_{\mathrm{dyn}}\,\|(\mathbf{x}_k - \bar{\mathbf{x}}_k, \mathbf{K}_k(\mathbf{x}_k - \bar{\mathbf{x}}_k)\|\,\left\|\begin{bmatrix}\mathbf{I}\\ \mathbf{K}_k\end{bmatrix}\right\| \\
&= M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|\left\|\begin{bmatrix}\mathbf{I}\\ \mathbf{K}_k\end{bmatrix}\right\|^2 \le M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|(1 + \|\mathbf{K}_k\|_{\mathrm{op}}^2) \\
&= M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|\left\|\begin{bmatrix}\mathbf{I}\\ \mathbf{K}_k\end{bmatrix}\right\|^2 \le M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\|(1 + \|\mathbf{K}_k\|_{\mathrm{op}}^2) \\
&\le 2R_{\mathbf{K}}^2 M_{\mathrm{dyn}}\|\mathbf{x}_k - \bar{\mathbf{x}}_k\| \\
&\le 4B_{\mathrm{stab}}R_{\mathbf{K}}^2 M_{\mathrm{dyn}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\|\beta_{\mathrm{stab}}^{k-1}, \quad \beta_{\mathrm{stab}} = (1 - \frac{\eta}{L_{\mathrm{stab}}}).
\end{aligned}$$

Invoking Lemma K.18 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 $4B_{\mathrm{stab}}R_{\mathbf{K}}^2 M_{\mathrm{dyn}}\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le 1/4B_{\mathrm{stab}}L_{\mathrm{stab}}$, which requires $\|\mathbf{x}_1 - \bar{\mathbf{x}}_1\| \le 1/16B_{\mathrm{stab}}^2 R_{\mathbf{K}}^2 L_{\mathrm{stab}} M_{\mathrm{dyn}}$.

The last part of the lemma uses $(R_{\mathrm{dyn}}, L_{\mathrm{dyn}}, M_{\mathrm{dyn}})$-regularity of $\bar{\boldsymbol{\rho}}$ and (K.12).

K.6 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 Jacobian stability, Definition K.3, 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 [51, 76].

Definition K.4 (Stabilizability)**.** A control trajectory $\boldsymbol{\rho} = (\mathbf{x}_{1:K+1}, \mathbf{u}_{1:K}) \in \mathscr{P}_K$ is $L_{\mathcal{S}/\mathcal{O}}$-Jacobian-Stabilizable if $\max_k \mathcal{V}_k(\boldsymbol{\rho}) \le L_{\mathcal{S}/\mathcal{O}}$, where for $k \in [K+1]$, $\mathcal{V}_k(\boldsymbol{\rho})$ is defined by

$$\begin{aligned}
\mathcal{V}_k(\boldsymbol{\rho}) := \sup_{\xi:\|\xi\le 1}&\left(\inf_{\tilde{\mathbf{u}}_{1:s}} \|\tilde{\mathbf{x}}_{K+1}\|^2 + \eta\sum_{j=k}^{K}\|\tilde{\mathbf{x}}_j\|^2 + \|\tilde{\mathbf{u}}_j\|^2\right) \\
&\text{s.t. } \tilde{\mathbf{x}}_k = \xi, \quad \tilde{\mathbf{x}}_{j+1} = \tilde{\mathbf{x}}_j + \eta\,(\mathbf{A}_j(\boldsymbol{\rho})\tilde{\mathbf{x}}_j + \mathbf{B}_j(\boldsymbol{\rho})\tilde{\mathbf{u}}_j),
\end{aligned}$$

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 [51]).

Definition K.5 (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

$$\mathbf{P}_{K+1}^{\mathrm{ric}}(\boldsymbol{\rho}) = \mathbf{I}, \quad \mathbf{P}_k^{\mathrm{ric}}(\boldsymbol{\rho}) = (\mathbf{I} + \eta \mathbf{A}_{\mathrm{cl},k}(\boldsymbol{\rho}))^\top \mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})(\mathbf{I} + \eta \mathbf{A}_{\mathrm{cl},k}(\boldsymbol{\rho})) + \eta(\mathbf{I} + \mathbf{K}_k(\boldsymbol{\rho})\mathbf{K}_k(\boldsymbol{\rho})^\top)$$
$$\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho}) = (\mathbf{I} + \eta \mathbf{B}_k^\top \mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\mathbf{B}_k)^{-1}(\mathbf{B}_k^\top \mathbf{P}_{k+1}(\boldsymbol{\rho}))(\mathbf{I} + \eta \mathbf{A}_k)$$
$$\mathbf{A}_{\mathrm{cl},k}^{\mathrm{ric}}(\boldsymbol{\rho}) = \mathbf{A}_k + \mathbf{B}_k \mathbf{K}_k(\boldsymbol{\rho}).$$

The main result of this section is that the parameters $(R_{\mathbf{K}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$ in Definition K.3 can be bounded in terms of L_{dyn} in Definition K.2, and the bound $L_{\mathcal{S}/\mathcal{O}}$ defined above.

Proposition K.10 (Instantiating the Lyapunov Lemma)**.** Let $L_{\mathrm{dyn}}, L_{\mathcal{S}/\mathcal{O}} \geq 1$, 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 $L_{\mathcal{S}/\mathcal{O}}$-Jacobian Stabilizable. Suppose further that $\eta \leq 1/5L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}$. Then, $(\boldsymbol{\rho}, \mathbf{K}_{1:K}^{\mathrm{ric}})$-is $(R_{\mathbf{K}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-Jacobian Stable, where

$$R_{\mathbf{K}} = \frac{4}{3} L_{\mathcal{S}/\mathcal{O}} L_{\mathrm{dyn}}, \quad B_{\mathrm{stab}} = \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{S}/\mathcal{O}}, \quad L_{\mathrm{stab}} = 2L_{\mathcal{S}/\mathcal{O}}$$

Proposition K.10 is proven in Appendix K.6.1 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.

K.6.1 Proof of Proposition K.10 (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 $L_{\mathrm{dyn}}, L_{\mathcal{S}/\mathcal{O}} \geq 1$. We begin by translating our stabilizability assumption (Definition K.4) into the the $\mathbf{P}$-matrices in Definition K.5. The following statement recalls Lemma F.1 in [51], an instantiation of well-known solutions to linear-quadratic dynamic programming (e.g. [7]).

Lemma K.11 (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 $L_{\mathcal{S}/\mathcal{O}}$-stabilizable,

$$\max_{k \in [K+1]} \|\mathbf{P}_k(\boldsymbol{\Theta})\|_{\mathrm{op}} \leq L_{\mathcal{S}/\mathcal{O}}.$$

Lemma K.12 (Lyapunov Lemma, Lemma F.10 in [51])**.** 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 $\max_k \|\mathbf{I} - \mathbf{X}_k\|_{\mathrm{op}} \leq \kappa\eta$ for some $\kappa \leq 1/2\eta$,

$$\|\boldsymbol{\Phi}_{j,k}\|^2 \leq \max\{1, 2\kappa\} \max_{k \in [K+1]} \|\boldsymbol{\Lambda}_k\|(1 - \eta\alpha)^{j-k}, \quad \alpha := \frac{1}{\max_{k \in [K+1]} \|\boldsymbol{\Lambda}_{1:K+1}\|}.$$

Claim K.13. 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 $\max_{k \in [K]} \max\{\|\mathbf{A}_k\|, \|\mathbf{B}_k\|\} \leq L_{\mathrm{dyn}}$.

Proof. For any $k \in [K]$,

$$\max\{\|\mathbf{A}_k\|, \|\mathbf{B}_k\|\} = \max\left\{\left\|\frac{\partial}{\partial x} f(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\right\|, \left\|\frac{\partial}{\partial u} f(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\right\|\right\} \leq \|\nabla f(\bar{\mathbf{x}}_k, \bar{\mathbf{u}}_k)\| \leq L_{\mathrm{dyn}},$$

where the last inequality follows by regularity. □

Claim K.14. 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 $L_{\mathcal{S}/\mathcal{O}}$-stabilizable and $(0, L_{\mathrm{dyn}}, \infty)$-regular, and if $\eta \le 1/3L_{\mathrm{dyn}}$,

$$\|\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho})\| \le \frac{4}{3} L_{\mathcal{S}/\mathcal{O}} L_{\mathrm{dyn}}$$

Proof. We bound

$$\begin{aligned}
\|\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho})\| &\le \|\mathbf{B}_k\|\|\mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\|(1 + \eta\|\mathbf{A}_k\|) \\
&\le L_{\mathrm{dyn}}(1 + \eta L_{\mathrm{dyn}})\|\mathbf{P}_{k+1}^{\mathrm{ric}}(\boldsymbol{\rho})\| && \text{(Claim K.13)} \\
&\le L_{\mathcal{S}/\mathcal{O}} L_{\mathrm{dyn}}(1 + \eta L_{\mathrm{dyn}}) && \text{(Lemma K.11, } L_{\mathcal{S}/\mathcal{O}} \ge 1) \\
&\le \frac{4}{3} L_{\mathcal{S}/\mathcal{O}} L_{\mathrm{dyn}} && (\eta \le 1/3L_{\mathrm{dyn}}.)
\end{aligned}$$

□

Proof of Proposition K.10. We want to show that $\mathbf{K}_{1:K}^{\mathrm{ric}}(\boldsymbol{\rho})$ is $(R_{\mathbf{K}}, B_{\mathrm{stab}}, L_{\mathrm{stab}})$-stabilizing.Claim K.14 has already established that $\max_{k \in [K]} \|\mathbf{K}_k^{\mathrm{ric}}(\boldsymbol{\rho})\| \le R_{\mathbf{K}} = \frac{4}{3} L_{\mathcal{S}/\mathcal{O}} L_{\mathrm{dyn}}$.

To prove the other conditions, we apply Lemma K.12 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 K.5, let have that the term $\boldsymbol{\Lambda}_k$ in Lemma K.12 is precise equal to $\mathbf{P}_k(\boldsymbol{\Theta})$. From Lemma K.11,

$$\max_{k \in [K+1]} \|\mathbf{P}_k(\boldsymbol{\Theta})\|_{\mathrm{op}} = \max_{k \in [K+1]} \mathcal{V}_k(\boldsymbol{\rho}) \le L_{\mathcal{S}/\mathcal{O}}.$$

This implies that if $\max_k \|\mathbf{X}_k - \mathbf{I}\| \le \kappa\eta \le 1/2$, we have

$$\|\boldsymbol{\Phi}_{\mathrm{cl},j,k}(\boldsymbol{\Theta})\|^2 = \|(\mathbf{X}_j \cdot \mathbf{X}_{j-1} \cdot \ldots \mathbf{X}_k)\| \le \max\{1, 2\kappa\} L_{\mathcal{S}/\mathcal{O}} \left(1 - \frac{\eta}{L_{\mathcal{S}/\mathcal{O}}}\right)^{j-k}.$$

It suffices to find an appropriate upper bound κ. We have

$$\begin{aligned}
\|\mathbf{X}_k - \mathbf{I}\| = \|\eta\mathbf{A}_{\mathrm{cl},k}(\boldsymbol{\Theta})\| &\le \eta(\|\mathbf{A}_k\| + \|\mathbf{B}_k\|\|\mathbf{K}_k(\boldsymbol{\Theta})\|) \\
&\le \eta L_{\mathrm{dyn}}(1 + \|\mathbf{K}_k(\boldsymbol{\Theta})\|) \\
&\le \eta L_{\mathrm{dyn}}(1 + \frac{4}{3} L_{\mathrm{dyn}} L_{\mathcal{S}/\mathcal{O}}) && \text{(Claim K.14)} \\
&\le \frac{7}{3}\eta L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}} && (L_{\mathcal{S}/\mathcal{O}}, L_{\mathrm{dyn}} \ge 1)
\end{aligned}$$

Setting $\kappa = \frac{7}{3} L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}$., we have that as $\eta \le \frac{1}{5L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}} \le \min\{\frac{3}{14L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}}, \frac{1}{3L_{\mathrm{dyn}}}\}$ (recall $L_{\mathrm{dyn}}, L_{\mathcal{S}/\mathcal{O}} \ge 1$), we can bound

$$\max\{1, 2\kappa\} \le \max\left\{1, \frac{14}{3} L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}\right\} \le \max\left\{1, 5L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}\right\} = 5L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}^2,$$

where again recall $L_{\mathcal{S}/\mathcal{O}}, L_{\mathrm{dyn}} \ge 1$. In sum, for $\eta \le \frac{1}{5L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}}$, we have

$$\|\boldsymbol{\Phi}_{\mathrm{cl},j,k}\|^2 \le 5L_{\mathrm{dyn}}^2 L_{\mathcal{S}/\mathcal{O}}^2 \left(1 - \frac{\eta}{L_{\mathcal{S}/\mathcal{O}}}\right)^{j-k}.$$

Hence, using the elementary inequality $\sqrt{1-a} \le (1 - a/2)$,

$$\|\boldsymbol{\Phi}_{\mathrm{cl},j,k}\| \le \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{S}/\mathcal{O}} \left(1 - \frac{\eta}{L_{\mathcal{S}/\mathcal{O}}}\right)^{(j-k)/2} \le \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{S}/\mathcal{O}} \left(1 - \frac{\eta}{2L_{\mathcal{S}/\mathcal{O}}}\right)^{j-k},$$

which shows that we can select $B_{\mathrm{stab}} = \sqrt{5} L_{\mathrm{dyn}} L_{\mathcal{S}/\mathcal{O}}$ and $L_{\mathrm{stab}} = 2L_{\mathcal{S}/\mathcal{O}}$. □

K.7 Solutions to recursions

This section contains the solutions to various recursions.

Lemma K.15 (First Key Recursion)**.** Let $C_1 > 0, C_2 \geq 1/2$, $\beta_{\text{stab}} \in (0,1)$, and suppose $\varepsilon_1, \varepsilon_2, \dots$ is a sequence satisfying $\varepsilon_1 \leq \bar{\varepsilon}_1$, and

$$\varepsilon_{k+1} \leq C_2 \beta_{\text{stab}}^k \bar{\varepsilon}_1 + C_1 \eta \sum_{j=1}^{k} \beta_{\text{stab}}^{k-j} \varepsilon_j^2$$

Then, as long as $C_1 \leq \beta(1-\beta)/2\eta$, it holds that $\varepsilon_k \leq 2C_2 \beta_{\text{stab}}^{k-1} \bar{\varepsilon}_1$ for all k.

Proof. Consider the sequence $\nu_k = 2C_2 \beta_{\text{stab}}^{k-1} \bar{\varepsilon}_1$. Since $C_2 \geq 1/2$, we have $\nu_1 \geq \bar{\varepsilon}_1 \geq \varepsilon_1$. Moreover,

$$\begin{aligned} C_2 \beta_{\text{stab}}^k \bar{\varepsilon}_1 + C_1 \sum_{j=1}^{k} \beta_{\text{stab}}^{k-j} \nu_j &= C_2 \beta_{\text{stab}}^k \bar{\varepsilon}_1 + 2C_1 C_2 \sum_{j=1}^{k} \beta_{\text{stab}}^{k+j-2} \bar{\varepsilon}_1 \\ &= C_2 \beta_{\text{stab}}^k \bar{\varepsilon}_1 \left(1 + \frac{2C_1}{\beta} \sum_{j=0}^{k-1} \beta_{\text{stab}}^j \right) \\ &\leq C_2 \beta_{\text{stab}}^k \bar{\varepsilon}_1 \left(1 + \frac{2C_1 \eta}{\beta(1-\beta)}\right) \end{aligned}$$

Hence, for $C_1 \leq \beta(1-\beta)/2\eta$, we have $C_2 \beta_{\text{stab}}^k \bar{\varepsilon}_1 + C_1 \sum_{j=1}^{k} \beta_{\text{stab}}^{k-j} \nu_j \leq 2C_2 \bar{\varepsilon}_1 \beta_{\text{stab}}^k \leq \nu_{k+1}$. This shows that the (ν_k) sequence dominates the (ε_k) sequence, as needed. □

Lemma K.16 (Second Key Recursion)**.** Let $c, \Delta, \eta > 0$, $\beta_{\text{stab}} \in (0,1)$ and let $\varepsilon_1, \varepsilon_2, \dots$ satisfy $\varepsilon_1 \leq c$ and

$$\varepsilon_{k+1} \leq c\beta_{\text{stab}}^k + c\eta\Delta\beta_{\text{stab}}^{k-1} \sum_{j=1}^{k} \varepsilon_j.$$

Then, if $\Delta \leq \frac{\beta(1-\beta)}{2c\eta}$, $\varepsilon_{k+1} \leq 2c\beta_{\text{stab}}^k$ for all k.

Proof. Consider the sequence $\nu_k = 2c\beta_{\text{stab}}^{k-1}$. Since $\varepsilon_1 \leq c$, $\nu_1 \geq \varepsilon_1$. Moreover,

$$\begin{aligned} c\beta_{\text{stab}}^k + c\eta\Delta\beta_{\text{stab}}^{k-1} \sum_{j=1}^{k} \nu_j &\leq c\beta_{\text{stab}}^k + 2c^2\eta\Delta\beta_{\text{stab}}^{k-1} \sum_{j=1}^{k} \beta_{\text{stab}}^{j-1} \\ &\leq c\beta_{\text{stab}}^k + 2c^2\eta\Delta\beta_{\text{stab}}^{k-1} \frac{1}{1-\beta} \\ &\leq c\beta_{\text{stab}}^k \left(1 + 2c\Delta \frac{\eta}{\beta(1-\beta)}\right). \end{aligned}$$

Hence, for $\Delta \leq \frac{\beta(1-\beta)}{2c\eta}$, the above is at most $2c\beta_{\text{stab}}^k \leq \nu_{k+1}$. This shows that the (ν_k) sequence dominates the (ε_k) sequence, as needed. □

Lemma K.17 (Third Key Recursion)**.** Let $\eta > 0$, $\beta = (1 - \frac{\eta}{L})$, $L \geq 2\eta$, and let $C, C', \Delta_2, \Delta_1 > 0$. Suppose that $\varepsilon_1, \varepsilon_2, \dots$ satisfies,

$$\varepsilon_{t+1} \leq \eta \sum_{s=1}^{t} \beta^{t-s} (\Delta_2 + C\varepsilon_s^2)) + \beta^t \Delta_1$$

whenever $\max_{1 \leq s \leq t} \varepsilon_t \leq C'$. Suppose that

$$\Delta_2 \leq \frac{1}{\max\{8CL, 4LC'\}}, \quad \Delta_1 \leq \frac{1}{\max\{16CL, 8C'\}}$$

Then, for all t,

$$\varepsilon_t \leq 4\Delta_1 \beta^{t-1} + 2L\Delta_2.$$

Proof. Consider $\bar{\varepsilon}_t = \alpha_1\beta^{t-1} + \alpha_2$, with $\alpha_1 \geq \Delta_1$ and $\alpha_2 > 0$. As long as $\alpha_1 + \alpha_2 \leq C'$, we have $\bar{\varepsilon}_t \leq C'$ for all t. To show $\bar{\varepsilon}_t \geq \varepsilon_t$, it suffices that $\bar{\varepsilon}_t \geq \eta\sum_{s=1}^t \beta^{t-s}(\Delta_2 + C\bar{\varepsilon}_s^2)) + \beta^t\Delta_1$. To have this occur, we need

$$\begin{aligned}
&\eta\sum_{s=1}^t \beta^{t-s}(\Delta_2 + C\bar{\varepsilon}_s^2)) + \beta^t\Delta_1 \\
&\quad\leq \eta\sum_{s=1}^t \beta^{t-s}(\Delta_2 + 2C\alpha_1^2\beta^{2(s-1)} + 2C\alpha_2^2) + \beta^t\Delta_1 \\
&\quad\leq \eta\sum_{s=1}^t \beta^{t-s}(\Delta_2 + 2C\alpha_1^2\beta^{2(s-1)} + 2C\alpha_2^2) + \beta^t\Delta_1 \\
&\quad= (\Delta_2 + 2C\alpha_2^2)\cdot(\eta\sum_{s=1}^t \beta^{t-s}) + 2C\alpha_1^2\cdot(\eta\sum_{s=1}^t \beta^{t-s}\beta^{2(s-2)}) + \beta^t\Delta_1 \\
&\quad= (\Delta_2 + 2C\alpha_2^2)\cdot(\eta\sum_{s=1}^t \beta^{t-s}) + \beta^{t-1}(2C\alpha_1^2\cdot(\eta\sum_{s=1}^t \beta^{s-1}) + \Delta_1) \\
&\quad\leq L(\Delta_2 + 2C\alpha_2^2) + \beta^{t-1}(2C\alpha_1^2 L + \Delta_1),
\end{aligned}$$

where the last step upper bounds the geometric series with $\eta = (1 - \eta/L)$. Assuming $\eta \leq L/2$, the above is at most

$$L(\Delta_2 + 2C\alpha_2^2) + 2\beta^t(2C\alpha_1^2 L + \Delta_1).$$

Matching terms, it is enough that

$$\alpha_2 \geq L(\Delta_2 + 2C\alpha_2^2), \quad \alpha_1 \geq 2(2C\alpha_1^2 L + \Delta_1), \quad \alpha_1 + \alpha_2 \leq C'$$

Let's choose $\alpha_2 = 2L\Delta_2$ and $\alpha_1 = 4\Delta_1$. Then, it is enough that

$$L\Delta_2 \geq 8CL^2\Delta_2^2, \quad 2\Delta_1 \geq 32C\Delta_1^2 L, \quad (2L\Delta_2 + 4\Delta_1) \leq \frac{1}{C'}$$

For this, it suffices that $\Delta_2 \leq \frac{1}{\max\{8CL, 4LC'\}}$ and $\Delta_1 \leq \frac{1}{\max\{16CL, 8C'\}}$. ☐

Lemma K.18 (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,

$$\begin{aligned}
\|\mathbf{\Phi}'_{k+1,1}\|_{\mathrm{op}} &\leq c\eta\sum_{j=1}^k \beta_{\mathrm{stab}}^{k-j}\|\mathbf{X}_j - \mathbf{X}'_j\|\|\mathbf{\Phi}'_{j,1}\| + c\beta_{\mathrm{stab}}^k \\
&\leq c\eta\beta_{\mathrm{stab}}^{k-1}\Delta\sum_{j=1}^k \|\mathbf{\Phi}'_{j,1}\| + c\beta_{\mathrm{stab}}^k. \qquad (\|\mathbf{X}_j - \mathbf{X}'_j\| \leq \eta\Delta\beta_{\mathrm{stab}}^{j-1})
\end{aligned}$$

Define $\varepsilon_j = \|\mathbf{\Phi}'_{j,1}\|$. Then, $\varepsilon_1 = 1 \leq c$, so Lemma K.16 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. ☐

L Sampling and Score Matching

In this section, we provide a rigorous guarantee on the quality of sampling from the learned DDPM under Assumption C.1. 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. [18] to show that if we have learned the conditional score function, then sampling can be done efficiently. While Block et al. [13] 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.

We organize the section as follows:

- We then state the main result of the section, Theorem 13, 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 (L.1), we give the exact polynomial dependence of the sampling parameters α and J on the parameters of the problem.
- Before embarking on the proof, Appendix L.1 introduces simplifying notation; notably, dropping the dependence on subscript h, replacing the score dependence on j with a class Θ_j, and denoted $\mathcal{D}_{\sigma,h,[t]}$ as simply $q_{[t]}$.
- We break the proof of Theorem 13 into two sections. First, in Appendix L.2, we recall a result of Chen et al. [18], Lee et al. [41] 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 L.4, 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 L.3, we apply statistical learning techniques, similar to those in Block et al. [13], to show that with sufficiently many samples, we can effectively learn the score. We detail in Remark L.8 how the realizability part of Assumption C.1 can be relaxed.
- Finally, we conclude the proof of Theorem 13 by combining the two intermediate results detailed above.

To begin, we define our notion of statistical complexity:

We now state the main result of this section.

Theorem 13. Fix $1 \leq h \leq H$, let q denote $\mathcal{D}_{\sigma,h}$, d denote $d_{\mathcal{A}}$, and suppose that $(\mathsf{a}_i, \mathsf{o}_{h,i}) \sim q$ are independent for $1 \leq i \leq n$. Suppose that the projection of q onto the first coordinate has support (as defined in Definition F.3) contained in the euclidean ball of radius $R \geq 1$ in $\mathbb{R}^d$. For $\varepsilon > 0$, set

$$J = c\frac{d^3 R^4 (R + \sqrt{d})^4 \log\left(\frac{dR}{\varepsilon}\right)}{\varepsilon^{20}}, \qquad \alpha = c\frac{\varepsilon^8}{d^2 R^2 (R+\sqrt{d})^2}. \tag{L.1}$$

for some universal constant $c > 0$. Suppose that for all $1 \leq j \leq J$, 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 $\delta > 0$:

$$\sup_{\substack{\theta,\theta' \in \Theta_j \\ ||\mathsf{a}|| \vee ||\mathsf{a}'|| \leq R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)} \\ ||\mathsf{o}_h|| \leq R}} ||\mathbf{s}_\theta(\mathsf{a}, \mathsf{o}_h, t) - \mathbf{s}_{\theta'}(\mathsf{a}', \mathsf{o}_h, t)|| \leq c\frac{d^2 (R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8}.$$

- Assumption C.1 holds and thus, for all $j \in [J]$, it holds that $\mathcal{R}_n(\Theta_j) \leq C_\Theta \alpha^{-1} n^{-1/\nu}$ for some $\nu \geq 2$ and all $n \in \mathbb{N}$ and, moreover, the linear growth condition is satisfied.

- The parameter $\widehat{\theta} = \widehat{\theta}_{1:J}$ is defined to be the empirical minimizer of $\mathcal{L}_{\text{DDPM}}$ from Section 3.

If

$$n \geq c \left(\frac{C_\Theta \alpha^{-1} dR(R \vee \sqrt{d}) \log(dn)}{\varepsilon^4} \right)^{4\nu} \vee \left(\frac{d^6 (R^4 \vee d^2 \log^3 \left(\frac{ndR}{\delta\varepsilon} \right))}{\varepsilon^{24}} d^2 \right)^{4\nu},$$

then with probability at least $1 - \delta$, it holds that

$$\mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}} \left[\inf_{\mu \in \mathscr{C}(\mathit{DDPM}(\mathbf{s}_{\widehat{\theta}}, \mathsf{o}_h), q(\cdot|\mathsf{o}_h))} \mathbb{P}_{(\widehat{\mathsf{a}}, \mathsf{a}^*) \sim \mu} \left(||\widehat{\mathsf{a}} - \mathsf{a}^*|| \geq \varepsilon \right) \right] \leq 3\varepsilon.$$

Remark L.1. We emphasize that the exact value of the polynomial dependence (and in particular its pessimism) stem from the guarantees of Chen et al. [18], Lee et al. [41] 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. [18], Lee et al. [41] 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 L.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 $\mathcal{R}_n(\Theta) \leq C'_\Theta (\alpha/n)^{-1/\nu}$, then the sample complexity above becomes

$$n \geq c \left(\frac{C'_\Theta \log(dn)}{\alpha} \right)^{4\nu} \vee \left(\frac{d^2 (R^2 \vee d^2 \log^3 \left(\frac{ndR}{\varepsilon\delta} \right))}{\alpha^2 \varepsilon^{16}} \right)^{4\nu}.$$

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 discussed in Remark C.2, it is precisely this weaker guarantee that makes the problem challenging.

L.1 Simplifying Notation

We substantially simplify the notation in this appendix to suppress all dependence on h. In particular, we fix some $h \in [H]$ and consider $\mathsf{o}_h \sim \mathcal{D}_{\sigma,h,[0]}$. We let q denote $\mathcal{D}_{\sigma,h}$ and d denote $d_{\mathcal{A}}$. We further fix some $\sigma > 0$ and let $q_{[t]}$ denote the law of $\mathsf{a} \mid \mathsf{o}_h$ according to $\mathcal{D}_{\sigma,h,[t]}$ for the sake of notational simplicity. Furthermore, to emphasize that our analysis of the statistical learning theory decomposes accross DDPM time steps, we denote by Θ_j the function class $\mathbf{s}_\theta(\cdot, \cdot, \alpha j)$. We (redundantly) keep the dependence on t in the function evaluation for the sake of clarity. All other notation is defined *in situ*.

We emphasize that while our theoretical analysis treats each $\mathbf{s}_{\theta,h}$ separately, empirically one sees better success in training the score estimates jointly; on the other hand, the focus of this paper is not on sampling and score estimation and so we make the simplifying assumption for the sake of convenience.

L.2 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 $0 \leq t \leq T$:

$$d\mathsf{a}^t = -\mathsf{a}^t dt + \sqrt{2} dB_t, \quad \mathsf{a}^0 \sim q,$$

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 14 (Corollary 4, Chen et al. [18])**.** Suppose that a distribution q on $\mathbb{R}^d$ is supported on some ball of radius $R \geq 1$. Let C be a universal constant, fix $\varepsilon > 0$, and let α, J be set as in (L.1). If we have a score estimator $\mathbf{s}_\theta : \mathbb{R}^d \times [\tau] \to \mathbb{R}^d$ such that

$$\max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, j) - \nabla \log q_{[\alpha j]}(\mathsf{a}) \right|\right|^2 \right] \leq \varepsilon^4,$$

then

$$\sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \leq 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathrm{Law}(\mathsf{a}^J)} [f(\widehat{\mathsf{a}})] - \mathbb{E}_{\mathsf{a}^* \sim q} [f(\mathsf{a}^*)] \leq \varepsilon^2,$$

where a^J is sampled as in (2.2).

Remark L.3. As a technical aside, we note that Chen et al. [18, 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. [18, 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 14 in the way above as opposed to the more complicated version with the early stopping.

Remark L.4. While [18, 41] 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. [18] 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. [18, 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 14 applies to unconditional sampling, it is easy to derive conditional sampling guarantees as a corollary.

Corollary L.1. Suppose that q is a joint distribution on actions a and observations $\mathsf{o}_h \in \mathbb{R}^{d'}$. 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 (L.1) and

$$\mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}} \left[\max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}(\cdot|\mathsf{o}_h)} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, j, \mathsf{o}_h) - \nabla \log q_{[\alpha j]}(\mathsf{a}|\mathsf{o}_h) \right|\right|^2 \right] \right] \leq \varepsilon^4, \tag{L.2}$$

then

$$\mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}} \left[\inf_{\mu \in \mathscr{C}(\mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h), q(\cdot|\mathsf{o}_h))} \mathbb{P}_{(\widehat{\mathsf{a}}, \mathsf{a}^*) \sim \mu} \left(||\widehat{\mathsf{a}} - \mathsf{a}^*|| \geq \varepsilon \right) \right] \leq 3\varepsilon$$

Proof. We begin by showing an intermediate result,

$$\mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}} \left[\sup_{f : ||f||_\infty \vee ||||\nabla f||||_\infty \leq 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h)} [f(\widehat{\mathsf{a}})] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot|\mathsf{o}_h)} [f(\mathsf{a}^*)] \right] \leq 3\varepsilon^2. \tag{L.3}$$

using Theorem 14. Let

$$\mathcal{A} = \left\{ \max_{j \in [J]} \mathbb{E}_{\mathsf{a} \sim q_{[\alpha j]}(\cdot|\mathsf{o}_h)} \left[\left|\left| \mathbf{s}_\theta(\mathsf{a}, j, \mathsf{o}_h) - \nabla \log q_{[\alpha j]}(\mathsf{a}|\mathsf{o}_h) \right|\right|^2 \right] \leq \varepsilon^2 \right\}.$$

By Markov's inequality and (L.2), it holds that

$$\mathbb{P}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}}\left(\mathcal{A}^c\right) \le \frac{\varepsilon^4}{\varepsilon^2} = \varepsilon^2$$

and thus

$$\begin{aligned}
&\mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}}\left[\sup_{f\colon ||f||_\infty \vee ||||\nabla f||||_\infty \le 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h)}\left[f(\widehat{\mathsf{a}})\right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot|\mathsf{o}_h)}\left[f(\mathsf{a}^*)\right]\right] \\
&= \mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}}\left[\mathbf{I}[\mathcal{A}] \sup_{f\colon ||f||_\infty \vee ||||\nabla f||||_\infty \le 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h)}\left[f(\widehat{\mathsf{a}})\right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot|\mathsf{o}_h)}\left[f(\mathsf{a}^*)\right]\right] \\
&+ \mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}}\left[\mathbf{I}[A^c] \sup_{f\colon ||f||_\infty \vee ||||\nabla f||||_\infty \le 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h)}\left[f(\widehat{\mathsf{a}})\right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot|\mathsf{o}_h)}\left[f(\mathsf{a}^*)\right]\right] \\
&\le \mathbb{E}_{\mathsf{o}_h \sim q_{\mathsf{o}_h}}\left[\mathbf{I}[\mathcal{A}] \inf_{q' \in \Delta(\mathbb{R}^d)} W_2(q(\cdot|\mathsf{o}_h), q') + \mathrm{TV}\left(q', \mathrm{Law}(\pi^\tau)\right)\right] + 2\varepsilon^2.
\end{aligned}$$

For each o_h, we may apply Theorem 14 and observe that for $\mathsf{o}_h \in \mathcal{A}$,

$$\sup_{f\colon ||f||_\infty \vee ||||\nabla f||||_\infty \le 1} \mathbb{E}_{\widehat{\mathsf{a}} \sim \mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h)}\left[f(\widehat{\mathsf{a}})\right] - \mathbb{E}_{\mathsf{a}^* \sim q(\cdot|\mathsf{o}_h)}\left[f(\mathsf{a}^*)\right] \le \varepsilon^2,$$

which proves (L.3). Now, for any fixed o_h, by Markov's inequality and the definition of Wasserstein distance,

$$\inf_{\mu \in \mathscr{C}(\mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h), q(\cdot|\mathsf{o}_h))} \mathbb{P}_{(\widehat{\mathsf{a}}, \mathsf{a}^*) \sim \mu}\left(||\widehat{\mathsf{a}} - \mathsf{a}^*|| \ge \varepsilon\right) \le \frac{W_1(\mathtt{DDPM}(\mathbf{s}_\theta, \mathsf{o}_h), q(\cdot|\mathsf{o}_h))}{\varepsilon}.$$

The result follows. □

Note that the guarantee in Corollary L.1 is precisely what we need to control the one step imitation error in Theorem 4; 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.

L.3 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 [13], 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 L.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 o_h and $t = \alpha j$, we would like to minimize $\mathbb{E}_{\mathsf{a} \sim q_{[t]}}\left[\left|\left|\mathbf{s}_\theta(\mathsf{a}, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}|\mathsf{o}_h)\right|\right|^2\right]$, 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 [32, 74, 62], this objective can be shown to be equivalent to minimizing

$$\mathcal{L}_{\mathrm{DDPM}}(\theta, \mathsf{a}, \mathsf{o}, t) = \mathbb{E}_{\mathsf{a} \sim q_{[t]}}\left[\left|\left|\mathbf{s}_\theta\left(e^{-t}\mathsf{a} + \sqrt{1 - e^{-2t}}\gamma, \mathsf{o}_h, t\right) + \frac{1}{\sqrt{1 - e^{-2t}}}\gamma\right|\right|^2\right].$$

Because we are really interested in the expectation over the joint distribution $(\mathsf{a}, \mathsf{o}_h)$, we may take the expectation over o_h and recover (4.1) as the empirical approximation. We now prove the following result for a single time step t:

Proposition L.1. Suppose that q is a distribution such that $q(\cdot|\mathsf{o}_i)$ is supported on a ball of radius R for q-almost every o_h. For fixed $j \in [J]$ and α from (L.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 $\theta \in \Theta_j$,

$$\sup_{\substack{\theta,\theta' \in \Theta_j \\ ||\mathsf{a}|| \vee ||\mathsf{a}'|| \leq R \\ ||\mathsf{o}_h|| \leq R}} ||\mathbf{s}_\theta(\mathsf{a}, \mathsf{o}_h, t) - \mathbf{s}_{\theta'}(\mathsf{a}', \mathsf{o}_h, t)|| \leq c \frac{d^2 (R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8}$$

for some universal constant $c > 0$. Recall the Rademacher term $\mathcal{R}_n(\Theta_j)$ defined in Definition C.2, and let

$$\widehat{\theta} \in \operatorname*{arg\,min}_{\theta \in \Theta} \sum_{i=1}^n \mathcal{L}_{\mathrm{DDPM}}(\theta, \mathsf{a}_i, \mathsf{o}_i, t)$$

for independent and identically distributed $(\mathsf{a}_i, \mathsf{o}_i) \sim q$. Then it holds with probability at least $1 - \delta$ over the data that

$$\mathbb{E}_{(\mathsf{a}_t, \mathsf{o}_h) \sim q_{[t]}} \left[\left| \left| \mathbf{s}_{\widehat{\theta}}(\mathsf{a}_t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}_t | \mathsf{o}_h) \right| \right|^2 \right]$$
$$\leq c \cdot \sqrt{\frac{\log(dn)}{1 - e^{-2t}}} \left(\mathcal{R}_n(\Theta) + \frac{d^2 (R + \sqrt{d \log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8} \cdot \sqrt{\frac{d \log\left(\frac{4dn}{\delta}\right)}{n}} \right).$$

Remark L.5. We note that while we assume a linearly growing score function for the sake of simplicity, our analysis easily handles any polynomial growth with a mild resulting change in the constants, omitted for the sake of simplicity.

Before we provide a proof, we recall the following result:

Lemma L.2. Suppose that q is supported in a ball of radius R and let $t \geq \alpha$ for α as in (L.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,

$$\sup_{\substack{||\mathsf{a}|| \vee ||\mathsf{a}'|| \leq R \\ \mathsf{o}_h}} \left| \left| \nabla \log q_{[t]}(\mathsf{a} | \mathsf{o}_h) - \nabla \log q_{[t]}(\mathsf{a}' | \mathsf{o}_h) \right| \right| \leq 2LR$$

and there exists some assignment of Θ and $\mathbf{s}_\theta$ that satisfies the boundedness condition in Proposition L.1.

Proof. The first statement follows from eplacing the ε in Chen et al. [18, Lemma 20 (c)] by ε^2. The second statement follows immediately from the first. □

Remark L.6. Note that a slight variation of this result is what leads to the dependence on α in the growth parameter in Assumption C.1 allowing for realizability. Indeed, by Chen et al. [18, Lemma 20], it holds that the true score of $q_{[\alpha]}$ is

$$L = \frac{1}{1 - e^{-2\alpha}} \vee \frac{\left| 1 - e^{-2\alpha}(1 + R^2) \right|}{(1 - e^{-2\alpha})^2}$$

Lipschitz, which is $O(\alpha^{-1})$ for $\alpha \ll 1$.

We also require the following standard result:

Lemma L.3. If $\mathcal{R}_n(\Theta_j)$ is defined as in Definition C.2, then

$$\mathbb{E}_{\gamma_1, \ldots, \gamma_n} \left[\sup_{\substack{\theta \in \Theta_j \\ 1 \leq j \leq J}} \frac{1}{n} \cdot \sum_{i=1}^n \langle \mathbf{s}_\theta(\mathsf{a}, \mathsf{o}_i, j), \gamma_i \rangle \right] \leq \sqrt{\pi \log(dn)} \cdot \mathcal{R}_n(\Theta_j)$$

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 $\mathbb{E}\left[\max_{i,j}(\boldsymbol{\gamma}_i)_j\right] \le \sqrt{\pi \log(dn)}$ by classical Gaussian concentration. See Van Handel [69] for more details. $\square$

Proof of Proposition L.1. Let P_n denote the empirical measure on n independent samples $\{(\mathsf{a}_i, \mathsf{o}_i, \boldsymbol{\gamma}_i)\}$ 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,

$$P_n\left(\left|\left|C_t \mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \mathsf{o}_h, t) - \boldsymbol{\gamma}\right|\right|^2\right) \le \cdot P_n\left(\left|\left|C_t \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h) - \boldsymbol{\gamma}\right|\right|^2\right). \tag{L.4}$$

We emhasize that by Lemma L.2, realizability does not make the result vaccuous. Adding and subtracting $C_t \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h)$ from the left hand inequality, expanding and rearranging, we see that

$$\begin{aligned}C_t^2 P_n\left(\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h)\right|\right|^2\right) &\le 2C_t \cdot P_n\left(\left\langle \mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h), \boldsymbol{\gamma}\right\rangle\right) \\ &\le 2C_t \cdot P_n\left(\sup_{\theta\in\Theta_j} \left\langle \mathbf{s}_{\theta}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h), \boldsymbol{\gamma}\right\rangle\right).\end{aligned}$$

We now claim that with probability at least $1-\delta$, it holds that

$$\begin{aligned}P_n\left(\sup_{\theta\in\Theta} \left\langle \mathbf{s}_{\theta}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h), \boldsymbol{\gamma}\right\rangle\right) &\le \mathbb{E}\left[P_n\left(\sup_{\theta\in\Theta_j} \left\langle \mathbf{s}_{\theta}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h), \boldsymbol{\gamma}\right\rangle\right)\right] \\ &\quad + B \cdot \sqrt{\frac{d \log\left(\frac{2d}{\delta}\right)}{n}},\end{aligned}$$

where

$$B = c\frac{d^2(R + \sqrt{d\log\left(\frac{2nd}{\delta}\right)})^2}{\varepsilon^8} \tag{L.5}$$

for some universal constant $c > 0$. To see this, we claim that with probability at least $1 - \frac{\delta}{2}$, it holds that $||\mathsf{a}_i^t|| \le c\left(R + \sqrt{d\log\left(\frac{2nd}{\delta}\right)}\right)$ for all $1 \le i \le n$. Indeed, this follows by Gaussian concentration in Jin et al. [34, Lemmata 1 & 2]. We may now apply Lemma L.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 [75, Theorem 4.10] that if we let the function

$$G = G(\mathsf{a}_1, \mathsf{o}_1, \ldots, \mathsf{a}_n, \mathsf{o}_n) = P_n\left(\sup_{\theta\in\Theta_j} \left\langle \mathbf{s}_{\theta}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h), \boldsymbol{\gamma}\right\rangle\right),$$

then for any i, on the event of bounded norm, replacing $(\mathsf{a}_i, \mathsf{o}_i)$ with $(\mathsf{a}_i', \mathsf{o}_i')$ and leaving other terms unchanged changes ensures that $|G - G'| \le \frac{2B}{n}\boldsymbol{\gamma}_i$. Thus by Jin et al. [34, Corollary 7] and a union bound, the claim holds. Because $\boldsymbol{\gamma}$ is mean zero, we have

$$\begin{aligned}\mathbb{E}\left[P_n\left(\sup_{\theta\in\Theta} \left\langle \mathbf{s}_{\theta}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h), \boldsymbol{\gamma}\right\rangle\right)\right] &\le \mathbb{E}\left[P_n\left(\sup_{\theta\in\Theta} \left\langle \mathbf{s}_{\theta}(\mathsf{a}^t, \mathsf{o}_h, t), \boldsymbol{\gamma}\right\rangle\right)\right] \\ &\le \sqrt{\pi \log(dn)} \cdot \mathcal{R}_n(\Theta_j),\end{aligned}$$

where the last inequality follows by Lemma L.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

$$P_n\left(\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t, \mathsf{o}_h, t) - \nabla \log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h)\right|\right|^2\right) \le \frac{2}{C_t}\sqrt{\pi \log(nd)} \cdot \mathcal{R}_n(\Theta) + \frac{B}{C_t} \cdot \sqrt{\frac{d\log\left(\frac{2nd}{\delta}\right)}{n}},$$

with B given in (L.5). We now use a uniform norm comparison between population and empirical norms to conclude the proof. Indeed, it holds by Rakhlin et al. [54, Lemma 8.i & 9] that there exists a critical radius

$$r_n \le cB\log^3(n)\mathcal{R}_n(\Theta_j)^2$$

such that with probability at least $1-\delta$,

$$\mathbb{E}_{(\mathsf{a}^t,\mathsf{o}_h)\sim q_{[t]}}\left[\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t,\mathsf{o}_h,t)-\nabla\log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h)\right|\right|^2\right]$$
$$\leq 2\cdot P_n\left(\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t,\mathsf{o}_h,t)-\nabla\log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h)\right|\right|^2\right)+cr_n+c\frac{\log\left(\frac{1}{\delta}\right)+\log\log n}{n},$$

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

$$\mathbb{E}_{(\mathsf{a}^t,\mathsf{o}_h)\sim q_{[t]}}\left[\left|\left|\mathbf{s}_{\widehat{\theta}}(\mathsf{a}^t,\mathsf{o}_h,t)-\nabla\log q_{[t]}(\mathsf{a}^t|\mathsf{o}_h)\right|\right|^2\right]\leq\frac{4}{C_t}\sqrt{\pi\log(nd)}\cdot\mathcal{R}_n(\Theta_j)+\frac{2B}{C_t}\cdot\sqrt{\frac{d\log\left(\frac{4nd}{\delta}\right)}{n}}$$
$$+cB\log^3(n)\cdot\mathcal{R}_n^2(\Theta_j)+c\frac{\log\left(\frac{2}{\delta}\right)+\log\log(n)}{n}$$

with probability at least $1-\delta$. This concludes the proof. □

Remark L.7. For the sake of simplicity, in the proof of Proposition L.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. [13] with a slightly different loss. While in our proof the appeal to Rakhlin et al. [54] 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 L.8. 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 (L.4) with minimal modification to the proof.

With Proposition L.1, and a union bound, we recover the following result:

Proposition L.4. Suppose that the conditions on $\mathbf{s}_\theta$ in Proposition L.1 continue to hold. Suppose further that $||\mathbf{s}_\theta(\mathsf{a},\mathsf{o}_h,t)||\leq C_{\mathrm{grow}}(1+||\mathsf{a}||+||\mathsf{o}_h||)$ for all a and some universal constant $C_{\mathrm{grow}}>0$. Let J and α be as in (L.1) and suppose that $\alpha\leq\frac{1}{2}$. Then, with probability at least $1-\delta$ over $\mathcal{D}'$, it holds that

$$\mathbb{E}_{\mathsf{o}_h\sim q_{\mathsf{o}_h}}\left[\max_{j\in[J]}\mathbb{E}_{\mathsf{a}\sim q_{[\alpha j]}(\cdot|\mathsf{o}_h)}\left[\left|\left|\mathbf{s}_\theta(\mathsf{a},j,\mathsf{o}_h)-\nabla\log q_{[\alpha j]}(\mathsf{a}|\mathsf{o}_h)\right|\right|^2\right]\right]$$
$$\leq c\frac{d(R\vee\sqrt{d})^2\log(dn)}{\varepsilon^4}\mathcal{R}_n(\Theta)+c\frac{d^3\left((R\vee\sqrt{d})^2+d\log\left(\frac{ndR}{\delta\varepsilon}\right)\right)}{\varepsilon^{12}}\sqrt{\frac{d\log\left(\frac{4dnR}{\delta\varepsilon}\right)}{n}}$$

In particular if

$$\mathcal{R}_n(\Theta_j)\leq C_\Theta n^{-1/\nu}$$

for some $\nu\geq 2$ and all $j\in[J]$, then for

$$n\geq cC_{\mathrm{grow}}\left(\frac{C_\Theta\alpha^{-1}d(R\vee\sqrt{d})^2\log(dn)}{\varepsilon^4}\right)^{4\nu}\vee\left(\frac{d^6(R^4\vee d^3\log^3\left(\frac{ndR}{\delta\varepsilon}\right))}{\varepsilon^{24}}d^2\right)^{4\nu}$$

it holds that with probability at least $1-\delta$,

$$\mathbb{E}_{\mathsf{o}_h\sim q_{\mathsf{o}_h}}\left[\max_{j\in[J]}\mathbb{E}_{\mathsf{a}\sim q_{[\alpha j]}(\cdot|\mathsf{o}_h)}\left[\left|\left|\mathbf{s}_\theta(\mathsf{a},j,\mathsf{o}_h)-\nabla\log q_{[\alpha j]}(\mathsf{a}|\mathsf{o}_h)\right|\right|^2\right]\right]\leq\varepsilon^4.$$

Proof. We begin by proving the result on the event that $||\mathsf{a}||\vee||\mathsf{o}_h||\leq C(R\vee\sqrt{d})\log\left(\frac{Jn(R\vee\sqrt{d})}{\delta\varepsilon}\right)$. Note that

$$1-e^{-2t}\geq 1-e^{-2\alpha}\geq\alpha$$

because $2\alpha \le 1$. We now apply Proposition L.1 and take a union bound over $j \in [J]$. All that remains is to demonstrate that the contribution of the event that a^j is outside the above defined ball is negligable. To do this, observe that by Lee et al. [41, Lemma 4.15], there is some $C > 0$ such that a^j is $C(R \vee \sqrt{d})$-subGaussian. By the sublinearity of the growth of $\mathbf{s}_\theta$ in a, as well as the Lipschitzness of $q_{[\alpha j]}$ following from Chen et al. [18, Lemma 20], bounding a maximum by a sum, and the elementary computation in Lemma L.5, we have that the expectation of this term on this event is bounded by $\frac{CC_{\mathrm{grow}}}{n}$. 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 [75].

Proof of Theorem 13. We note that the proof follows immediately from combining Corollary L.1 with Proposition L.4. □

We conclude the section with the following elementary computation used above:

Lemma L.5. Suppose that X is a σ-subGaussian random variable on $\mathbb{R}$. Then for any $r \ge 1$,

$$\mathbb{E}\left[|X| \cdot \mathbf{I}[|X| > r]\right] \le C\frac{\sigma}{r} \cdot e^{-\frac{r^2}{2\sigma^2}}$$

Proof. This is an elementary computation. Indeed,

$$\begin{aligned}
\mathbb{E}\left[|X| \cdot \mathbf{I}[|X| > r]\right] &= \int_r^\infty \mathbb{P}\left(|X| > t\right) dt \le C \int_r^\infty e^{-\frac{t^2}{2\sigma^2}} dt \\
&\le C \cdot \int_r^\infty \frac{t}{r} e^{-\frac{t^2}{2\sigma^2}} dt \\
&\le C\frac{\sigma}{r} \cdot e^{-\frac{r^2}{2\sigma^2}}.
\end{aligned}$$

The result follows. □

M Proofs for Generic Incrementally Stable Primitive Controllers

This section proves Theorems 5 and 6, gneeralizing our guarantees to general primitive controllers. Note that, in this more general setting, we can no longer expect to bound the norm of the difference between two controllers evaluated at some point $\mathbf{x}$ $\kappa_t(\mathbf{x}) - \kappa_t(\mathbf{x}')$ by differences in their parameter values. Instead, we opt for the more local notion of distance considered in Theorems 5 and 6, via the localized distance $\mathsf{d}_{\mathrm{loc},\alpha}$ considered in Definition E.1. To this end, Appendix M.1 begins by generalizing the analysis of the composite MDP to allow the distance $\mathsf{d}_{\mathcal{A}}$ take an additional state-argument (in order to capture the localization of the distance in $\mathsf{d}_{\mathrm{loc},\alpha}$). Appendix M.2 then converts our assumption of incremental stability, Assumption 3.1b, into the IPS stability conditioned required in the composite MDP. Finally, we conclude of our intended results in Appendix M.3, following the same arguments as for affine primitive controllers in Appendix J.

M.1 Generalization of analysis in the composite MDP

Here, we consider a generalization of the analysis of the composite MDP where we allow $\mathsf{d}_{\mathcal{A}}$ to depend on state. Our analysis follows Appendix H.3 and the proof of Theorem 9. All notation here borrows from that section. Formally, we consider

$$\mathsf{d}_{\mathcal{A};\mathcal{S}}(\cdot, \cdot \mid \cdot) : (\mathcal{A} \times \mathcal{A}) \times \mathcal{S} \to \mathbb{R}_{\ge 0}.$$

We recall the direct decomposition in Definition H.1 of $\mathcal{S} = \mathcal{Z} \oplus \mathcal{S}_{/\mathcal{O}}$, where we recall that $\mathcal{Z}$ is the component that coincides with the 'o_h' component of the state in our instantiation. Further, recall that ϕ_{o} is the projection onto the $\mathcal{Z}$ component.

Condition M.1 (Measurability)**.** We require that $\mathsf{d}_{\mathcal{A};\mathcal{S}}$ is measurable, and that, for all s, the set $\{(\mathsf{a}',\mathsf{a}) : \mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}',\mathsf{a};\mathsf{s}) > \varepsilon\}$ is open. and that $(\mathsf{a}',\mathsf{a},\mathsf{s}) \mapsto \mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}',\mathsf{a};\mathsf{s})$ is measurable. We also assume that $\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}',\mathsf{a};\mathsf{s})$ only depends on s through $\phi_{\mathrm{o}}(\mathsf{s})$.

We re-define a state-conditioned input stability as follows

Definition M.1. We say that a sequence $(\mathsf{s}_{1:H+1},\mathsf{a}_{1:H})$ is state-conditioned input-stable with respect to an auxilliary sequence $\tilde{\mathsf{s}}_{1:H+1}$ if

$$\mathsf{d}_{\mathcal{S}}(\mathsf{s}'_{h+1},\mathsf{s}_{h+1}) \vee \mathsf{d}_{\mathrm{TVC}}(\mathsf{s}'_{h+1},\mathsf{s}_{h+1}) \le \max_{1\le j\le h} \mathsf{d}_{\mathcal{A};\mathcal{S}}\left(\mathsf{a}'_j,\mathsf{a}_j \mid \tilde{\mathsf{s}}_j\right), \ \forall h \in [H]$$

We now define a one-step error which is state dependent (allowing for $\mathsf{d}_{\mathcal{A};\mathcal{S}}$). To simplify the exposition, we define marginal gaps which ignore the now-state-dependent $\mathsf{d}_{\mathcal{A};\mathcal{S}}$.

Definition M.2 (Modified Imitation Gaps)**.** Define the state

$$\mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_h(\mathsf{s}),\pi^\star_h(\mathsf{s}) \mid \mathsf{s}') := \inf_{\mu_2} \mathbb{P}_{\mu_2}\left[\mathsf{d}_{\mathcal{A};\mathcal{S}}(\hat{\mathsf{a}}_h,\mathsf{a}^\star_h \mid \mathsf{s}') \le \varepsilon\right],$$

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

$$\Gamma_{\mathrm{marg},\mathcal{S},\varepsilon} := \inf_{\mu_1} \mathbb{P}_{\mu_1}\left[\max_{h\in[H]} \max\{\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1},\hat{\mathsf{s}}_{h+1}) > \varepsilon\right], \quad \Gamma_{\mathrm{joint},\mathcal{S},\varepsilon} := \max_{h\in[H]} \inf_{\mu_1} \mathbb{P}_{\mu_1}[\mathsf{d}_{\mathcal{S}}(\mathsf{s}^\star_{h+1},\hat{\mathsf{s}}_{h+1}) > \varepsilon]$$

where above μ_1 ranges over the same couplings as in Definition D.1.

Guarantees under TVC of $\hat{\pi}$. We now generalize Proposition D.2 under the assumption that $\hat{\pi}$ is TVC.

Proposition M.1 (Generalization of Proposition D.2)**.** Let $\pi^\star$ be state-conditioned input-stable w.r.t. $(\mathsf{d}_{\mathcal{S}},\mathsf{d}_{\mathcal{A};\mathcal{S}})$ and let $\hat{\pi}$ be γ-TVC. Then, for all $\varepsilon > 0$,

$$\Gamma_{\mathrm{joint},\mathcal{S},\varepsilon}(\hat{\pi} \parallel \pi^\star) \le H\gamma(\varepsilon) + \sum_{h=1}^{H} \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h} \mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_h(\mathsf{s}^\star_h) \parallel \pi^\star_h(\mathsf{s}^\star_h) \mid \mathsf{s}^\star_h).$$

Proof Sketch. The proof is nearly identical to the proof of Proposition D.2 in Appendix G. The only difference is that, when we measure the distance between $\mathsf{a}^\star_h \sim \pi^\star_h(\mathsf{s}^\star_h)$ and $\hat{\mathsf{a}}^{\mathrm{inter}}_h \sim \hat{\pi}_h(\mathsf{s}^\star_h)$, this distance is specified at $\mathsf{s}^\star_h$. Hence, we replace $\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}^{\mathrm{inter}}_h,\mathsf{a}^\star_h)$ with $\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}^{\mathrm{inter}}_h,\mathsf{a}^\star_h \mid \mathsf{s}^\star_h)$. This leads to use replacing $\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_h(\mathsf{s}^\star_h) \parallel \pi^\star(\mathsf{s}^\star_h))$. with $\mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_h(\mathsf{s}^\star_h) \parallel \pi^\star_h(\mathsf{s}^\star_h) \mid \mathsf{s}^\star_h)$ in the final bound. ☐

Guarantees with smoothing kernel. Next, we turn to the generalization of Theorems 4 and 9 to allow for state-conditioned action distances.

Definition M.3. Given non-decreasing maps $\gamma_{\mathrm{IPS,TVC}},\gamma_{\mathrm{IPS},\mathcal{S}} : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ 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 $(\gamma_{\mathrm{IPS,TVC}},\gamma_{\mathrm{IPS},\mathcal{S}},\mathsf{d}_{\mathrm{IPS}},r_{\mathrm{IPS}})$*-state-conditioned-restricted IPS* if it satisfies the conditions of Definition H.9, with the only modification that for the constructed $\mathsf{s}_{1:H+1},\mathsf{a}_{1:H},\tilde{\mathsf{s}}_{1:H}$, the condition that $\mathsf{s}_{1:H+1},\mathsf{a}_{1:H}$ is input-stable is replaced with "state-conditioned input stable with respect to the sequence $\tilde{\mathsf{s}}_{1:H}$." More precisely, the condition is met 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

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

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 state-conditioned input stable with respect to the sequence $\tilde{\mathsf{s}}_{1:H}$

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

Theorem 15. Consider the setting of Theorem 9, but with $\pi^\star$ satisfies $(\gamma_{\text{IPS,TVC}}, \gamma_{\text{IPS},\mathcal{S}}, \mathsf{d}_{\text{IPS}}, r_{\text{IPS}})$-state-conditioned-restricted IPS (Definition M.3) rather than (standard) restricted IPS (Definition H.9). Again, let $\varepsilon > 0$ and $r \in (0, \frac{1}{2} r_{\text{IPS}}]$, and efine

$$p_r := \sup_{\mathsf{s}} \mathbb{P}_{\mathsf{s}' \sim \mathsf{W}_\sigma(\mathsf{s})}[\mathsf{d}_{\text{IPS}}(\mathsf{s}', \mathsf{s}) > r], \quad \varepsilon' := \varepsilon + \gamma_{\text{IPS},\mathcal{S}}(2r)$$

Then, for any policy $\hat{\pi}$, both $\Gamma_{\text{joint},\mathcal{S},\varepsilon}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\text{rep}})$ and $\Gamma_{\text{marg},\mathcal{S},\varepsilon'}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ are upper bounded by

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

Proof. The proof follows by modifying Theorem 8, and hence Theorem 9 as a consequence. The key change is that we replace the event $\mathcal{B}_{\text{est},h} = \left\{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\text{tel,inter}}, \mathsf{a}_h^{\text{tel}}) > \varepsilon\right\}$ [15] in Definition H.6 with

$$\mathcal{B}_{\text{est},h} = \left\{\mathsf{d}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\text{tel,inter}}, \mathsf{a}_h^{\text{tel}} \mid \tilde{\mathsf{s}}_h^{\text{tel}}) > \varepsilon\right\}, \tag{M.1}$$

and replace the event $\mathcal{Q}_{\text{IS}}$ in Definition H.8 with

$$\mathcal{Q}_{\text{IS}} := \left\{\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H} \text{ is state-conditioned input-stable w.r.t. } \tilde{\mathsf{s}}^{\text{rep}}_{1:H}\right\}$$

We also define the following event

$$\mathcal{Q}'_{\text{IS},h} := \left\{\mathsf{s}^{\text{rep}}_{1:h+1}, \mathsf{a}^{\text{rep}}_{1:h} \text{ is state-conditioned input-stable w.r.t. } \tilde{\mathsf{s}}^{\text{tel}}_{1:h}\right\},$$

which considers input stability for $h \le H$ steps and shifts the reference sequence from $\tilde{\mathsf{s}}^{\text{rep}}_{1:h}$ to $\tilde{\mathsf{s}}^{\text{tel}}_{1:h}$. What changes as a result of these argument is as follows:

- We check that Lemma H.7 goes through:

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

  The first modification here is that, when $\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H}$ is input stable with respect to $\tilde{\mathsf{s}}^{\text{rep}}_{1:H}$,

  $$\mathsf{d}_{\mathcal{S}}(\hat{\mathsf{s}}_{h+1}, \mathsf{s}^{\text{rep}}_{h+1}) \vee \mathsf{d}_{\text{TVC}}(\hat{\mathsf{s}}_{h+1}, \mathsf{s}^{\text{rep}}_{h+1}) \le \max_{1 \le j \le h} \mathsf{d}_{\mathcal{A};\mathcal{S}}\left(\hat{\mathsf{a}}_j, \mathsf{a}^{\text{rep}}_j \mid \tilde{\mathsf{s}}^{\text{rep}}_j\right), \ \forall h \in [H]$$

  Since $\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}, \mathsf{a}' \mid \mathsf{s})$ depends only on s through $\phi_{\text{o}}(\mathsf{s})$, then when

  $$\bigcap_{1 \le j \le h} \mathcal{B}_{\text{tel},j} := \left\{\mathsf{a}^{\text{rep}}_j = \mathsf{a}^{\text{tel}}_j, \ \phi_{\text{o}}(\tilde{\mathsf{s}}^{\text{rep}}_j) = \phi_{\text{o}}(\tilde{\mathsf{s}}^{\text{tel}}_j)\right\}, \subset \bar{\mathcal{B}}_{\text{all},h},$$

  holds, we have

  $$\max_{1 \le j \le h} \mathsf{d}_{\mathcal{A};\mathcal{S}}\left(\hat{\mathsf{a}}_j, \mathsf{a}^{\text{rep}}_j \mid \tilde{\mathsf{s}}^{\text{rep}}_j\right) = \max_{1 \le j \le h} \mathsf{d}_{\mathcal{A};\mathcal{S}}\left(\hat{\mathsf{a}}_j, \mathsf{a}^{\text{tel}}_j \mid \tilde{\mathsf{s}}^{\text{tel}}_j\right), \ \forall h \in [H]$$

  Finally, on $\bar{\mathcal{B}}_{\text{all},h}$, we have $\hat{\mathsf{a}} = \hat{\mathsf{a}}^{\text{tel,inter}}$, so we get

  $$\max_{1 \le j \le h} \mathsf{d}_{\mathcal{A};\mathcal{S}}\left(\hat{\mathsf{a}}_j, \mathsf{a}^{\text{rep}}_j \mid \tilde{\mathsf{s}}^{\text{rep}}_j\right) = \max_{1 \le j \le h} \mathsf{d}_{\mathcal{A};\mathcal{S}}\left(\hat{\mathsf{a}}^{\text{tel,inter}}_j, \mathsf{a}^{\text{tel}}_j \mid \tilde{\mathsf{s}}^{\text{tel}}_j\right), \ \forall h \in [H],$$

  which is at most ε under our second definition of $\mathcal{B}_{\text{est},h}$.

- Lemma H.8 goes unchanged

- We check that Lemma H.9 goes through. This follows from the defintion of state-conditioned input-stable, using $\mathsf{s}^{\text{rep}}_{1:H+1}, \mathsf{a}^{\text{rep}}_{1:H}, \tilde{\mathsf{s}}^{\text{rep}}_{1:H}$ as $\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H}, \tilde{\mathsf{s}}_{1:H}$, so that when $\mathbb{P}_\mu[\mathcal{Q}^c_{\text{IS}} \cap \mathcal{Q}_{\text{close}}] = 0$.[16]

[15] This is the special case of $\mathcal{B}_{\text{est},h} = \left\{\vec{\mathsf{d}}_{\mathcal{A}}(\hat{\mathsf{a}}_h^{\text{tel,inter}}, \mathsf{a}_h^{\text{tel}}) \npreceq \vec{\varepsilon}\right\}$ with $\vec{\mathsf{d}}_{\mathcal{A}}$ being scalar valued (i.e. all coordinate identical).

[16] Here, we replace p_{IPS} in that lemma with failure probability 0.

- Recall that $\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a},\mathsf{a}';\mathsf{s})$ depends only on s through $\phi_{\mathrm{o}}(\mathsf{s})$. Hence, under the event

$$\bigcap_{1\le j\le h} \mathcal{B}_{\mathrm{tel},j} := \left\{\mathsf{a}_j^{\mathrm{rep}} = \mathsf{a}_j^{\mathrm{tel}},\ \phi_{\mathrm{o}}(\tilde{\mathsf{s}}_j^{\mathrm{rep}}) = \phi_{\mathrm{o}}(\tilde{\mathsf{s}}_j^{\mathrm{tel}})\right\}, \subset \bar{\mathcal{B}}_{\mathrm{all},h}$$

we have that $\mathcal{Q}_{\mathrm{IS}}$ implies $\mathcal{Q}'_{\mathrm{IS},h}$.

- Lemma H.10 replaces $\mathsf{d}_{\mathrm{os},\varepsilon}(\hat{\pi}_{\sigma,h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \parallel \pi^\star_{\mathrm{rep},h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}))$ with $\mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_{\sigma,h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \parallel \pi^\star_{\mathrm{rep},h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \mid \tilde{\mathsf{s}}_h^{\mathrm{tel}})$, i.e. the state-conditioned one-step error (that we condition on $\tilde{\mathsf{s}}_h^{\mathrm{tel}}$ comes from our re-definition of $\mathcal{B}_{\mathrm{est},h}$ in (M.1). Thus, we get $\Gamma_{\mathrm{joint},\mathcal{S},\varepsilon}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star_{\mathrm{rep}})$ and $\Gamma_{\mathrm{marg},\mathcal{S},\varepsilon'}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star)$ are upper bounded by

$$H\left(2p_r + 3\gamma_\sigma(\max\{\varepsilon,\gamma_{\mathrm{IPS},\mathrm{TVC}}(2r)\})\right) + \sum_{h=1}^{H} \mathbb{E}_{\tilde{\mathsf{s}}_h^{\mathrm{tel}}} \mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_{\sigma,h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \parallel \pi^\star_{\mathrm{rep},h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \mid \tilde{\mathsf{s}}_h^{\mathrm{tel}}). \tag{M.2}$$

- Because $\tilde{\mathsf{s}}_h^{\mathrm{tel}}$ has marginal $\mathsf{s}_h^\star \sim \mathsf{P}_h^\star$, we can replace the terms $\mathbb{E}_{\tilde{\mathsf{s}}_h^{\mathrm{tel}}} \mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_{\sigma,h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \parallel \pi^\star_{\mathrm{rep},h}(\tilde{\mathsf{s}}_h^{\mathrm{tel}}) \mid \tilde{\mathsf{s}}_h^{\mathrm{tel}})$ with $\mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star} \mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^\star) \parallel \pi^\star_{\mathrm{rep},h}(\mathsf{s}_h^\star) \mid \mathsf{s}_h^\star)$.

- Using the same data-processing argument as in the proof as in Theorem 8, we can bound $\mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star} \mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_{\sigma,h}(\mathsf{s}_h^\star) \parallel \pi^\star_{\mathrm{rep},h}(\mathsf{s}_h^\star) \mid \mathsf{s}_h^\star) \le \mathbb{E}_{\mathsf{s}_h^\star\sim\mathsf{P}_h^\star} \mathbb{E}_{\tilde{\mathsf{s}}_h^\star\sim\mathsf{W}_\sigma(\mathsf{s}_h^\star)} \mathsf{d}_{\mathrm{os},\mathcal{S},\varepsilon}(\hat{\pi}_h(\tilde{\mathsf{s}}_h^\star) \parallel \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}_h^\star) \mid \mathsf{s}_h^\star)$.

□

M.2 State-Conditioned Input-Stability and IPS in the Composite MDP via t-ISS

Lemma K.3 reduced IPS in the composite MDP to incremental stability in a form that applies primarily to affine primitive controllers. In this section, we generalize the lemma further to depend on a more localized distance reflecting the state-conditioned distance $\mathsf{d}_{\mathcal{A};\mathcal{S}}(\cdot,\cdot;\cdot)$ in the composite MDP. Recall the local-distance between composite actions $\mathsf{a} = \kappa_{1:\tau_{\mathrm{chunk}}}, \mathsf{a}' = \kappa'_{1:\tau_{\mathrm{chunk}}} \in \mathcal{A}$ at state $\mathbf{x}$ and scale $\alpha > 0$, defined in Definition E.1 as

$$\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a},\mathsf{a}' \mid \mathbf{x}) := \max_{1\le i\le \tau_{\mathrm{chunk}}} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\|\le\alpha} \|\kappa_i(\mathbf{x}_i+\delta\mathbf{x}) - \kappa'_i(\mathbf{x}_i+\delta\mathbf{x})\|,$$

where above $\mathbf{x}_1 = \mathbf{x}$, $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \kappa_t(\mathbf{x}))$ with $\mathsf{a} = \kappa_{1:\tau_{\mathrm{chunk}}}$.

Lemma M.2. Instantiate the composite MDP as in Definition K.1, with $\pi^\star$ as in Definition J.3. Furthermore, suppose that under $(\boldsymbol{\rho}_T, \mathsf{a}_{1:H}) \sim \mathcal{D}_{\mathrm{exp}}$ with $\boldsymbol{\rho}_T = (\mathbf{x}_{1:T+1}, \mathbf{u}_T)$, the following both hold with probability one:

- Each action a_h satisfies our notion of incremental stability (Definition 3.1) with moduli $\gamma(\cdot), \beta(\cdot,\cdot)$, constants $c_\gamma, c_\xi > 0$ (i.e. Assumption 3.1b holds)

Finally, let $\varepsilon_0 > 0$ satisfy (E.1), that is:

$$\gamma^{-1}(\beta(2\gamma(\varepsilon_0),\tau_{\mathrm{chunk}}) \le \varepsilon_0 \le \min\{c_\gamma, \gamma^{-1}(c_\xi/4)\},$$

Further, given $\tilde{\mathsf{s}} \in \mathcal{S} = \mathscr{P}_{\tau_{\mathrm{chunk}}}$ with last-step $\tilde{\mathbf{x}}_{\tau_{\mathrm{chunk}}}$, consider the distance-like function

$$\bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a},\mathsf{a};\alpha,\tilde{\mathsf{s}}) := \psi(\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a},\mathsf{a}' \mid \tilde{\mathbf{x}}_{\tau_{\mathrm{chunk}}})) \cdot \mathcal{I}_\infty\left\{\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a},\mathsf{a}' \mid \tilde{\mathbf{x}}_{\tau_{\mathrm{chunk}}}) \le \varepsilon_0\right\}, \quad \psi(u) := 2\beta(2\gamma(u),0).$$

Then, the following hold:

(a) $\pi^\star$ is state-conditioned input-stable (Definition M.1) with respect to $\mathsf{d}_{\mathcal{S}}, \mathsf{d}_{\mathrm{TVC}}$ as defined in Appendix D

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a},\mathsf{a}';\mathsf{s}) = \bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a},\mathsf{a}';\psi(\varepsilon),\mathsf{s}),$$

(b) For any $r_{\mathrm{IPS}} \le c_\xi/2$, $\pi^\star$ is $(r_{\mathrm{IPS}}, \gamma_{\mathrm{IPS},\mathrm{TVC}}, \gamma_{\mathrm{IPS},\mathcal{S}}, \mathsf{d}_{\mathrm{IPS}})$-state-conditioend restricted-IPS (Definition M.3) with

$$\mathsf{d}_{\mathcal{A}}(\mathsf{a},\mathsf{a}';\mathsf{s}) = \bar{\mathsf{d}}_{\mathcal{A}}(\mathsf{a},\mathsf{a}';\psi(\varepsilon)+\beta(r_{\mathrm{IPS}},0),\mathsf{s}), \quad \gamma_{\mathrm{IPS},\mathrm{TVC}}(r) = \beta(r,\tau_{\mathrm{chunk}}-\tau_{\mathrm{obs}}), \quad \gamma_{\mathrm{IPS},\mathcal{S}}(r) = \beta(r,0),$$

Proof of Lemma M.2. The proof is nearly identical to that of Lemma K.3, based on Lemma K.2. The only difference is that, rather than using the worst-case bound $\mathbf{x}_t \in \mathcal{X}_0$, we condition on the relevant states. For part (a), we consider $(\mathsf{s}_{1:H+1}, \mathsf{a}_{1:H})$ be drawn from the distribution induces by $\pi^\star$, and let $\mathsf{a}'_{1:H}$ be some other sequences of actions, and measure $\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}_h, \mathsf{a}_h; \mathsf{s}_h)$. Thus the relevant control-state to condition on is $\mathbf{x}_{t_h}$ in the construction of Lemma K.2. For verifying (b), we instead condition on $\tilde{\mathbf{x}}_{t_h}$ because, as in Definition M.3, we measure the input-state stability condition for restricted-state-conditioned-IPS with sequence to the states $\tilde{\mathsf{s}}_{1:H}$. □

M.3 Concluding the proof of Theorems 5 and 6

M.3.1 Proof of Theorem 6

The result is a direct consequence of the following points. First, with our instantion of the composite MDP, we can bound $\mathcal{L}_{\mathrm{marg},\varepsilon}(\hat{\pi}) \le \Gamma_{\mathrm{marg},\mathcal{S},\varepsilon}(\hat{\pi} \parallel \pi^\star) \le \Gamma_{\mathrm{joint},\mathcal{S},\varepsilon}(\hat{\pi} \parallel \pi^\star)$ by the same argument in Lemma J.1[17]; by a similar argument, we have $\mathcal{L}_{\mathrm{joint},\varepsilon}(\hat{\pi}) \le \Gamma_{\mathrm{joint},\mathcal{S},\varepsilon}(\hat{\pi} \parallel \pi^\star)$ when $\mathcal{D}_{\mathrm{exp}}$ has $\tau \le \tau_{\mathrm{obs}}$-bounded memory. The bound now follows from Proposition M.1, the fact that Lemma M.2 verifies the input-stability property (with $\varepsilon, \tau_{\mathrm{chunk}}$ satisfies (E.1)). □

M.3.2 Proof of Theorem 5

We begin with a lemma that upper bounds the imitation gaps by $\Delta_{\mathrm{ISS},\sigma,h}(\hat{\pi}; \varepsilon, \alpha(\varepsilon) + 2\upbeta(2r, 0))$ and other relevant terms. Essentially, the following lemma combines the general imitation guarantee in Theorem 15 with the incremental stability analysis in Lemmas K.2 and M.2.

Lemma M.3. Consider the instantiation of the composite MDP as in Definition K.1, let $r \le c_\xi/4$, and recall $\alpha(\varepsilon) = 2\upbeta(2\upgamma(\varepsilon), 0)$. Further suppose that ε satisfy Eq. (E.1). Then, the the modified imitation gaps (whose definition we recall Definition M.2) $\Gamma_{\mathrm{joint},\mathcal{S},\alpha(\varepsilon)}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star_{\mathrm{rep}})$ and $\Gamma_{\mathrm{marg},\mathcal{S},\alpha(\varepsilon)+\upbeta(2r,0)}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star)$ are bounded above by

$$H\left(4p_r + 3\gamma_\sigma\left(\max\{\alpha(\varepsilon), \upbeta(2r, \tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})\}\right)\right) + \sum_{h=1}^{H} \Delta_{\mathrm{ISS},\sigma,h}(\hat{\pi}; \varepsilon, \alpha(\varepsilon) + 2\upbeta(2r, 0)).$$

Before proving the lemma, lets quickly show how it implies the desired theorem. We bound

$$\begin{aligned}\mathcal{L}_{\mathrm{marg},2\upbeta(2\upgamma(\varepsilon),0)+2\upbeta(2r,0)} &\le \mathcal{L}_{\mathrm{marg},2\upbeta(2\upgamma(\varepsilon),0)+\upbeta(2r,0)} \\ &= \mathcal{L}_{\mathrm{marg},\alpha(\varepsilon)+\upbeta(2r,0)}(\hat{\pi}) \le \Gamma_{\mathrm{marg},\mathcal{S},\alpha(\varepsilon)+\upbeta(2r,0)}(\hat{\pi} \circ \mathsf{W}_\sigma \parallel \pi^\star),\end{aligned}$$

where the last inequality is due to as in the proof of Theorem 6, the intermediate inequality uses the definition of $\alpha(\varepsilon)$, and the first inequality uses anti-monotonicity of $\mathcal{L}_{\mathrm{marg},\varepsilon}$ in ε. Moreover, as shown in the proof of Theorem 2 in (J.4) and (J.5), we can take $\gamma_\sigma(u) = \frac{u\sqrt{2\tau_{\mathrm{obs}}-1}}{2\sigma}$ and for $p_r \le p$ when $r = \sigma\upomega_p$, $\upomega_p := 2\sqrt{5d_x + 2\log\left(\frac{1}{p}\right)}$ and $\mathsf{W}_\sigma(\cdot)$ is the Gaussian Kernel in (J.1). Hence, we conclude that if $\sigma \le c_\xi/4\upomega_p$,

$$\mathcal{L}_{\mathrm{marg},\varepsilon_1(p)} \le H\left(4p + \frac{3\sqrt{2\tau_{\mathrm{obs}}-1}}{2\sigma}\left(\max\left\{\varepsilon_2, \upbeta(2\sigma\upomega_p, \tau_{\mathrm{chunk}} - \tau_{\mathrm{obs}})\right\}\right)\right) + \sum_{h=1}^{H} \Delta_{\mathrm{ISS},\sigma,h}(\hat{\pi}; \varepsilon, \varepsilon_1(p)),$$

where above $\varepsilon_1(p) = 2\upbeta(2\upgamma(\varepsilon), 0) + 2\upbeta(2\sigma\upomega_p, 0)$ and $\varepsilon_2 = 2\upbeta(2\upgamma(\varepsilon), 0)$, as needed. Since $\upgamma(\varepsilon) \le 2\sigma$, in we can choose $p = \frac{\upgamma(\varepsilon)}{\sigma 2}$ and upper bound $\upomega_p \le \upomega(\varepsilon) := 2\sqrt{5d_x + 2\log\left(\frac{2\sigma}{\upgamma(\varepsilon)}\right)}$. The bound now follows from this upper bound and the bound $4p = 4\frac{2\cdot 2\upgamma(\varepsilon)}{\sigma 8} \le 4\frac{2\upbeta(2\upgamma(\varepsilon),0)}{\sigma 8} \le \frac{\varepsilon_2}{2\sigma}$, the first inequality follows from Observation K.4. □

[17] Here, $\Gamma_{\mathrm{marg},\mathcal{S},\varepsilon}, \Gamma_{\mathrm{joint},\mathcal{S},\varepsilon}$ are defined in in Definition M.2. The only difference between these the standard gaps $\Gamma_{\mathrm{marg},\varepsilon}, \Gamma_{\mathrm{joint},\varepsilon}$ consider in Definition M.2 is that they drop the closesness on composite actions, which is immaterial for $\mathcal{L}_{\mathrm{marg},\varepsilon}(\hat{\pi})$.

Proof of Lemma M.3. Recall the replica and deconvolution kernels $\mathsf{W}^\star_{\mathrm{dec},h}(\cdot), \mathsf{W}_{\mathrm{rep},h}(\cdot)$ defined in Definition H.5. We have that

$$\mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}\mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\mathcal{S},\alpha}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h) \mid \mathsf{s}^\star_h) = \mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}\mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \inf_\mu \mathbb{P}_\mu[\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}', \mathsf{a} \mid \mathsf{s}^\star_h) > \alpha] \tag{M.3}$$

where $\inf_\mu$ is over all couplings $\mathsf{a}_h \sim \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h), \mathsf{a}'_h \sim \hat{\pi}(\tilde{\mathsf{s}}^\star_h)$. By the gluing lemma (Lemma F.2), each coupling in μ is equivalent to a coupling $\hat{\mu}$ over $(\mathsf{s}^\star_h, \tilde{\mathsf{s}}^\star_h, \hat{\mathsf{s}}^\star_h, \mathsf{a}, \mathsf{a}')$ where

- $\mathsf{s}^\star_h \sim \mathsf{P}^\star_h$, $\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)$
- $\hat{\mathsf{s}}^\star_h \mid \tilde{\mathsf{s}}^\star_h \sim \mathsf{W}^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h)$ for the deconvolution kernel defined in
- $\mathsf{a}_h \sim \pi^\star_h(\hat{\mathsf{s}}^\star_h)$ and $\mathsf{a}'_h \sim \hat{\pi}(\tilde{\mathsf{s}}^\star_h)$

For couplings $\hat{\mu}$ of this form, and for $r > 0$, then, we can bound (M.3) via

$$\begin{aligned}
&\inf_{\hat{\mu}} \mathbb{P}_{\hat{\mu}}[\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \mathsf{s}^\star_h) \le \alpha] \\
&= \inf_{\hat{\mu}} \mathbb{E}_{\hat{\mu}} \mathbf{I}\{\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \mathsf{s}^\star_h) \le \alpha\} \\
&= \inf_{\hat{\mu}} \mathbb{E}_{\hat{\mu}} \left[\mathbf{I}\left\{\sup_{\hat{\mathsf{s}}:\mathsf{d}_{\mathrm{IPS}}(\mathsf{s},\hat{\mathsf{s}}^\star_h)\le 2r} \mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \hat{\mathsf{s}}) > \alpha\right\} + \mathbf{I}\{\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}^\star_h, \hat{\mathsf{s}}^\star_h) > 2r\}\right].
\end{aligned}$$

Because for any $\hat{\mu}$, $\hat{\mathsf{s}}^\star_h \mid \tilde{\mathsf{s}}^\star_h \sim \mathsf{W}^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h)$, we see that the joint distribution $\mathsf{s}^\star_h, \hat{\mathsf{s}}^\star_h$ is independent of the coupling $\hat{\mu}$ and follows the replica distribution: $\hat{\mathsf{s}}^\star_h \mid \mathsf{s}^\star_h \sim \mathsf{W}_{\mathrm{rep},h}(\mathsf{s}^\star_h)$. Consequently, by the Bayesian concentration lemma Lemma H.5, the expected value of the term $\mathbf{I}\{\mathsf{d}_{\mathrm{IPS}}(\mathsf{s}^\star_h, \hat{\mathsf{s}}^\star_h) > 2r\}$ is at most $2p_r$. Hence,

$$\mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}\mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\mathcal{S},\alpha}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h) \mid \mathsf{s}^\star_h) \le 2p_r + \inf_{\hat{\mu}} \mathbb{E}_{\hat{\mu}} \left[\mathbf{I}\left\{\sup_{\hat{\mathsf{s}}:\mathsf{d}_{\mathrm{IPS}}(\mathsf{s},\hat{\mathsf{s}}^\star_h)\le 2r} \mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \hat{\mathsf{s}}) > \alpha\right\}\right]$$

Again, marginalizing over $\mathsf{s}^\star_h$ and using the form of the conditions, the right hand side of the above

$$\mathbb{E}_{\mathsf{s}^\star_h \sim \mathsf{P}^\star_h}\mathbb{E}_{\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\mathsf{s}^\star_h)} \mathsf{d}_{\mathrm{os},\mathcal{S},\alpha}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\mathrm{dec},h}(\tilde{\mathsf{s}}^\star_h) \mid \mathsf{s}^\star_h) \le 2p_r + \inf_{\hat{\mu}} \mathbb{P}_{\hat{\mu}} \left[\sup_{\hat{\mathsf{s}}:\mathsf{d}_{\mathrm{IPS}}(\mathsf{s},\hat{\mathsf{s}}^\star_h)\le 2r} \mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \hat{\mathsf{s}}) > \alpha\right] \tag{M.4}$$

Next, we recall the function $\psi(u) := 2\beta(2\gamma(u), 0)$, instantiate $\alpha = \psi(\varepsilon)$, $\alpha' = \alpha + \beta(2r, 0)$, $r_{\mathrm{IPS}} = 2r$, and set $\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \mathsf{s})$ to be

$$\mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \mathsf{s}) = \psi(\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a}, \mathsf{a}' \mid \tilde{\mathbf{x}}_{\tau_{\mathrm{chunk}}})) \cdot \mathcal{I}_\infty \{\mathsf{d}_{\mathrm{loc},\alpha}(\mathsf{a}, \mathsf{a}' \mid \tilde{\mathbf{x}}_{\tau_{\mathrm{chunk}}}) \le \varepsilon_0\},$$

Using that $\mathsf{d}_{\mathrm{IPS}}$ measures the Euclidean distance between the last control state of composite-state, we

$$\begin{aligned}
\mathbb{P}_{\hat{\mu}} \left[\sup_{\mathsf{s}:\mathsf{d}_{\mathrm{IPS}}(\mathsf{s},\hat{\mathsf{s}}^\star_h)\le 2r} \mathsf{d}_{\mathcal{A};\mathcal{S}}(\mathsf{a}'_h, \mathsf{a}_h \mid \mathsf{s}) > \alpha\right] &= \mathbb{P}_{\hat{\mu}} \left[\sup_{\mathsf{s}:\mathsf{d}_{\mathrm{IPS}}(\hat{\mathsf{s}},\hat{\mathsf{s}}^\star_h)\le 2r} \mathsf{d}_{\mathrm{loc},\alpha'}(\mathsf{a}_h, \mathsf{a}'_h \mid \hat{\mathbf{x}}_{t_h}) > \varepsilon\right] \\
&= \mathbb{P}_{\hat{\mu}} \left[\sup_{\hat{\mathbf{x}}_{t_h}:\|\mathbf{x}_{t_h} - \hat{\mathbf{x}}_{t_h}\|\le 2r} \mathsf{d}_{\mathrm{loc},\alpha'}(\mathsf{a}_h, \mathsf{a}'_h \mid \hat{\mathbf{x}}_{t_h}) > \varepsilon\right] \\
&= \mathbb{P}_{\hat{\mu}} \left[\sup_{\hat{\mathbf{x}}_{t_h}:\|\mathbf{x}_{t_h} - \hat{\mathbf{x}}_{t_h}\|\le 2r} \max_{0\le i<\tau_{\mathrm{chunk}}} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\|\le\alpha'} \|(\kappa_{t_h+i} - \kappa'_{t_h+i})(\hat{\mathbf{x}}_{t_h+i})\| > \varepsilon\right] \\
&\le \mathbb{P}_{\hat{\mu}} \left[\sup_{\hat{\mathbf{x}}_{t_h}:\|\mathbf{x}_{t_h} - \hat{\mathbf{x}}_{t_h}\|\le 2r} \max_{0\le i<\tau_{\mathrm{chunk}}} \sup_{\delta\mathbf{x}:\|\delta\mathbf{x}\|\le\alpha''} \|(\kappa_{t_h+i} - \kappa'_{t_h+i})(\mathbf{x}_{t_h+i})\| > \varepsilon\right] \\
&\qquad \text{(the inequality just replaces } \alpha' \text{ with } \alpha''\text{)} \\
&= \mathbb{P}_{\hat{\mu}} \left[\sup_{\hat{\mathbf{x}}_{t_h}:\|\mathbf{x}_{t_h} - \hat{\mathbf{x}}_{t_h}\|\le 2r} \mathsf{d}_{\mathrm{loc},\alpha''}(\mathsf{a}_h, \mathsf{a}'_h \mid \mathbf{x}_{t_h}) > \varepsilon\right],
\end{aligned} \tag{M.5}$$

$\hat{\mathbf{x}}_{t_h}$ is the first state in $\hat{\mathsf{s}}$, and $\mathbf{x}_{t_h}$ the first state in $\hat{\mathsf{s}}^\star_h$, $\hat{\mathbf{x}}_{t_h:t_h+\tau_{\text{chunk}}-1} = \mathrm{rollout}(\mathsf{a}_h; \hat{\mathbf{x}}_{t_h})$, $\hat{\mathbf{x}}^\star_{t_h:t_h+\tau_{\text{chunk}}-1} = \mathrm{rollout}(\mathsf{a}_h; \hat{\mathbf{x}}^\star_{t_h})$, and finally,

$$\alpha'' := \underbrace{\alpha'}_{=\alpha+\beta(r_{\text{IPS}},0)} + \Delta, \quad \Delta := \sup_{\hat{\mathbf{x}}_{t_h}:\|\mathbf{x}_{t_h}-\hat{\mathbf{x}}_{t_h}\|\le 2r} \sup_{0\le i<\tau_{\text{chunk}}} \|\hat{\mathbf{x}}_{t_h:t_h+\tau_{\text{chunk}}-1} - \mathbf{x}_{t_h:t_h+\tau_{\text{chunk}}-1}\|. \tag{M.6}$$

Now, we see that $\hat{\mu}$ ranges all couplinigs of the form

- $\hat{\mathsf{s}}^\star_h \sim \mathsf{P}^\star_h$ and $\tilde{\mathsf{s}}^\star_h \sim \mathsf{W}_\sigma(\hat{\mathsf{s}}^\star_h)$ (by inverting the deconvolution)
- $\mathsf{a}'_h \sim \tilde{\mathsf{s}}^\star_h$ and $\mathsf{a}_h \sim \pi^\star_h(\hat{\mathsf{s}}^\star_h)$,

which we can see (under our instantiation of the composite MDP under Appendices D and J) is equivalently to $\hat{\mu}$ ranging over all couplings in $\overline{\mathscr{C}}_{\sigma,h}(\hat{\pi})$. Hence, by Assumption 3.1b (i.e. t-ISS of a_h at a_h), we can bound Δ in (M.6) (using $r \le c_\xi/4$) by $\Delta \le \beta(2r,0)$. Hence, we can bound $\alpha'' \le \alpha + 2\beta(2r,0)$, and thus we conclude (from (M.5) and (M.4)) that

$$\begin{aligned}\mathbb{E}_{\mathsf{s}^\star_h\sim\mathsf{P}^\star_h}\mathbb{E}_{\tilde{\mathsf{s}}^\star_h\sim\mathsf{W}_\sigma(\mathsf{s}^\star_h)}\mathsf{d}_{\text{os},\mathcal{S},\alpha}(\hat{\pi}_h(\tilde{\mathsf{s}}^\star_h) \parallel \pi^\star_{\text{dec},h}(\tilde{\mathsf{s}}^\star_h) \mid \mathsf{s}^\star_h) &\le 2p_r + \inf_{\hat{\mu}\in\overline{\mathscr{C}}_{\sigma,h}(\hat{\pi})} \mathbb{P}_{\hat{\mu}}\left[\mathsf{d}_{\text{loc},\alpha''}(\mathsf{a}_h, \mathsf{a}'_h \mid \mathrm{rollout}(\mathsf{a}_h;\mathbf{x}_{t_h}) > \varepsilon\right] \\ &= 2p_r + \Delta_{\text{ISS},\sigma,h}\left(\hat{\pi};\varepsilon,\alpha+2\beta(2r,0)\right),\end{aligned}$$

where the last equality is by definition of $\Delta_{\text{ISS},\sigma,h}\left(\hat{\pi};\varepsilon,\alpha+2\beta(2r,0)\right)$. Consequently, from Theorem 15, for any policy $\hat{\pi}$, both $\Gamma_{\text{joint},\mathcal{S},\alpha}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star_{\text{rep}})$ and $\Gamma_{\text{marg},\mathcal{S},\alpha+\gamma_{\text{IPS},\mathcal{S}}(2r)}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star)$ are upper bounded by

$$H\left(4p_r + 3\gamma_\sigma(\max\{\alpha, \gamma_{\text{IPS},\text{TVC}}(2r)\})\right) + \sum_{h=1}^{H} \Delta_{\text{ISS},\sigma,h}\left(\hat{\pi};\varepsilon,\alpha+2\beta(2r,0)\right).$$

Subsituting in $\gamma_{\text{IPS},\text{TVC}}(r) = \beta(r,\tau_{\text{chunk}}-\tau_{\text{obs}}), \quad \gamma_{\text{IPS},\mathcal{S}}(r) = \beta(r,0)$, we conclude that

$$\begin{aligned}&\Gamma_{\text{joint},\mathcal{S},\alpha}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star_{\text{rep}}) \vee \Gamma_{\text{marg},\mathcal{S},\alpha+\beta(2r,\tau_{\text{chunk}}-\tau_{\text{obs}})}(\hat{\pi}\circ\mathsf{W}_\sigma \parallel \pi^\star) \\ &\le H\left(4p_r + 3\gamma_\sigma(\max\{\alpha, \beta(2r,0)\})\right) + \sum_{h=1}^{H} \Delta_{\text{ISS},\sigma,h}\left(\hat{\pi};\varepsilon,\alpha+2\beta(2r,0)\right).\end{aligned}$$

Substituting in $\gamma_{\text{IPS},\text{TVC}}(r) \le \beta(r,\tau_{\text{chunk}}-\tau_{\text{obs}})$, $\gamma_{\text{IPS},\mathcal{S}}(r) \le \beta(r,0)$ from Lemma M.2, as well as $\alpha = \psi(\varepsilon) = 2\beta(2\gamma(\varepsilon),0)$ concludes. $\square$

N Extensions and Further Results

N.1 Removing the necessity for minimal chunk length via stronger synthesis oracle

Theorem 16. Support we replace Assumption 3.1 in Theorem 1 with the assumption that our trajectory oracle produces *entire sequences of gains* $\kappa_{1:T}$ which satisfy time-varying incremental stability (Definition 3.1) on the whole trajectory. Then,

- The conclusion of Theorem 1 holds
- we no longer need the condition $\tau_{\text{chunk}} \ge c_3$; taking $\tau_{\text{chunk}} = 1$ suffices.
- The constants c_1, c_2 depend only on c_γ and $\bar{c}_\gamma$. That is, c_ξ and terms associated with β can be vaucuously large.

Analogoues, if we replace Assumption 3.1b in Theorem 6 with the assumption that $\kappa_{1:T}$ satisfies the time-varying incremental stability condition, then

- The conclusion of Theorem 6 holds
- We no longer need the condition $\tau_{\text{chunk}} \ge c_3$; taking $\tau_{\text{chunk}} = 1$ suffices. Moreover, we can replace the condition (E.1) of ε with the simpler condition $\varepsilon \le c_\gamma$.

- Lastly, one can replace $\varepsilon_1 = 2\beta(2\gamma(\varepsilon), 0)$ in (E.2) with the term $\varepsilon_1 = \gamma(\varepsilon)$.

The proof of Theorem 16 follows by replacing Lemma K.2 with the following simpler lemma that recapitulates Pfrommer et al. [50, Proposition 3.1], and propogating the argument through the proof.

Lemma N.1. Consider two consistent trajectories $(\mathbf{x}_{1:T+1}, \mathbf{u}_{1:T})$ and $(\mathbf{x}'_{1:T+1}, \mathbf{u}'_{1:T})$, as well as sequences of primitive controller $\kappa_{1:T}, \kappa'_{1:T}$, such that $\mathbf{x}_1 = \mathbf{x}'_1$, and $\mathbf{u}_t = \kappa_t(\mathbf{x}_t)$, $\mathbf{u}'_t = \kappa'_t(\mathbf{x}'_t)$. Suppose that

$$\max_t \sup_{\mathbf{x}:\|\mathbf{x}-\mathbf{x}_t\| \le \gamma(\varepsilon)} \|\kappa_t(\mathbf{x}) - \kappa'_t(\mathbf{x}_t)\| \le \varepsilon.$$

Then, $\max_t \|\mathbf{u}_t - \mathbf{u}'_t\| \le \varepsilon$ and $\max_t \|\mathbf{x}_t - \mathbf{x}'_t\| \le \gamma(\varepsilon)$

N.2 Noisy Dynamics

We can directly extend our imitation guarantees in the composite MDP to settings with noise:

$$\mathsf{s}_{h+1} \sim F_h^{\mathrm{noise}}(\mathsf{s}_h, \mathsf{a}_h, \mathsf{w}_h), \quad \mathsf{w}_h \sim \mathsf{P}_{\mathrm{noise},h}, \tag{N.1}$$

where the noises are idependent of states and of each other. Indeed, (N.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 D.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.

N.3 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:

$$\mathsf{s}^\star_{h+1} \sim F_h^{\mathsf{adv}}(\mathsf{s}^\star_h, \mathsf{a}^\star_h, \mathsf{w}_h, 0), \quad \mathsf{w}_h \sim \mathsf{P}_{\mathrm{noise},h}, \quad \mathsf{a}^\star_h \sim \pi^\star_h(\mathsf{s}^\star_h), \quad \mathsf{s}^\star_1 \sim \mathsf{P}_{\mathrm{init}}.$$

To reduce the composite-MDP in Appendix D, 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 [50]).

N.4 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.2 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 $\widetilde{\mathbf{x}} = \mathbf{x} + \mathbf{w}$. Denote the deconvolution measure of $\mathbf{x}$ given $\widetilde{\mathbf{x}}$ as $p(\cdot|\widetilde{\mathbf{x}})$. We show that this measure is continuous in TV.

Proposition N.2. 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 F.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 N.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 $\mathbf{x} = \mathsf{o}_h$ and consider Q to be a Gaussian (for example) kernel. In this case, we see that the deconvolution policy of Definition C.1 is automatically TVC.

To prove Proposition N.2, we begin with the following lemma:

Lemma N.3. 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. □

We will now make use of the theory of Dini derivatives ([25]) to prove a bound on total variation.

Lemma N.4. For fixed $\widetilde{\mathbf{x}}, \widetilde{\mathbf{x}}'$ and $0 \leq t \leq 1$, 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

$$\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}')) \leq \int_0^1 D^+ \operatorname{TV}\left(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)\right) dt. \tag{N.2}$$

Proof. We compute:

$$\begin{aligned}
2\left|\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))\right| &= \left|\int |p(\mathbf{x}|\widetilde{\mathbf{x}}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h})| - |p(\mathbf{x}|\widetilde{\mathbf{x}}) - p(\widetilde{\mathbf{x}}_t)|\, d\mu(\mathbf{x})\right| \\
&\leq \int |p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t)|\, d\mu(\mathbf{x}). \qquad \text{(N.3)}
\end{aligned}$$

Observe that by the assumption on Q and Lemma N.3, $p(\mathbf{x}|\widetilde{\mathbf{x}}_t)$ 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 and Thomson [25, Theorem 10], (N.2) holds. □

We now bound the Dini derivatives:

Lemma N.5. Let $\widetilde{\mathbf{x}}, \widetilde{\mathbf{x}}' \in \mathbb{R}^d$ such that for all $t \in [0,1]$it holds that

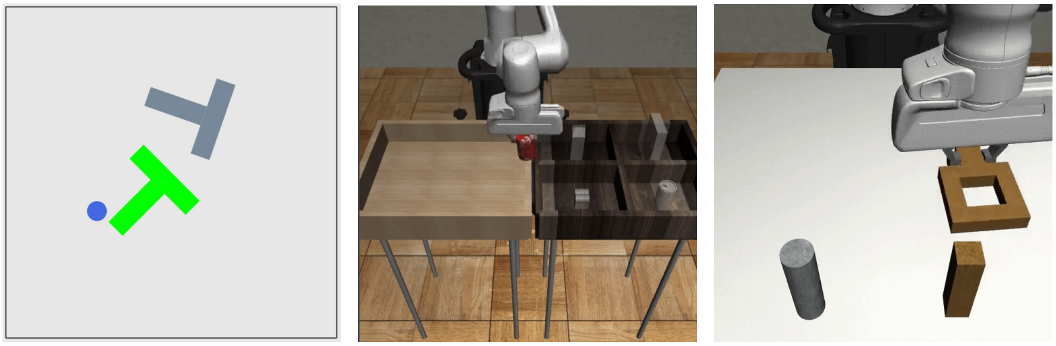
$$\sup_{\mathbf{x}\in\operatorname{supp} p} \left|\frac{d^2}{dt^2}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right)\right| = C < \infty,$$

where the derivative is applied on $\widetilde{\mathbf{x}}_t$. If the assumptions of Lemmas N.3 and N.5 hold, then

$$D^+ \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)) \leq ||\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}'|| \cdot \sup_{\substack{\mathbf{x}\in\operatorname{supp} p \\ t\in[0,1]}} ||\nabla \log Q(\widetilde{\mathbf{x}}_t - x)||.$$

Proof. By definition,

$$D^+ \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t)) = \limsup_{h\downarrow 0} \frac{\operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \operatorname{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))}{h}.$$



(a) PushT Enviroment [19]. 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 [42]. 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 [42]. The grasper must pick up the square nut (the position of which is randomized) and place it over the square peg.

Figure 11: Environment Visualizations.

Fix some t and some small h. By (N.3), it holds that

$$|\mathrm{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_{t+h})) - \mathrm{TV}(p(\cdot|\widetilde{\mathbf{x}}), p(\cdot|\widetilde{\mathbf{x}}_t))| \leq \frac{1}{2} \cdot \int |p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t)|\, d\mu(\mathbf{x}).$$

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

$$\frac{d}{dt}\left(p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\right) = \langle \widetilde{\mathbf{x}}' - \widetilde{\mathbf{x}}, \nabla p(\mathbf{x}|\widetilde{\mathbf{x}}_t)\rangle\,,$$

and thus,

$$|p(\mathbf{x}|\widetilde{\mathbf{x}}_{t+h}) - p(\mathbf{x}|\widetilde{\mathbf{x}}_t)| \leq h \cdot ||\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}'|| \cdot ||\nabla p(\mathbf{x}|\widetilde{\mathbf{x}}_t)|| + h^2 C$$

Now, applying Lemma N.3 and plugging into the previous computation concludes the proof. □

We are finally ready to state and prove our main result:

Proof of Proposition N.2. 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 N.5 holds. Applying Lemma N.4 then concludes the proof. □

O Experiment Details

O.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. [19] as well as the robomimic demonstration dataset Mandlekar et al. [42]. [18]

[18] 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.

Compute. We ran all experiments using 4 Nvidia V100 GPUs on an internal cluster node. For each environment running all experiments depicted in Figure 6 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.

O.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 O for visualizations of the environments.

PushT. The PushT environment introduced in [19] 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 agent's position towards a desired target position at each point in time via acceleration control. Similar to Chi et al. [19], 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 [42] 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 [42], 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.

2D Quadcopter. The 2D quadcopter system is described by the state vector: $(x, z, \phi, \dot{x}, \dot{z}, \dot{\phi})$, with input $u = (u_1, u_2)$, and dynamics:

$$\begin{aligned}
\ddot{x} &= -u_1 \sin(\phi)/m, \\
\ddot{z} &= u_1 \cos(\phi)/m - g, \\
\ddot{\phi} &= u_2/I_{xx}.
\end{aligned}$$

The specific constants we use are $m = 0.8$, $g = 9.8$, and $I_{xx} = 0.5$. We integrate these dynamics using forward Euler with step size $\tau = 0.01$. The task is to move the quadcopter to the origin state. The cost function we used for the MPC expert is:

$$c((x, z, \phi, \dot{x}, \dot{z}, \dot{\phi}), (u_1, u_2)) = x^2 + z^2 + \phi^2 + \dot{x}^2 + \dot{z}^2 + \dot{\phi}^2 + 0.5(u_1 - mg)^2 + 0.1u_2^2.$$

We constructed a per-timestep reward function using this cost function:

$$r((x, z, \phi, \dot{x}, \dot{z}, \dot{\phi}), (u_1, u_2))) = \exp(-c((x, z, \phi, \dot{x}, \dot{z}, \dot{\phi}), (u_1, u_2))),$$

such that the MPC cost minimization corresponds to maximizing the reward used to benchmark the trained models.

O.3 Gain Synthesis

For the quadcopter gain-diffusion experiments, we synthesize stabilizing gains for each $(\overline{x}_t, \overline{u}_t)$ pair in our training data by analytically differentiating the dynamics $x_{t+1} = f(x_t, u_t)$ given in O.2 at $\overline{x}_t, \overline{u}_t$ and applying infinite-horizon LQR to the linearized system $x_{t+1} = A(x_t - \overline{x}_t) + B(u_t - \overline{u}_t)$ where $A = \partial_x f(\overline{x}_t, \overline{u}_t), B = \partial_u f(\overline{x}_t, \overline{u}_t)$. In particular, we solve the discrete time algebraic Ricatti equation:

$$P = A^\top PA - (A^\top PB)(R + B^\top PB)^{-1} + Q,$$

where for simplicity we used identity matrices for R, Q. Using P we computed the gains:

$$K = -(R + B^\top PB)^{-1} B^\top PA.$$

Since the timesteps of the simulator are small, we experimentally find that this is sufficient in order to stabilize the system over the diffused chunks and produces significantly less variance in the gains than performing time-varying discrete LQR over the chunks to synthesize the gains.

O.4 DDPM Model and Training Details.

PushT, Can-Pick-and-Place, Square Nut Assembly. For these experiments we use the same 1-D convolutional UNet-style [56] architecture employed by [19], which is in turn adapted from Janner et al. [33]. This principally consists of 3 sets of downsampling 1-dimensional convolution operations using Mish activation functions [46], Group Normalization (with 8 groups) [77], 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) [49], with the timestep encoded using sin-positional encoding into a 64 dimensional vector.

During training and evaluation we utilize a squared cosine noise schedule [47] with 100 timesteps across all experiments. 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.

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 [19].

2D Quadcopter. For the 2D quadcopter experiments, we used a 5 layer MLP with hidden feature dimensions of $128, 128, 64$, and 64 for all experiments, with sine-positional encoding of dimension 64 and FiLM to condition on the diffusion timestep and observation chunk. We use the same optimizer setup as the PushT, Pick-and-Place, and Square Assembly experiments with a batch size of 64 and a total of 200 epochs or 20,000 training iterations, whichever is larger.

We predict sequences of 8 control inputs, conditioned on two previous observations, where the 2nd control input in this sequence corresponds to the current timestep. For gain diffusion experiments, this includes a sequence of 8 control inputs, reference states, and gains. Similar to our other experiments, the 2nd through 5th generated actions of this sequence are emitted.

Augmentation Procedure. 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.