The Power of Learned Locally Linear Models for Nonlinear Policy Optimization

Daniel Pfrommer [*1] **Max Simchowitz** [*1] **Tyler Westenbroek** [2] **Nikolai Matni** [3] **Stephen Tu** [4]

Abstract

A common pipeline in learning-based control is to iteratively estimate a model of system dynamics, and apply a trajectory optimization algorithm - e.g. `iLQR` - on the learned model to minimize a target cost. This paper conducts a rigorous analysis of a simplified variant of this strategy for general nonlinear systems. We analyze an algorithm which iterates between estimating local linear models of nonlinear system dynamics and performing `iLQR`-like policy updates. We demonstrate that this algorithm attains sample complexity polynomial in relevant problem parameters, and, by synthesizing locally stabilizing gains, overcomes exponential dependence in problem horizon. Experimental results validate the performance of our algorithm, and compare to natural deep-learning baselines.

1. Introduction

Machine learning methods such as model-based reinforcement learning have lead to a number of breakthroughs in key applications across robotics and control (Kocijan et al., 2004; Tassa et al., 2012; Nagabandi et al., 2019). A popular technique in these domains is learning-based model-predictive control (MPC) (Morari & Lee, 1999; Williams et al., 2017), wherein a model learned from data is used to repeatedly solve online planning problems to control the real system. It has long been understood that solving MPC *exactly*–both with perfectly accurate dynamics and minimization to globally optimality for each planning problem–enjoys numerous beneficial control-theoretic properties (Jadbabaie & Hauser, 2001).

Unfortunately, the above situation is not reflective of practice. For one, most systems of practical interest are *nonlinear*, and therefore exact global recovery of system dynamics suffers from a curse of dimensionality. And second, the nonlinear dynamics render any natural trajectory planning problem nonconvex, making global optimality elusive. In this work, we focus on learning-based trajectory optimization, the "inner-loop" in MPC. We ask *when can we obtain rigorous guarantees about the solutions to nonlinear trajectory optimization under unknown dynamics?*

We take as our point of departure the `iLQR` algorithm (Li & Todorov, 2004). Initially proposed under known dynamics, `iLQR` solves a planning objective by solving an iterative linear control problem around a first-order Taylor expansion (the *Jacobian linearization*) of the dynamics, and second-order Taylor expansion of the control costs. In solving this objective, `iLQR` synthesizes a sequence of locally-stabilizing feedback gains, and each `iLQR`-update can be interpreted as a gradient-step through the closed-loop linearized dynamics in feedback with these gains. This has the dual benefit of proposing a locally stabilizing policy (not just an open-loop trajectory), and of stabilizing the gradients to circumvent exponential blow-up in planning horizon. `iLQR`, and its variants (Todorov & Li, 2005; Williams et al., 2017), are now ubiquitous in robotics and control applications; and, when dynamics are unknown or uncertain, one can simply substitute the exact dynamics model with an estimate (e.g. Levine & Koltun (2013)). In this case, dynamics are typically estimated with neural networks. Thus, Jacobian linearizations can be computed by automated differentiation (AutoDiff) through the learned model.

Contributions. We propose and analyze an alternative to the aforementioned approach of first learning a deep neural model of dynamics, and then performing AutoDiff to conduct the `iLQR` update. We consider a simplified setting with fixed initial starting condition. Our algorithm maintains a *policy*, specified by an open-loop input sequence and a sequence of stabilizing gains, and loops two steps: **(a)** it learns local linear model of the closed-loop linearized dynamics (in feedback with these gains), which we use to perform a gradient update; **(b)** it re-estimates a linear model after the gradient step, and synthesizes a new set of set gains from this new model. In contrast to past approaches,

[*]Equal contribution [1]Massachusetts Institute of Technology [2]University of Texas, Austin [3]University of Pennsylvania [4]Google Brain. Correspondence to: Daniel Pfrommer <dpfrom@mit.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

our algorithm *only ever estimates linear models of system dynamics.*

For our analysis, we treat the underlying system dynamics as continuous and policy as discrete; this reflects real physical systems, is representative of discrete-time simulated environments which update on smaller timescales than learned policies, and renders explicit the effect of discretization size on sample complexity. We consider an interaction model where we query an oracle for trajectories corrupted with measurement (but not process) noise. Our approach enjoys the following theoretical properties. **1.** Using a number of iterations and oracle queries *polynomial* in relevent problem parameters and tolerance ϵ, it computes a policy π whose input sequence is an ϵ-first order stationary point for the `iLQR` approximation of the planning objective (i.e., the gradient through the closed-loop linearized dynamics has norm $\leq \epsilon$). Importantly, learning the linearized model at each iteration obviates the need for global dynamics models, allowing for sample complexity polynomial in dimension.

2. We show that contribution **1** implies convergence to a local-optimality criterion we call an ϵ-approximate *Jacobian Stationary Point* (ϵ-`JSP`); this roughly equates to the open-loop trajectory under π having cost within ϵ-*globally optimal* for the linearized dynamics about its trajectory.

`JSP`s are purely a property of the open-loop inputs, allowing comparison of the quality of the open-loop plan with differing gains. Moreover, the results of Westenbroek et al. (2021) show that an approximate `JSP`s for certain planning objective enjoy favorable *global properties*, despite (as we show) being computable from (local) gradient-based search (see Appendix B.2 for elaboration).

Experimental Findings. We validate our algorithms on standard models of the quadrotor and inverted pendulum, finding an improved performance as iteration number increases, and that the synthesized gains prescribed by `iLQR` yield improved performance over vanilla gradient updates.

Proof Techniques. Central to our analysis are novel perturbation bounds for controlled nonlinear differential equations. Prior results primarily focus on the open-loop setting (Polak, 2012, Theorem 5.6.9), and implicitly hide an exponential dependence on the time horizon for open-loop unstable dynamics. We provide what is to the best of our knowledge the first analysis which demonstrates that local feedback can overcome this pathology. Specifically, we show that if the feedback gains stabilize the Jacobian-linearized dynamics, then (a) the Taylor-remainder of the first-order approximation of the dynamics does *not* scale exponentially on problem horizon (Proposition 4.3), and (b) small perturbations to the nominal input sequence preserve closed-loop stability of the linearized dynamics. These findings are detailed in Appendix A.6, and enable us to bootstrap the many recent advances in statistical learning for linear systems to our nonlinear setting.

1.1. Related Work

`iLQR` (Li & Todorov, 2004) is a more computationally expedient variant of differential dynamic programming (`DPP`) (Jacobson & Mayne, 1970); numerous variants exist, notably `iLQG` (Todorov & Li, 2005) and `iLQR` (Li & Todorov, 2004), which better address problem stochasticity. `iLQR` is a predominant approach for the "inner loop" trajectory optimization step in MPC, with applications in robotics (Tassa et al., 2012), quadrotors (Torrente et al., 2021), and autonomous racing (Kabzan et al., 2019).

A considerable literature has combined `iLQR` with learned dynamics models; here, the Jacobian linearization matrices are typically derived through automated differentiation (Levine & Koltun, 2013; Levine & Abbeel, 2014; Koller et al., 2018), though local kernel least squares regression has also been studied (Rosolia & Borrelli, 2019; Papadimitriou et al., 2020). In these works, the dynamics models are refined/re-estimated as the policy is optimized; thus, these approaches are one instantiation of the broader iterative learning control (ILC) paradigm (Arimoto et al., 1984); other instantiations of ILC include (Kocijan et al., 2004; Dai et al., 2021; Aswani et al., 2013; Bechtle et al., 2020).

Recent years have seen multiple rigorous guarantees for learning system identification and control (Dean et al., 2017; Simchowitz et al., 2018; Oymak & Ozay, 2019; Agarwal et al., 2019; Simchowitz & Foster, 2020), though a general theory of learning for nonlinear control remains elusive. Recent progress includes nonlinear imitation learning (Pfrommer et al., 2022), learning systems with known nonlinearities in the dynamics (Sattar & Oymak, 2022; Foster et al., 2020; Mania et al., 2020) or perception model (Mhammedi et al., 2020; Dean & Recht, 2021).

Lastly, there has been recent theoretical attention given to the study of first-order trajectory optimization methods. Roulet et al. (2019) perform an extension theoretical study of the convergence properties of `iLQR`, `iLQG`, and `DPP` with *exact* dynamics models, and corroborate their findings experimentally. Westenbroek et al. (2021) show further that for certain classes of nonlinear systems, all ϵ-first order stationary points of a suitable trajectory optimization objective induce trajectories which converge exponentially to desired system equilbria. In some cases, there may be multiple spurious local minima, each of which is nevertheless exponentially stabilizing. Examining the proof (Westenbroek et al., 2021) shows the result holds more generally for all ϵ-`JSP`s, and therefore we use their work justify the `JSP` criterion proposed in this paper.

2. Setting

We consider a continuous-time nonlinear control system with state $\mathbf{x}(t) \in \mathbb{R}^{d_\mathbf{x}}$, input $\mathbf{u}(t) \in \mathbb{R}^{d_\mathbf{u}}$ with finite horizon $T > 0$, and fixed initial condition $\xi_{\text{init}} \in \mathbb{R}^{d_\mathbf{x}}$. We denote the space of bounded input signals $\mathcal{U} := \{\mathbf{u}(\cdot) : [0,T] \to \mathbb{R}^{d_\mathbf{u}} : \sup_{t\in[0,T]} \|\mathbf{u}(t)\| < \infty\}$. We endow $\mathcal{U}$ with an inner product $\langle \mathbf{u}(\cdot), \mathbf{u}'(\cdot)\rangle_{\mathcal{L}_2(\mathcal{U})} := \int_0^T \langle \mathbf{u}(s), \mathbf{u}'(s)\rangle \mathrm{d}s$, where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product, which induces a norm $\|\mathbf{u}(\cdot)\|^2_{\mathcal{L}_2(\mathcal{U})} := \langle \mathbf{u}(\cdot), \mathbf{u}(\cdot)\rangle_{\mathcal{L}_2(\mathcal{U})}$. For $\mathbf{u} \in \mathcal{U}$, the open-loop dynamics are governed by the ordinary differential equation (ODE)

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t \mid \mathbf{u}) = f_{\text{dyn}}(\mathbf{x}(t \mid \mathbf{u}), \mathbf{u}(t)), \;\; \mathbf{x}(0 \mid \mathbf{u}) = \xi_{\text{init}},$$

where $f_{\text{dyn}} : \mathbb{R}^{d_\mathbf{x}} \times \mathbb{R}^{d_\mathbf{u}} \to \mathbb{R}^{d_\mathbf{x}}$ a $\mathcal{C}^2$ map. Given a terminal cost $V(\cdot) : \mathbb{R}^{d_\mathbf{x}} \to \mathbb{R}$ and running $Q(\cdot,\cdot,\cdot) : \mathbb{R}^{d_\mathbf{x}} \times \mathbb{R}^{d_\mathbf{u}} \times [0,T] \to \mathbb{R}$, we optimize the control objective

$$\mathcal{J}_T(\mathbf{u}) := V(\mathbf{x}(T \mid \mathbf{u})) + \textstyle\int_{t=0}^T Q(\mathbf{x}(t \mid \mathbf{u}), \mathbf{u}(t), t)\mathrm{d}t.$$

We make the common assumption that the costs are strongly $\mathcal{C}^2$, and that $\mathcal{Q}$ is strongly convex:

Assumption 2.1. For all $t \in [0,T]$, $V(\cdot)$ and $Q(\cdot,\cdot,t)$ are twice-continuously differentiable ($\mathcal{C}^2$), and $x \mapsto V(x)$ and $(x,u) \mapsto Q(x,u,t) - \frac{\alpha}{2}(\|x\|^2 + \|u\|^2)$ are convex.

Given a continuously differentiable function $\mathcal{F} : \mathcal{U} \to \mathbb{R}^n$ and perturbation $\delta\mathbf{u} \in \mathcal{U}$, we define its *directional derivative* $\mathrm{D}\mathcal{F}(\mathbf{u})[\delta\mathbf{u}] := \lim_{\eta\to 0} \eta^{-1}(\mathcal{F}(\mathbf{u}+\eta\delta\mathbf{u}) - \mathcal{F}(\mathbf{u}))$. The *gradient* $\nabla\mathcal{F}(\mathbf{u}) \in \mathcal{U}$ is the (almost-everywhere) unique element of $\mathcal{U}$ such that $\forall \delta\mathbf{u} \in \mathcal{U}$, $\int_0^T \nabla\mathcal{F}(\mathbf{u})(t)\delta\mathbf{u}(t)\mathrm{d}t = \mathrm{D}\mathcal{F}(\mathbf{u})[\delta\mathbf{u}]$. We denote the gradients of $\mathbf{u} \mapsto \mathbf{x}(t \mid \mathbf{u})$ as $\nabla_\mathbf{u}\,\mathbf{x}(t \mid \mathbf{u})$, and of $\mathbf{u} \mapsto \mathcal{J}_T(\mathbf{u})$ as $\nabla_\mathbf{u}\mathcal{J}_T(\mathbf{u})$.

Discretization and Feedback Policies. Because digital controllers cannot represent continuous open-loop inputs, we compute ϵ-JSPs $\mathbf{u} \in \mathcal{U}$ which are the zero-order holds of discrete-time control sequences. We let $\tau \in (0,T]$ be a discretization size, and set $K = \lfloor T/\tau \rfloor$. Going forward, we denote discrete-time quantities in `colored, bold-seraf font`.

For $k \geq 1$, define $t_k = (k-1)\tau$, and define the intervals $\mathcal{I}_k = [t_k, t_{k+1})$. For $t \in [0,T]$, let $k(t) := \sup\{k : t_k \leq t\}$. We let $\mathbf{U} := (\mathbb{R}^{d_\mathbf{u}})^K$, whose elements are denoted $\vec{\mathtt{u}} = \mathtt{u}_{1:K}$, and let $\mathsf{ct} : \mathbf{U} \to \mathcal{U}$ denote the natural inclusion $\mathsf{ct}(\vec{\mathtt{u}})(t) := \mathtt{u}_{k(t)}$.

Next, to mitigate the curse of horizon, we study *policies* which (a) have discrete-time open-loop inputs and (b) have discrete-time feedback gains to stabilize around the trajectories induced by the nominal inputs. In this work, Π_τ denotes the set of all policies $\pi = (\mathtt{u}^\pi_{1:K}, \mathtt{K}^\pi_{1:K})$ defined by a discrete-time open-loop policy $\mathtt{u}^\pi_{1:K} \in \mathbf{U}$, and a sequence of feedback gains $(\mathtt{K}^\pi_k)_{k\in[K]} \in (\mathbb{R}^{d_\mathbf{x} d_\mathbf{u}})^K$. A policy π induces nomimal dynamics $\mathbf{u}^\pi(\cdot) = \mathsf{ct}(\mathtt{u}^\pi_{1:K})$ and $\mathbf{x}^\pi(t) = \mathbf{x}(t \mid \mathbf{u}^\pi)$; we set $\mathtt{x}^\pi_k = \mathbf{x}^\pi(t_k)$. It also induces the following dynamics by stabilizing around the policy.

Definition 2.1. Given a continuous-time input $\bar{\mathbf{u}} \in \mathcal{U}$, we define the *stabilized trajectory* $\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) := \mathbf{x}(t \mid \tilde{\mathbf{u}}^{\pi,\mathrm{ct}})$, where $\tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) := \bar{\mathbf{u}}(t) + \mathtt{K}^\pi_{k(t)}(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t_{k(t)} \mid \bar{\mathbf{u}}) - \mathtt{x}^\pi_{k(t)})$. This induces a stabilized objective: $\mathcal{J}^\pi_T(\bar{\mathbf{u}}) := V(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}})) + \int_0^T Q(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}), \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathbf{u}), t)\mathrm{d}t$. We define the shorthand $\nabla\mathcal{J}_T(\pi) := \nabla_{\bar{\mathbf{u}}}\mathcal{J}^\pi_T(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^\pi}$

Notice that, while π is specified by *discrete-time* inputs, $\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(\cdot), \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(\cdot)$ are *continuous-time* inputs and trajectories stabilized by π and the gradient $\nabla\mathcal{J}^\pi_T(\cdot)$ is defined over *continuous-time perturbations*.

Optimization Criteria. Due to nonlinear dynamics, the objectives $\mathcal{J}_T, \mathcal{J}^\pi_T$ are nonconvex, so we can only aim for local optimality. Approximate first-order stationary points (FOS) are a natural candidate (Roulet et al., 2019).

Definition 2.2. We say $\mathbf{u}$ is an ϵ-FOS of a function $\mathcal{F} : \mathcal{U} \to \mathbb{R}$ if $\|\nabla_\mathbf{u}\mathcal{F}(\mathbf{u})\|_{\mathcal{L}_2(\mathcal{U})} \leq \epsilon$. We say π is ϵ-*stationary* if $\|\nabla J_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} := \|\nabla_{\bar{\mathbf{u}}}\mathcal{J}^\pi_T(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^\pi}\|_{\mathcal{L}_2(\mathcal{U})} \leq \epsilon$.

Our primary criterion is to compute ϵ-stationary policies π. However, this depends both on the policy inputs $\mathbf{u}^\pi$ (and induced trajectory $\mathbf{x}^\pi$), *as well as* the gains. We therefore propose a secondary optimization criterion which depends only on the policies inputs/trajectory. It might be tempting to hope that $\mathbf{u}^\pi$ is an ϵ-FOS of the original objective $\mathcal{J}_T(\mathbf{u})$. However, when the Jacobian linearized trajectory (Definition 2.3 below) of the dynamics around $(\mathbf{x}^\pi, \mathbf{u}^\pi)$ are unstable, the open-loop gradient $\|\nabla\mathcal{J}_T(\mathbf{u}^\pi)\|_{\mathcal{L}_2(\mathcal{U})}$ can be a factor of e^T larger than the stabilized gradient $\|\nabla\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})}$ despite the fact that, definitionally, $\mathcal{J}_T(\mathbf{u}^\pi) = \mathcal{J}^\pi_T(\mathbf{u}^\pi)$ (see Appendix B.1). We therefore propose an alternative definition in terms of Jacobian-linearized trajectory.

Definition 2.3. Given $\mathbf{u}, \bar{\mathbf{u}} \in \mathcal{U}$, define the Jacobian-linearized (JL) *trajectory* $\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}; \mathbf{u}) = \mathbf{x}(t \mid \mathbf{u}) + \langle \nabla_\mathbf{u}\,\mathbf{x}(t \mid \mathbf{u}), \bar{\mathbf{u}} - \mathbf{u}\rangle$, and *cost* $\mathcal{J}^{\mathrm{jac}}_T(\bar{\mathbf{u}}; \mathbf{u}) := V(\mathbf{x}^{\mathrm{jac}}(T \mid \bar{\mathbf{u}}; \mathbf{u})) + \int_{t=0}^T Q(\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}; \mathbf{u}), \bar{\mathbf{u}}(t), t)\mathrm{d}t$.

In words, the JL trajectory is just the first-order Taylor expansion of the dynamics around an input $\mathbf{u} \in \mathcal{U}$, and the cost is the cost functional applied to those JL dynamics. We propose an optimization criterion which requires that $\mathbf{u}$ is near-*globally* optimal for the JL dynamics around $\mathbf{u}$:

Definition 2.4. We say $\mathbf{u} \in \mathcal{U}$ is an ϵ-Jacobian Stationary Point (JSP) if $\mathcal{J}_T(\mathbf{u}) \leq \inf_{\bar{\mathbf{u}}\in\mathcal{U}} \mathcal{J}^{\mathrm{jac}}_T(\bar{\mathbf{u}}; \mathbf{u}) + \epsilon$.

The consideration of JSPs has three advantages: (1) as noted above, JSPs depend only on a trajectory and not on feedback gains; (2) a JSP is sufficient to ensure that the exponential-stability guarantees derived in Westenbroek

et al. (2021) (and mentioned in the introduction above) hold for certain systems; this provides a link between the local optimality derived in this work and *global* trajectory behavior (see Appendix B.2 for further discussion); (3) despite the potentially exponential-in-horizon gap between gradients of $\mathcal{J}_T$ and $\mathcal{J}_T^\pi$, the following result enables us to compare stationary points of the two objectives in a manner that is *independent* of the horizon T.

Propostion 4.1 (informal). *Suppose π is ϵ-stationary, and τ is sufficiently small. Then, $\mathbf{u}^\pi$ is an ϵ'-JSP of $\mathcal{J}_T$, where* $\epsilon' = \mathcal{O}\left(\epsilon^2/(2\alpha(1+\max_k \|\mathtt{K}_k^\pi\|^2)\right)$.

Oracle Model and Problem Desideratum. In light of the above discussion, we aim to compute a approximately stationary policy, whose open-loop is therefore an approximate JSP for the original objective. To do so, we assume access to an oracle which can perform feedback with respect to gains $\mathtt{K}_k^\pi$.

Definition 2.5 (Oracle Dynamics)**.** Given $\vec{\mathtt{u}} = \mathtt{u}_{1:K} \in \mathbf{U}$, we define the *oracle dynamics* $\mathbf{x}^\pi_{\mathrm{orac}}(t \mid \vec{\mathtt{u}}) := \mathbf{x}(t \mid \mathtt{ct}(\mathtt{u}^\pi_{\mathrm{orac},1:K}(\vec{\mathtt{u}})))$, where we define $\mathtt{u}^\pi_{\mathrm{orac},k}(\vec{\mathtt{u}}) := \mathtt{u}_k + \mathtt{K}_k^\pi \mathbf{x}^\pi_{\mathrm{orac}}(t_k \mid \vec{\mathtt{u}})$, and define $\mathtt{x}^\pi_{\mathrm{orac},k}(\vec{\mathtt{u}}) := \mathbf{x}^\pi_{\mathrm{orac}}(t_k \mid \vec{\mathtt{u}})$.

Oracle 2.1. *We assume access to an oracle* orac *with variance $\sigma^2_{\mathrm{orac}} > 0$, which given any $\pi \in \Pi_\tau$ and $\vec{\mathtt{u}} = \mathtt{u}_{1:K}$, returns, $\mathtt{orac}_{\pi,x}(\vec{\mathtt{u}}) \sim \mathcal{N}(\mathtt{x}^\pi_{\mathrm{orac},1:K+1}(\vec{\mathtt{u}}), \mathbf{I}_{(K+1)d_{\mathtt{x}}}\sigma^2_{\mathrm{orac}})$ and $\mathtt{orac}_{\pi,u}(\vec{\mathtt{u}}) = \mathtt{u}^\pi_{\mathrm{orac},1:K}(\vec{\mathtt{u}})$*

In words, Oracle 2.1 returns entire trajectories by applying feedback along the gains $\mathtt{K}_k^\pi$. The addition of measurement noise is to introduce statistical tradeoffs that prevent near-exact zero-order differentiation; we discuss extensions to process noise in Appendix B.4. Because of this, the oracle trajectory in Definition 2.5 differs from the trajectory dynamics in Definition 2.1 in that the feedback does not subtract off the normal $\mathtt{x}_k^\pi$; thus, the oracle can be implemented without noiseless access to the nominal trajectory. Still, we assume that the feedback applied by the oracle is exact. Having defined our oracle, we specify the following problem desideratum (note below that M is scaled by $1/\tau$ to capture the computational burden of finer discretization).

Desideratum 1. *Given ϵ, ϵ' and unknown dynamical system $f_{\mathrm{dyn}}(\cdot,\cdot)$, compute a policy π for which (a) π is ϵ-stationary, and (b) $\mathbf{u}^\pi$ is an ϵ'-JSP of $\mathcal{J}_T$, using M calls to Oracle 2.1, where M/τ is polynomial in $1/\epsilon$, $1/\epsilon'$, and relevant problem parameters.*

Notation. We let $[j : k] := \{j, j+1, \ldots, k\}$, and $[k] = [1 : k]$. We use standard-bold for continuous-time quantities $(\mathbf{x}, \mathbf{u})$, and bold-serif for discrete (e.g. $\mathtt{u}_k^\pi$). We let $\mathtt{u}^\pi_{j:k} = (\mathtt{u}_j, \mathtt{u}_{j+1}, \ldots, \mathtt{u}_k)$. Given vector $\mathtt{v}$ and matrices $\mathtt{X}$, let $\|\mathtt{v}\|$ and $\|\mathtt{X}\|$ Euclidean and operator norm, respectively; for clarity, we write $\|\mathtt{u}^\pi_{j:k}\|^2_{\ell_2} = \sum_{i=j}^k \|\mathtt{u}_i^\pi\|^2$. As denoted above, $\langle\cdot,\cdot\rangle_{\mathcal{L}_2(\mathcal{U})}$ and $\|\cdot\|_{\mathcal{L}_2(\mathcal{U})}$ denote inner products and norms in $\mathcal{L}_2(\mathcal{U})$. We let $x \vee y := \max\{x, y\}$, and $x \wedge y := \min\{x, y\}$.

3. Algorithm

Our iterative approach is summarized in Algorithm 1 and takes in a time step $\tau > 0$, horizon T, a per iteration sample size N, iteration number n_{iter}, a noise variance σ_w, a gradient step size $\eta > 0$ and a controllability parameter k_0. The algorithm produces a sequence of polices $\pi^{(n)} = (\mathtt{u}^{(n)}_{1:K}, \mathtt{K}^{(n)}_{1:K})$, where $K = \lfloor T/\tau \rfloor$ is the number of time steps per roll-out. Our algorithm uses the primitive EstMarkov$(\pi; N, \sigma_w)$ (Algorithm 2), which makes N calls to the oracle to produce estimates $\hat{\mathtt{x}}_{1:K+1}$ of the nominal state trajectory, and another N calls with randomly-perturbed inputs of perturbation-variance σ_w to produce estimates $(\hat{\Psi}_{j,k})_{k<j}$ of the closed-loop Markov parameters associated to the current policy $\Psi^\pi_{\mathrm{cl},j,k}$, defined in Definition 4.6. We use a method-of-moments estimator for simplicity. At each iteration n, Algorithm 1 calls calls EstMarkov$(\pi; N, \sigma_w)$ first to produce an estimate of the gradient of the closed-loop objective with respect to the current discrete-time nominal inputs. The gradient with respect to the k-th input $\mathtt{u}_k^{(n)}$ is given by:

$$\hat{\nabla}_k^{(n)} = \hat{\Psi}_{K+1,k}^\top V_x(\hat{\mathtt{x}}_{K+1}) + Q_u(\hat{\mathtt{x}}_k, \mathtt{u}_k^\pi, t_k) + \tau \sum_{j=k+1}^K \hat{\Psi}_{j,k}^\top \left(Q_x(\hat{\mathtt{x}}_j, \mathtt{u}_j^\pi, t_j) + (\mathtt{K}_j^\pi)^\top Q_u(\hat{\mathtt{x}}_j, \mathtt{u}_j^\pi, t_j)\right) \tag{3.1}$$

The form of this estimate corresponds to a natural plug-in estimate of the gradient of the discrete-time objective defined in Definition 4.4. We use this gradient in Eq. (3.1) to update the current input; this update is rolled-out in feedback with the current feedback controller to produce the nominal input $\mathtt{u}^{(n+1)}_{1:K}$ for the next iteration (Algorithm 1, Line 5). Finally, we call EstGains (Algorithm 3), which synthesizes gains for the new policy using a Ricatti-type recursion along a second estimate of the linearized dynamics, produced by unrolling the system with the new nominal input and old gains described above. The algorithm then terminates at n_{iter} iterations and chooses the policy with the smallest estimated gradient that was observed.

4. Algorithm Analysis

For simplicity, we assume $K = \lfloor T/\tau \rfloor \in \mathbb{N}$ is integral. In order to state uniform regularity conditions on the dynamics and costs, we fix an *feasible radius* $R_{\mathrm{feas}} > 0$ and restrict to states and inputs bounded thereby.

Definition 4.1. We say $(x, u) \in \mathbb{R}^{d_{\mathtt{x}} \times d_{\mathtt{u}}}$ are *feasible* if $\|x\| \vee \|u\| \le R_{\mathrm{feas}}$. We say a policy π is feasible if $(2\mathbf{x}^\pi(t), 2\mathbf{u}^\pi(t))$ are feasible for all $t \in [0, T]$.

We adopt the following boundedness condition.

Algorithm 1 Trajectory Optimization

Initialize time step $\tau > 0$, horizon $T \geq \tau$, $K \leftarrow \lfloor T/\tau \rfloor$, initial policy $\pi^{(1)}$, sample size N, noise variance σ_w, gradient step size η, controllability parameter k_0, iteration number n_{iter}.
for iterations $n = 1, 2, \ldots, n_{\text{iter}}$ **do**
  $(\hat{\Psi}_{j,k})_{k<j}, \hat{\mathtt{x}}_{1:K+1} = \textsc{EstMarkov}(\pi; N, \sigma_w)$.
  Compute $\hat{\nabla}_k^{(n)}$ in Eq. (3.1)
  Gradient update $\mathtt{u}_{1:K}^{(n+1)} \leftarrow \mathtt{orac}_{\pi^{(n)},u}(\tilde{\mathtt{u}}_{1:K}^{(n)})$, where $\tilde{\mathtt{u}}_k^{(n)} := \mathtt{u}_k^{(n)} - \frac{\eta}{\tau}\hat{\nabla}_k^{(n)} - \mathtt{K}_k^{\pi^{(n)}}\hat{\mathtt{x}}_k$.
  Estimate $\mathtt{K}_{1:K}^{(n+1)} = \textsc{EstGains}(\tilde{\pi}^{(n)}; \sigma_w, N, k_0)$, where $\tilde{\pi}^{(n)} = (\mathbf{u}^{(n+1)}(\cdot), \mathtt{K}_{1:K}^{(n)})$
  Update policy $\pi^{(n+1)} = (\mathtt{u}_{1:K}^{(n+1)}, \mathtt{K}_{1:K}^{(n+1)})$
return $\pi^{(n_{\text{out}})}$, $n_{\text{out}} \in \arg\min_{n \in [n_{\text{iter}}]} \|\hat{\nabla}_k^{(n)}\|$.

Algorithm 2 $\textsc{EstMarkov}(\pi; N, \sigma_w)$

`% estimate nominal trajectory`
for samples $i = 1, 2, \ldots, N$ **do**
  Collect trajectory $\mathtt{x}_{1:K+1}^{(i)} \sim \textbf{TrajOrac}_\pi(\mathtt{u}_{1:K}^{\pi})$.
Average $\hat{\mathtt{x}}_{1:K+1} = \frac{1}{N}\sum_{i=1}^N \mathtt{x}_{1:K+1}^{(i)}$
`% estimate perturbed trajectory`
for samples $i = 1, 2, \ldots, N$ **do**
  Draw $\mathtt{w}_{1:K}^{(i)}$ uniformly from $\sigma_w \cdot (\{-1, 1\}^{d_u})^K$.
  Let $\mathtt{u}_k^{(i)} = \mathtt{u}_k^{\pi} + \mathtt{w}_k^{(i)} - \mathtt{K}_k^{\pi}\hat{\mathtt{x}}_k$, for $k \in [K]$
  Collect trajectory $\mathtt{y}_{1:K+1}^{(i)} \sim \mathtt{orac}_{\pi,x}(\mathtt{u}_{1:K}^{(i)})$.
Estimate transition operators $\hat{\Psi}_{j,k} := \frac{1}{N\sigma_w^2}\sum_{i=1}^N (\mathtt{y}_j^{(i)} - \hat{\mathtt{x}}_j)(\mathtt{w}_k^{(i)})^\top$, $k < j$
return $(\hat{\Psi}_{j,k})_{k<j}, \hat{\mathtt{x}}_{1:K+1}$

Condition 4.1. For all n, the policies $\pi^{(n)}$ and $\tilde{\pi}^{(n)}$ produced by Algorithm 1 are feasible.

If π and $\tilde{\pi}^{(n)}$ produce bounded inputs, and the resulting state trajectories also remain bounded, then Condition 4.1 will hold for $R_{\text{feas}} > 0$ sufficiently large. This is a common assumption in the control literature (see e.g. Jadbabaie & Hauser (2001)), as physical systems, such as those with Lagrangian dynamics, will remain bounded under bounded inputs (see Appendix B.3 for discussion).

Assumption 4.1 (Dynamics regularity)**.** f_{dyn} is $\mathcal{C}^2$, and for all feasible (x, u), the following hold $\|f_{\text{dyn}}(x,u)\| \leq \kappa_f$, $\|\partial_x f_{\text{dyn}}(x,u)\| \vee \|\partial_u f_{\text{dyn}}(x,u)\| \leq L_f$, $\|\nabla^2 f_{\text{dyn}}(x,u)\| \leq M_f$.

Assumption 4.2 (Cost regularity)**.** For all feasible (x, u), the following hold $0 \leq V(x) \vee Q(x,u,t) \leq \kappa_{\text{cost}}$, $\|\partial_x V(x)\| \vee \|\partial_x Q(x,u,t)\| \vee \|\partial_u Q(x,u,t)\| \leq L_{\text{cost}}$, $\|\nabla^2 V(x)\| \vee \|\nabla^2 Q(x,u,t)\| \leq M_{\text{cost}}$.

To take advantage of stabilizing gains, we require two additional assumptions, which are defined in terms of the JL dynamics.

Algorithm 3 $\textsc{EstGains}(\pi; N, \sigma_w, k_0)$

Initialize number of samples N, noise variance σ_w, (discrete) controllability window $k_0 \in \mathbb{N}$
Estimate Markov Parameters $(\hat{\Psi}_{j,k})_{k<j} = \textsc{EstMarkov}(\pi; N, \sigma_w)$
`% Define` $\hat{\mathcal{C}}_{k|j_2,j_1} := [\hat{\Psi}_{k+1,j_2} \mid \hat{\Psi}_{k+1,j_2-1} \mid \ldots \hat{\Psi}_{k+1,j_1}]$
for $k = k_0, k_0+1, \ldots, K$ **do**
  Define $\hat{\mathbf{B}}_k = \hat{\Psi}_{k+1,k}$
  `% Define` $\hat{\mathcal{C}}_{k,\text{in}} := \hat{\mathcal{C}}_{k-1|k-1,k-k_0+1}$, $\hat{\mathcal{C}}_{k,\text{out}} := \hat{\mathcal{C}}_{k|k-1,k-k_0+1}$
  Define $\hat{\mathtt{A}}_k := \hat{\mathcal{C}}_{k,\text{out}}\hat{\mathcal{C}}_{k,\text{in}}^\dagger - \hat{\mathbf{B}}_k\mathtt{K}_k^\pi$
Set $\hat{\mathtt{P}}_{K+1} = \mathbf{I}_{d_x}$.
for $k = K, K-1, \ldots, k_0$ **do**
  $\hat{\mathtt{K}}_k := (\mathbf{I}_{d_u} + \hat{\mathtt{B}}_k^\top\hat{\mathtt{P}}_{k+1}\hat{\mathtt{B}}_k)^{-1}\left(\hat{\mathtt{B}}_k^\top\hat{\mathtt{P}}_{k+1}\hat{\mathtt{A}}_k\right)$.
  $\hat{\mathtt{P}}_k = (\hat{\mathtt{A}}_k + \hat{\mathtt{B}}_k\hat{\mathtt{K}}_k)^\top\hat{\mathtt{P}}_{k+1}(\hat{\mathtt{A}}_k + \hat{\mathtt{B}}_k\hat{\mathtt{K}}_k) + \tau(\mathbf{I}_{d_x} + \hat{\mathtt{K}}_k^\top\hat{\mathtt{K}}_k)$.
Set $\hat{\mathtt{K}}_k = 0$ for $k \leq k_0$.
Return $\hat{\mathtt{K}}_{1:K}$.

Definition 4.2 (Open-Loop Linearized Dynamics)**.** We define the (open-loop) JL dynamic matrices about π as $\mathbf{A}_{\text{ol}}^\pi(t) = \partial_x f_{\text{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t))$ and $\mathbf{B}_{\text{ol}}^\pi(t) = \partial_u f_{\text{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t))$. We define the *open-loop* JL transition function $\mathbf{\Phi}_{\text{ol}}^\pi(s,t)$, defined for $t \geq s$ as the solution to $\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{\Phi}_{\text{ol}}^\pi(s,t) = \mathbf{A}_{\text{ol}}^\pi(s)\mathbf{\Phi}_{\text{ol}}^\pi(s,t)$, with initial condition $\mathbf{\Phi}_{\text{ol}}^\pi(t,t) = \mathbf{I}$.

We first require that stabilizing gains can be synthesized; this is formulated in terms of an upper bound on the cost-to-go for the LQR control problem (Anderson & Moore (2007, Section 2)) induced by the JL dynamics.

Assumption 4.3 (Stabilizability)**.** Given a *policy* π, and a sequence of controls $\tilde{\mathbf{u}}(\cdot) \in \mathcal{U}$, let $V^\pi(t \mid \tilde{\mathbf{u}}, \xi) = \int_{s=t}^T (\|\tilde{\mathbf{x}}(s)\|^2 + \|\tilde{\mathbf{u}}(s)\|^2)\mathrm{d}s + \|\tilde{\mathbf{x}}(T)\|^2$, under the linearized dynamics $\frac{\mathrm{d}}{\mathrm{d}s}\tilde{\mathbf{x}}(s) = \mathbf{A}_{\text{ol}}^\pi(s)\tilde{\mathbf{x}}(s) + \mathbf{B}_{\text{ol}}^\pi(s)\tilde{\mathbf{u}}(s)$, $\tilde{\mathbf{x}}(t) = \xi$. We assume that, for all feasible policies, $\sup_{t\in[0,T]} V^\pi(t \mid \tilde{\mathbf{u}}, \xi) \leq \mu_{\text{ric}}\|\xi\|^2$. Moreover, we assume (for simplicity) that the initial policy has (a) no gains: $\mathtt{K}_k^{\pi^{(1)}} = 0$ for all $k \in [K]$, and (b) satisfies $V^{\pi^{(1)}}(t \mid 0, \xi) \leq \mu_{\text{ric}}\|\xi\|^2$.

The assumption on $\pi^{(1)}$ can easily be generalized to accomodate initial policies with stabilizing gains. Our final assumption is controllability (see e.g. Anderson & Moore (2007, Appendix B)), which is necesssary for identification of system parameters to synthesize stabilizing gains.

Assumption 4.4 (Controllability)**.** There exists constants $t_{\text{ctrl}}, \nu_{\text{ctrl}} > 0$ such that, for all feasible π and $t \in [t_{\text{ctrl}}, T]$, $\int_{s=t-t_{\text{ctrl}}}^t \mathbf{\Phi}_{\text{ol}}^\pi(t,s)\mathbf{B}_{\text{ol}}^\pi(s)\mathbf{B}_{\text{ol}}^\pi(s)^\top\mathbf{\Phi}_{\text{ol}}^\pi(t,s)^\top\mathrm{d}s \succeq \nu_{\text{ctrl}}\mathbf{I}_{d_x}$

For simplicity, we assume $k_{\mathrm{ctrl}} := t_{\mathrm{ctrl}}/\tau$ is integral. Finally, to state our theorem, we adopt an asymptotic notation which suppresses all parameters except $\{T, \tau, \alpha\}$.

Definition 4.3 (Asymptotic Notation)**.** We let $\mathcal{O}_\star(\cdot)\cdot$ term a term which hides polynomial dependences on $d_{\mathsf{x}}, d_{\mathsf{u}}, R_{\mathrm{feas}}, \kappa_f, M_f, L_f, \kappa_{\mathrm{cost}}, L_{\mathrm{cost}}, M_{\mathrm{cost}}, \mu_{\mathrm{ric}}, \nu_{\mathrm{ctrl}}, t_{\mathrm{ctrl}}$, and on $\exp(t_0 L_f)$, where $t_0 = \tau k_0 \geq t_{\mathrm{ctrl}}$.

Notice that we suppress an *exponential* dependence on our proxy t_0 for the controllability horizon t_{ctrl}; this is because the system cannot be stabilized until the dynamics can be accurately estimated, which requires waiting as long as the controllability window (Chen & Hazan, 2021; Tsiamis et al., 2022). We discuss this dependence further in Appendix B.5. Finally, we state a logarithmic term which addresses high-probability confidence:

$$\iota(\delta) := \log \frac{24T^2 n_{\mathrm{iter}} \max\{d_{\mathsf{x}}, d_{\mathsf{u}}\}}{\tau^2 \delta}. \tag{4.1}$$

We can now state our main theorem, which establishes that, with high probability, for a small enough step size τ, and large enough sample size N and iteration number n_{iter}, we obtain an ϵ-stationary policy and ϵ'-JSP, where ϵ^2, ϵ' scale as $\mathrm{poly}(T)(\tau^2 + \frac{1}{\tau^2\sqrt{N}})$:

Theorem 1. *Fix $\delta \in (0,1)$, and suppose for the sake of simplicity that $\tau \leq 1 \leq T$. Then, there are constants $c_1, \ldots, c_5 = \mathcal{O}_\star(1)$ such that if we tune $\eta = 1/c_1\sqrt{T}$, $\sigma_w = (\sigma_{\mathrm{orac}}^2 \iota(\delta)/N)^{\frac{1}{4}}$ and $k_0 \geq k_{\mathrm{ctrl}} + 2$, then as long as*

$$\tau \leq \tfrac{1}{c_2}, \quad N \geq c_3 \iota(\delta) \max\left\{ \tfrac{T^2}{\tau^2}, \tfrac{1}{\tau^4}, \sigma_{\mathrm{orac}}^2 \tfrac{T^4}{\tau^2}, \tfrac{\sigma_{\mathrm{orac}}^2}{\tau^8} \right\}.$$

Then, with probability $1-\delta$, if Condition 4.1 and all aforementioned Assumptions hold,

(a) For all $n \in [n_{\mathrm{iter}}]$, and $\pi' \in \{\pi^{(n)}, \tilde{\pi}^{(n)}\}$, $\mu_{\pi',\star} \leq 8\mu_{\mathrm{ric}}$ and $L_{\pi'} \leq 6\max\{1, L_f\}\mu_{\mathrm{ric}}$.

(b) $\pi = \pi^{(n_{\mathrm{out}})}$ is ϵ-stationary, where $\epsilon^2 = c_4(T\tau^2 + \frac{T^{\frac{3}{2}}}{n_{\mathrm{iter}}}) + c_4(\frac{T^{\frac{7}{2}}}{\tau^2}(\frac{\iota(\delta)^2}{N} + \sigma_{\mathrm{orac}}\sqrt{\frac{\iota(\delta)}{N}}) + \sigma_{\mathrm{orac}}^2 \frac{T^{\frac{3}{2}}\iota(\delta)^2}{N})$.

(c) For $\pi = \pi^{(n_{\mathrm{out}})}$, $\mathbf{u}^\pi$ is an ϵ'-JSP, where $\epsilon' = c_5 \frac{\epsilon^2}{\alpha}$.

As a corollary, we achieve Desideratum 1.

Corollary 4.1. *For any $\epsilon, \epsilon' > 0$ and $\delta \in (0,1)$, there exists an appropriate choices of $\{\tau, N, \eta, \sigma_w\}$ such that Algorithm 1 finds, with probability $\geq 1-\delta$, an ϵ-stationary policy π with $\mathbf{u}^\pi$ being an ϵ'-JSP using at most M oracle calls, where $M/\tau = \mathcal{O}_\star(\mathrm{poly}(T, 1/\epsilon, 1/\epsilon', \log(1/\delta)))$.*

4.1. Analysis Overview

In this section, we provide a high-level sketch of the analysis. Appendix A provides the formal proof, and carefully outlines the organization of the subsequent appendices which establish the subordinate results.

As our policies consists of zero-order hold discrete-time inputs, our analysis is mostly performed in discrete-time.

Definition 4.4 (Stabilized trajectories, discrete-time inputs)**.** Let $\vec{\mathbf{u}} \in \mathsf{U}$, and recall the continuous-input trajectories $\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}, \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}$ in Definition 2.1. We define $\tilde{\mathbf{x}}^\pi(t \mid \vec{\mathbf{u}}) := \tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \mathsf{ct}(\vec{\mathbf{u}}))$ and $\tilde{\mathbf{u}}^\pi(t \mid \vec{\mathbf{u}}) := \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathsf{ct}(\vec{\mathbf{u}}))$, and their discrete samplings $\tilde{\mathbf{x}}_k^\pi(\vec{\mathbf{u}}) := \tilde{\mathbf{x}}^\pi(t_k \mid \vec{\mathbf{u}})$ and $\tilde{\mathbf{u}}_k^\pi(\vec{\mathbf{u}}) := \tilde{\mathbf{u}}^\pi(t_k \mid \vec{\mathbf{u}})$. We define a discretized objective $\mathcal{J}_T^{\pi,\mathrm{disc}}(\vec{\mathbf{u}}) := V(\tilde{\mathbf{x}}_{K+1}^\pi(\vec{\mathbf{u}})) + \tau\sum_{k=1}^K Q(\tilde{\mathbf{x}}_k^\pi(\vec{\mathbf{u}}), \tilde{\mathbf{u}}_k^\pi(\vec{\mathbf{u}}), t_k)$, and the shorthand $\mathcal{J}_T^{\mathrm{disc}}(\pi) = \mathcal{J}_T^{\pi,\mathrm{disc}}(\mathbf{u}_{1:K}^\pi)$ and $\nabla\mathcal{J}_T^{\pi,\mathrm{disc}}(\pi) := \nabla_{\vec{\mathbf{u}}}\mathcal{J}_T^{\pi,\mathrm{disc}}(\vec{\mathbf{u}})\big|_{\vec{\mathbf{u}}=\mathbf{u}_{1:K}^\pi}$.

What we shall show is that our algorithm **(a)** finds a policy π such that $\|\nabla\mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2} \leq \epsilon$ is small, **(b)** by discretization, $\|\nabla\mathcal{J}_T^\pi(\mathbf{u}^\pi)\|_{\mathcal{L}_2(\mathcal{U})} \leq \epsilon + \mathcal{O}(\tau)$ is small (i.e. π is approximately stationary), and that **(c)** this implies that $\mathbf{u}^\pi$ is an approximate-JSP of $\mathcal{J}_T(\mathbf{u}^\pi)$. Part **(a)** requires the most effort, part **(b)** is a tedious discretization, and part **(c)** is by Proposition 4.1 stated below. Key in these steps are certain regularity conditions on the policy π. The first is the magnitude of the gains:

Definition 4.5. We define an upper bound on the gains of policy π as $L_\pi := \max\{1, \max_{k\in[K]} \|\mathtt{K}_k^\pi\|\}$.

This term suffices to translate stationary policies to JSPs:

Proposition 4.1. *Suppose Assumptions 2.1, 4.1 and 4.2, π is feasible, $\tau \leq \frac{1}{16 L_\pi L_f}$. Then, if $\|\nabla\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \leq \epsilon$, $\mathbf{u}^\pi(t)$ is an ϵ'-JSP of $\mathcal{J}_T$ for $\epsilon' = 64\epsilon^2 L_\pi^2/\alpha$.*

Proof Sketch. We construct a Jacobian linearization $\mathcal{J}_T^{\pi,\mathrm{jac}}$ of $\mathcal{J}_T^\pi$ by analogy to $\mathcal{J}^{\mathrm{jac}}$, and define ϵ-JSPs of $\mathcal{J}_T^\pi$ analogously. We show by inverting the gains that an ϵ-JSP of $\mathcal{J}_T^\pi$ is precisely an ϵ-JSP of $\mathcal{J}_T$. We then establish strong convexity of $\mathcal{J}_T^{\pi,\mathrm{jac}}$ (non-trivial due to the gains), and use the PL inequality for strongly convex functions to conclude. The formal proof is given in Appendix H.2. □

To establish parts **(a)** and **(b)**, we need to measure the stability of the policies. To this end, we first introduce *closed-loop* (discrete-time) linearizations of the dynamics, in terms of which we define a Lyapunov stability modulus.

Definition 4.6 (Closed-Loop Linearizations)**.** We discretize the open-loop linearizations in Definition 4.2 defining $\mathtt{A}_{\mathrm{ol},k}^\pi = \mathbf{\Phi}_{\mathrm{ol}}^\pi(t_{k+1}, t_k)$ and $\mathtt{B}_{\mathrm{ol},k}^\pi := \int_{s=t_k}^{t_{k+1}} \mathbf{\Phi}_{\mathrm{ol}}^\pi(t_{k+1}, s)\mathbf{B}_{\mathrm{ol}}^\pi(s)\mathrm{d}s$. We define an *discrete-time closed-loop* linearization $\mathtt{A}_{\mathrm{cl},k}^\pi := \mathtt{A}_{\mathrm{ol},k}^\pi + \mathtt{B}_{\mathrm{ol},k}^\pi \mathtt{K}_k^\pi$, and a discrete closed-loop *transition operator* is defined, for $1 \leq k_1 \leq k_2 \leq K+1$, $\mathtt{\Phi}_{\mathrm{cl},k_2,k_1}^\pi = \mathtt{A}_{\mathrm{cl},k_2-1}^\pi \cdot \mathtt{A}_{\mathrm{cl},k_2-2}^\pi \cdots \mathtt{A}_{\mathrm{cl},k_1}^\pi$, with the convention $\mathtt{\Phi}_{\mathrm{cl},k_1,k_1}^\pi = \mathbf{I}$. For $1 \leq k_1 < k_2 \leq K+1$, we define the closed-loop *Markov operator* $\mathtt{\Psi}_{\mathrm{cl},k_2,k_1}^\pi := \mathtt{\Phi}_{\mathrm{cl},k_2,k_1+1}^\pi \mathtt{B}_{\mathrm{ol},k_1}^\pi$.

Definition 4.7 (Lyapunov Stability Modulus)**.** Given a policy π, define $\Lambda^{\pi}_{K+1} = \mathbf{I}$, and $\Lambda^{\pi}_{k} = (\mathtt{A}^{\pi}_{\mathrm{cl},k})^{\top}\Lambda^{\pi}_{k+1}\mathtt{A}^{\pi}_{\mathrm{cl},k} + \tau\mathbf{I}$. We define $\mu_{\pi,\star} := \max_{k\in\{k_0,\dots,K+1\}} \|\Lambda^{\pi}_{k}\|$.

Notice that the stability modulus is taken after step k_0, which is where we terminate the Riccati recrusion in Algorithm 3. We shall show that, with high probability, Algorithm 1 synthesizes policies π which satisfy

$$L_\pi \le 6\max\{1, L_f\}\mu_{\mathrm{ric}}, \quad \mu_{\pi,\star} \le 8\mu_{\mathrm{ric}}, \tag{4.2}$$

so that $L_\pi, \mu_{\pi,\star} = \mathcal{O}_\star(1)$. Going forward, we let $\mathcal{O}_\pi(\cdot)\cdot$ denote a term suppressing polynomials in L_π, $\mu_{\pi,\star}$ and terms $\mathcal{O}_\star(1)$; when π satisfies Eq. (4.2), then $\mathcal{O}_\pi(\cdot) = \mathcal{O}_\star(\cdot)$. We say $x \le 1/\mathcal{O}_\pi(y)$, if $x \le 1/y'$, where $y' = \mathcal{O}_\pi(y)$. In Appendix I.3, we translate discrete-time stationary points to continuous-time ones, establishing part **(b)** of the argument.

Proposition 4.2. *For π feasible, $\|\nabla\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \le \frac{1}{\sqrt{\tau}}\|\nabla\mathcal{J}^{\mathrm{disc}}_T(\pi)\|_{\ell_2} + \mathcal{O}_\pi(\tau\sqrt{T})$.*

A more precise statement and explanation of the proof are given in Appendix A.5. The rest of the analysis boils down to **(a)**: finding an approximate stationary point of the time-discretized objective.

4.2. Finding a stationary point of $\mathcal{J}^{\pi,\mathrm{disc}}_T$

Taylor expansion of the dynamics. To begin, we derive perturbation bounds for solutions to the stabilized ordinary differential equations. Specifically, we provide bounds for when $\mathtt{u}^{\pi}_{1:K}$ is perturbed by a sufficiently small input $\delta\mathtt{u}_{1:K}$. Our formal proposition, Appendix A.6 states perturbations in both the ℓ_∞ and normalized ℓ_2-norms; for simplicity, state the special case for ℓ_∞-perturbation.

Proposition 4.3. *Let $\mathtt{u}_{1:K} = \mathtt{u}^{\pi}_{k} + \delta\mathtt{u}_{1:K}$, and suppose $\max_k \|\delta\mathtt{u}_k\| \le B_\infty \le 1/\mathcal{O}_\pi(1)$. Then, for all $k \in [K+1]$, $\|\tilde{\mathtt{x}}^{\pi}_{k}[\mathtt{u}_{1:K}] - \mathtt{x}^{\pi}_{k} - \sum_{j=1}^{k-1}\Psi^{\pi}_{\mathrm{cl},k,j}\delta\mathtt{u}_j\| \le \mathcal{O}_\pi(B^2_\infty)$.*

We also show, that if $B_\infty = 1/\mathcal{O}_\pi(T)$, then the policy with π' with the same gains $\mathtt{K}^{\pi}_{k} = \mathtt{K}^{\pi}_{k}$ as π, but the perturbed inputs $\mathtt{u}^{\pi'}_{k} = \mathtt{u}_k$ at most double its Lyapunov stability modulus $\mu_{\pi',\star} \le 2\mu_{\pi,\star}$. This allows small gradient steps to preserve stability.

Estimation of linearizations and gradients. We then argue that by making σ_w small, then to first order, the estimation procedure in Algorithm 2 recovers the *linearization* of the dynamics. The proof combines standard method-of-moments analysis based on matrix Chernoff concentration (Tropp, 2012) and Proposition 4.3 to argue the dynamics can be approximated by their linearization. Specifically, Appendix A.7 argues that, for all rounds $n \in [n_{\mathrm{iter}}]$ and $1 \le j < k \le K+1$, it holds that $\|\Psi^{\pi}_{\mathrm{cl},k,j} - \hat{\Psi}_{k,j}\| \le \mathrm{Err}_\Psi(\delta)$ where $\mathrm{Err}_\Psi(\delta) = \mathcal{O}_\pi(\sqrt{\frac{\iota(\delta)}{N}}(1 + \frac{\sigma_{\mathrm{orac}}}{\sigma_w} + \sigma_w)$, which can be made to scales as $N^{-\frac{1}{4}}$ by tuning $\sigma_w = (\sigma^2_{\mathrm{orac}}\iota(\delta)/N)^{\frac{1}{4}}$. From the Markov-recovery error, as well as a simpler bound for recovering $\mathtt{x}^{\pi}_{1:K}$ in Algorithm 2 (Lines 1-3), we show accurate recovery of the gradients: $\max_k \|\hat{\nabla}^{(n)}_k - (\nabla\mathcal{J}^{\mathrm{disc}}_T(\pi^{(n)}))_k\| \le T\mathcal{O}_\pi(\mathrm{Err}_\Psi(\delta))$.

The last step here is to argue that we also approximately recover $\mathtt{A}^{\pi}_{\mathrm{ol},k}, \mathtt{B}^{\pi}_{\mathrm{ol},k}$ in Algorithm 3 for synthesizing the gains: for all $k \ge k_0$,

$$\|\hat{\mathtt{B}}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\| \vee \|\hat{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{ol},k}\| \le \mathcal{O}_\pi(\tfrac{\mathrm{Err}_{\Psi,\pi}(\delta)}{\tau}).$$

This consists of two steps: (1) using controllability to show the matrices $\hat{\mathcal{C}}_{k,\mathrm{in}}$ in Algorithm 3 are well-conditioned and (2) using closeness of the Markov operators to show that $\hat{\mathcal{C}}_{k,\mathrm{in}}$ and $\hat{\mathcal{C}}_{k,\mathrm{out}}$ concentrate around their idealized values. Crucially, we only estimate system matrices for $k \ge k_0$ to ensure $\hat{\mathcal{C}}_{k,\mathrm{in}}$ is well-defined, and we use window $k_0 \ge k_{\mathrm{ctrl}} + 2$ to ensure $\hat{\mathcal{C}}_{k,\mathrm{out}}$ is sufficiently well-conditioned.

Concluding the proof. Appendices A.8 and A.9 conclude the proof with two steps: first, we show that cost-function decreases during the gradient step Algorithm 1 (Line 5) at round $n \in [n_{\mathrm{iter}}]$ in proportion to $-\|\hat{\nabla}^{(n)}_k\|^2$ (a consequence ofthe standard smooth descent argument). Here, we also apply the aforementioned result that small gradient steps preserve stability: $\mu_{\tilde{\pi}^{(n)},\star} \le 2\mu_{\pi^{(n)},\star}$. Second, we argue that the gains synthesized by Algorithm 3 ensure that the Lyapunov stability modulus of $\pi^{(n+1)}$ and the magnitude of its gains stay bounded by an algorithm-independent constant: $\mu_{\pi^{(n+1)},\star} \le 4\mu_{\mathrm{ric}} = \mathcal{O}_\star(1)$ and $L_{\pi^{(n+1)}} \le \mathcal{O}_\star(1)$; we use a novel certainty-equivalence analysis for discretized, time-varying linear systems which may be of independent interest (Appendix F). By combining these two results, we inductively show that all policies constructed satisfy (4.2), namely they have $\mu_{\pi,\star}$ and L_π at most $\mathcal{O}_\star(1)$. We then combine this with the typical analysis of nonconvex smooth gradient descent to argue that the policy $\pi^{(n_{\mathrm{out}})}$ has small discretized gradient, as needed.

5. Experiments

Our experiments evaluate the performance of our proposed trajectory optimization algorithm (Algorithm 1) and compare it with the well-established model-based baseline of trajectory optimization (`iLQR`) on top of learned dynamics (e.g. Levine & Koltun (2013)). Though our analysis considers a fixed horizon, we perform experiments in a receeding horizon control (RHC) fashion. We consider two control tasks: **(a)** a pendulum swing up task, and **(b)** a 2D quadrotor stabilization task. We implement our experiments using the `jax` (Bradbury et al., 2018) ecosystem. More details regarding the environments, tasks, and experimental setup details are found in Appendix J. Though our analysis considers the noisy oracle model, all experiments assume *noiseless* observations.

Least-squares vs. Method-of-Moments. Algorithm 1 prescribes the method-of-moments estimator to simplify the analysis; in our implementation, we find that estimating the transition operators using regularized least-squares instead yields to more sample efficient gradient estimation. This choice can also be analyzed with minor modifications (see e.g. Oymak & Ozay (2019); Simchowitz et al. (2019)).

`iLQR` **baseline.** We first collect a training dataset according to a prescribed exploration strategy, then train a neural network dynamics model on these dynamics, and finally optimize our policy by applying the `iLQR` algorithm directly on the learned model. We consider several variants of our `iLQR` baseline which use different exploration strategies and different supervision signals for model learning.

(1) Sampling strategies: We consider two sampling strategies; (a) `Agg` alternates between collecting data, fitting a dynamics model, and running `iLQR` to collect more data, and (b) `Rand` executes rollouts with random inputs starting from random initial conditions. The rationale is that the `Agg` strategy provides better data coverage for the desired task than `Rand`.

(2) Loss supervision: The standard loss supervision for learning dynamics is to regress against the next state transition. Inspired by our analysis, we also consider an idealized oracle that augments the supervision to also include noiseless the Jacobians of the ground truth model with respect to both the state and control input; we refer to this augmentation as `JacReg`.

(2) Model architecture: We use a fully connected three layer MLP network to for fitting the dynamics of the environment. Specifically, our model takes in input $(\mathtt{x}_k, \mathtt{u}_k)$ and predicts the state difference $\mathtt{x}_{k+1} - \mathtt{x}_k$.

Figure 1 shows the results of Algorithm 1 compared with several `iLQR` baselines on the pendulum and quadrotor tasks, respectively. In these figures, the x-axis plots the number of trajectories available to each algorithm, and the y-axis plots the cost suboptimality $(\mathcal{J}_T^{\text{alg}} - \mathcal{J}_T^\star)/\mathcal{J}_T^\star$ incurred by each algorithm; where $\mathcal{J}_T^{\text{alg}}$ is algorithmic cost and $\mathcal{J}_T^\star$ is the cost obtained via `iLQR` with the ground truth dynamics. The error bars in the plot are meidan, first and third quartile intervals computed over 20 different evaluation seeds.

Discussion. We observe that Algorithm 1 with feedback-gains consistently outperforms Algorithm 1 without gains, validating the important of locally-stabilized dynamics. Second, we see that the performance of the `iLQR` baselines does not significantly improve as more trajectory data is collected. We find that our learned models achieve very low train and test error, over the sampling distribution (i.e., `Agg` or `Rand`) used for learning. For `Rand`, we postulate that the distribution shift incurred by performing RHC via trajectory optimization on the learned model limits the closed-loop performance of our baseline. However, we note that `Agg+JacReg` achieves stellar performance early on, suggesting that (a) the `Agg` data collection method suffices for strong closed-loop performance (notice that `Rand+JacReg` fares far worse), and (b) that a second limiting factor is that estimating *dynamics* and performing automated differentiation is less favorable than directly estimating *Jacobians*, which are the fundamental quantities used by the `iLQR` algorithm. This gap between estimation of dynamics and derivatives has been observed in prior work (Pfrommer et al., 2022).

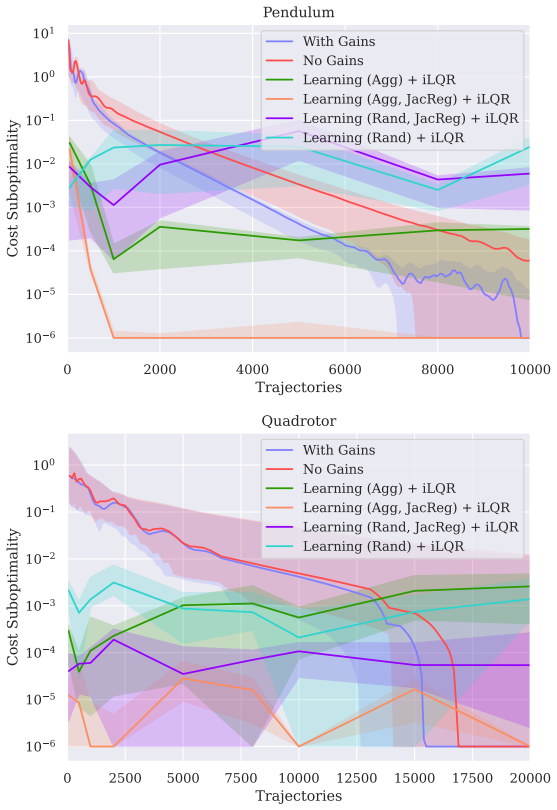

Figure 1: Cost suboptimality $(\mathcal{J}_T^{\text{alg}} - \mathcal{J}_T^\star)/\mathcal{J}_T^\star$ versus number of trajectories available to both Algorithm 1 and `iLQR` baselines. For visualization, the suboptimality is clipped to $(10^{-6}, \infty)$.

Though we find that our method outperforms deep-learning baselines (excluding `Agg+JacReg`) on the simpler inverted pendulum environment, the learning+`iLQR` approaches fare better on the quadrotor in the $\leq$ 10000 trajectories regime. We suspect that this is attributable to data-reuse, as Algorithm 1 estimates an entirely new model of system dynamics at each iteration. We believe that finding a way to combine the advantages of directly estimating linearized dynamics (observed in Algorithm 1, as well as `Agg+JacReg`) with the advantages of data-reuse would yield significant sample efficiency improvements.

References

Agarwal, N., Bullins, B., Hazan, E., Kakade, S., and Singh, K. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pp. 111–119. PMLR, 2019.

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

Arimoto, S., Kawamura, S., and Miyazaki, F. Bettering operation of robots by learning. *Journal of Robotic systems*, 1(2):123–140, 1984.

Aswani, A., Gonzalez, H., Sastry, S. S., and Tomlin, C. Provably safe and robust learning-based model predictive control. *Automatica*, 49(5):1216–1226, 2013.

Babuschkin, I., Baumli, K., Bell, A., Bhupatiraju, S., Bruce, J., Buchlovsky, P., Budden, D., Cai, T., Clark, A., Danihelka, I., Fantacci, C., Godwin, J., Jones, C., Hemsley, R., Hennigan, T., Hessel, M., Hou, S., Kapturowski, S., Keck, T., Kemaev, I., King, M., Kunesch, M., Martens, L., Merzic, H., Mikulik, V., Norman, T., Quan, J., Papamakarios, G., Ring, R., Ruiz, F., Sanchez, A., Schneider, R., Sezener, E., Spencer, S., Srinivasan, S., Wang, L., Stokowiec, W., and Viola, F. The DeepMind JAX Ecosystem, 2020. URL http://github.com/deepmind.

Bechtle, S., Lin, Y., Rai, A., Righetti, L., and Meier, F. Curious ilqr: Resolving uncertainty in model-based rl. In *Conference on Robot Learning*, pp. 162–171. PMLR, 2020.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q. JAX: composable transformations of Python+NumPy programs, 2018. URL http://github.com/google/jax.

Chen, X. and Hazan, E. Black-box control for linear dynamical systems. In *Conference on Learning Theory*, pp. 1114–1143. PMLR, 2021.

Dai, H., Landry, B., Yang, L., Pavone, M., and Tedrake, R. Lyapunov-stable neural-network control. *arXiv preprint arXiv:2109.14152*, 2021.

Dean, S. and Recht, B. Certainty equivalent perception-based control. In *Learning for Dynamics and Control*, pp. 399–411. PMLR, 2021.

Dean, S., Mania, H., Matni, N., Recht, B., and Tu, S. On the sample complexity of the linear quadratic regulator working draft. 2017.

Foster, D., Sarkar, T., and Rakhlin, A. Learning nonlinear dynamical systems from a single trajectory. In *Learning for Dynamics and Control*, pp. 851–861. PMLR, 2020.

Frostig, R., Sindhwani, V., Singh, S., and Tu, S. trajax: differentiable optimal control on accelerators, 2021. URL http://github.com/google/trajax.

Hennigan, T., Cai, T., Norman, T., and Babuschkin, I. Haiku: Sonnet for JAX, 2020. URL http://github.com/deepmind/dm-haiku.

Jacobson, D. H. and Mayne, D. Q. *Differential dynamic programming*. Number 24. Elsevier Publishing Company, 1970.

Jadbabaie, A. and Hauser, J. On the stability of unconstrained receding horizon control with a general terminal cost. In *Proceedings of the 40th IEEE Conference on Decision and Control (Cat. No. 01CH37228)*, volume 5, pp. 4826–4831. IEEE, 2001.

Kabzan, J., Hewing, L., Liniger, A., and Zeilinger, M. N. Learning-based model predictive control for autonomous racing. *IEEE Robotics and Automation Letters*, 4(4):3363–3370, 2019.

Karimi, H., Nutini, J., and Schmidt, M. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. In *Joint European conference on machine learning and knowledge discovery in databases*, pp. 795–811. Springer, 2016.

Kocijan, J., Murray-Smith, R., Rasmussen, C. E., and Girard, A. Gaussian process model based predictive control. In *Proceedings of the 2004 American control conference*, volume 3, pp. 2214–2219. IEEE, 2004.

Koller, T., Berkenkamp, F., Turchetta, M., and Krause, A. Learning-based model predictive control for safe exploration. In *2018 IEEE conference on decision and control (CDC)*, pp. 6059–6066. IEEE, 2018.

Levine, S. and Abbeel, P. Learning neural network policies with guided policy search under unknown dynamics. *Advances in neural information processing systems*, 27, 2014.

Levine, S. and Koltun, V. Guided policy search. In *International conference on machine learning*, pp. 1–9. PMLR, 2013.

Li, W. and Todorov, E. Iterative linear quadratic regulator design for nonlinear biological movement systems. In *ICINCO (1)*, pp. 222–229. Citeseer, 2004.

Mania, H., Jordan, M. I., and Recht, B. Active learning for nonlinear system identification with guarantees. *arXiv preprint arXiv:2006.10277*, 2020.

Mhammedi, Z., Foster, D. J., Simchowitz, M., Misra, D., Sun, W., Krishnamurthy, A., Rakhlin, A., and Langford, J. Learning the linear quadratic regulator from nonlinear observations. *Advances in Neural Information Processing Systems*, 33:14532–14543, 2020.

Morari, M. and Lee, J. H. Model predictive control: past, present and future. *Computers & Chemical Engineering*, 23(4-5):667–682, 1999.

Nagabandi, A., Konoglie, K., Levine, S., and Kumar, V. Deep dynamics models for learning dexterous manipulation. arxiv. *arXiv preprint arXiv:1909.11652*, 10, 2019.

Oymak, S. and Ozay, N. Non-asymptotic identification of lti systems from a single trajectory. In *2019 American control conference (ACC)*, pp. 5655–5661. IEEE, 2019.

Papadimitriou, D., Rosolia, U., and Borrelli, F. Control of unknown nonlinear systems with linear time-varying mpc. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pp. 2258–2263. IEEE, 2020.

Pfrommer, D., Zhang, T. T., Tu, S., and Matni, N. Tasil: Taylor series imitation learning. *arXiv preprint arXiv:2205.14812*, 2022.

Polak, E. *Optimization: algorithms and consistent approximations*, volume 124. Springer Science & Business Media, 2012.

Rosolia, U. and Borrelli, F. Learning how to autonomously race a car: a predictive control approach. *IEEE Transactions on Control Systems Technology*, 28(6):2713–2719, 2019.

Roulet, V., Srinivasa, S., Drusvyatskiy, D., and Harchaoui, Z. Iterative linearized control: stable algorithms and complexity guarantees. In *International Conference on Machine Learning*, pp. 5518–5527. PMLR, 2019.

Sattar, Y. and Oymak, S. Non-asymptotic and accurate learning of nonlinear dynamical systems. *Journal of Machine Learning Research*, 23(140):1–49, 2022.

Simchowitz, M. and Foster, D. Naive exploration is optimal for online lqr. In *International Conference on Machine Learning*, pp. 8937–8948. PMLR, 2020.

Simchowitz, M., Mania, H., Tu, S., Jordan, M. I., and Recht, B. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pp. 439–473. PMLR, 2018.

Simchowitz, M., Boczar, R., and Recht, B. Learning linear dynamical systems with semi-parametric least squares. In *Conference on Learning Theory*, pp. 2714–2802. PMLR, 2019.

Stewart, G. W. On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM review*, 19(4):634–662, 1977.

Tassa, Y., Erez, T., and Todorov, E. Synthesis and stabilization of complex behaviors through online trajectory optimization. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4906–4913. IEEE, 2012.

Todorov, E. and Li, W. A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems. In *Proceedings of the 2005, American Control Conference, 2005.*, pp. 300–306. IEEE, 2005.

Torrente, G., Kaufmann, E., Föhn, P., and Scaramuzza, D. Data-driven mpc for quadrotors. *IEEE Robotics and Automation Letters*, 6(2):3769–3776, 2021.

Tropp, J. A. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12 (4):389–434, 2012.

Tsiamis, A., Ziemann, I. M., Morari, M., Matni, N., and Pappas, G. J. Learning to control linear systems can be hard. In *Conference on Learning Theory*, pp. 3820–3857. PMLR, 2022.

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

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

Williams, G., Wagener, N., Goldfain, B., Drews, P., Rehg, J. M., Boots, B., and Theodorou, E. A. Information theoretic mpc for model-based reinforcement learning. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1714–1721. IEEE, 2017.

Xu, X. On the perturbation of the moore–penrose inverse of a matrix. *Applied Mathematics and Computation*, 374: 124920, 2020.

Contents

Part I
Analysis

A. Formal Analysis

A.1. Organization of the Appendix

First, we begin with an outline of Appendix A:

- Appendix A.2 reviews essential notation.
- Appendix A.3 gives a restatement of our main result, Theorem 1, as Theorem 2.

The rest of Appendix A carries out the proof of Theorem 2. Speficially,

- Appendix A.4 defines numerous problem parameters on which our arguments depend.

- Appendix A.5 proves Corollary A.1, a precise statement of Proposition 4.2 in the main text. It does so via an intermediate result, Proposition A.4, which bounds the $\mathcal{L}_\infty$ difference between the continuous-time gradient, and the imagine of the discrete-time gradient under the continuous-time inclusion map $\mathtt{ct}(\cdot)$.
- Appendix A.6 states key results based on Taylor expansions of dynamics around their linearizations, and norms of various derivative-like quantities.
- Appendix A.7 contains the main statements of the various estimation guarantees, notably, the recovery of nominal trajectories, Markov operators, discretized gradients, and linearized transition matrices $(\mathtt{A}^\pi_{\mathrm{ol},k}, \mathtt{B}^\pi_{\mathrm{ol},k})$.
- Appendix A.8 leverages the previous section to demonstrate (a) a certain descent condition holds for each gradient step and (b) that sufficiently accurate estimates of transition matrices lead to the synthesis of gains for which the corresponding policies have bounded stability moduli.
- Finally, Appendix A.9 concludes the proof, as well as states a more granular guarantee in terms of specific problem parameters and not general $\mathcal{O}_\star(\cdot)$ notation.

The rest of Part I of the Appendix provides the proofs of constituent results. Specifically,

- Appendix B presents various discussion of main results, as well as gesturing to extensions. Specifcally, Appendix B.1 describes the exponential gap between FOSs of $\mathcal{J}_T$ and JSPs, and Appendix B.2 explains the consequences of combining our result with (Westenbroek et al., 2021). We discuss how to implement a projection step to ensure Definition 4.7 in Appendix B.3. Finally, we discuss extensions to an oracle with process noise in Appendix B.4.
- Appendix C presents various computations of Jacobian linearizations, establishing that they do accurately capture first-order expansions.
- Appendix D proves all the Taylor-expansion like results listed in Appendix A.6.
- Appendix E proves all the estimation-error bounds in Appendix A.7.
- Appendix F provides a general certainty-equivalence and Lyapunov stability perturbation results for time-varying, discrete-time linear systems, in the regime that naturally arises when the state matrices are derived from discretizations of continuous-time dynamics.
- Appendix G instantiates the bounds in Appendix F to show that the gains synthesized by Algorithm 2 do indeed lead to policies with bounded stability modulus.
- Appendix H contains the proofs of optimization-related results: the proof of the descent lemma (Lemma A.13 (in Appendix H.1) and the proof of the conversion between stationary points and JSPs, Proposition 4.1 (in Appendix H.2)
- Finally, Appendix I contains various time-discretization arguments, and in particular establishes the aforementiond Proposition A.4 from Appendix A.5.

A.2. Notation Review

In this section, we review our basic notation.

Dynamics. Recall the nominal system dynamics are given by

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t \mid \mathbf{u}) = f_{\mathrm{dyn}}(\mathbf{x}(t \mid \mathbf{u}), \mathbf{u}(t)), \;\; \mathbf{x}(0 \mid \mathbf{u}) = \xi_{\mathrm{init}}.$$

We recall the definition of various stabilized dynamics.

Definition 2.1. Given a continuous-time input $\bar{\mathbf{u}} \in \mathcal{U}$, we define the *stabilized trajectory* $\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) := \mathbf{x}(t \mid \tilde{\mathbf{u}}^{\pi,\mathrm{ct}})$, where $\tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) := \bar{\mathbf{u}}(t) + \mathtt{K}^\pi_{k(t)}(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t_{k(t)} \mid \bar{\mathbf{u}})) - \mathtt{x}^\pi_{k(t)})$. This induces a stabilized objective: $\mathcal{J}^\pi_T(\bar{\mathbf{u}}) := V(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}})) + \int_0^T Q(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}), \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathbf{u}), t)\mathrm{d}t$. We define the shorthand $\nabla\mathcal{J}_T(\pi) := \nabla_{\bar{\mathbf{u}}}\mathcal{J}^\pi_T(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^\pi}$

Definition 4.4 (Stabilized trajectories, discrete-time inputs)**.** Let $\vec{\mathtt{u}} \in \mathsf{U}$, and recall the continuous-input trajectories $\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}, \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}$ in Definition 2.1. We define $\tilde{\mathbf{x}}^{\pi}(t \mid \vec{\mathtt{u}}) := \tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \mathsf{ct}(\vec{\mathtt{u}}))$ and $\tilde{\mathbf{u}}^{\pi}(t \mid \vec{\mathtt{u}}) := \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathsf{ct}(\vec{\mathtt{u}}))$, and their discrete samplings $\tilde{\mathtt{x}}^{\pi}_k(\vec{\mathtt{u}}) := \tilde{\mathbf{x}}^{\pi}(t_k \mid \vec{\mathtt{u}})$ and $\tilde{\mathtt{u}}^{\pi}_k(\vec{\mathtt{u}}) := \tilde{\mathbf{u}}^{\pi}(t_k \mid \vec{\mathtt{u}})$. We define a discretized objective $\mathcal{J}_T^{\pi,\mathrm{disc}}(\vec{\mathtt{u}}) := V(\tilde{\mathtt{x}}^{\pi}_{K+1}(\vec{\mathtt{u}})) + \tau \sum_{k=1}^{K} Q(\tilde{\mathtt{x}}^{\pi}_k(\vec{\mathtt{u}}), \tilde{\mathtt{u}}^{\pi}_k(\vec{\mathtt{u}}), t_k)$, and the shorthand $\mathcal{J}_T^{\mathrm{disc}}(\pi) = \mathcal{J}_T^{\pi,\mathrm{disc}}(\mathtt{u}^{\pi}_{1:K})$ and $\nabla\mathcal{J}_T^{\pi,\mathrm{disc}}(\pi) := \nabla_{\vec{\mathtt{u}}}\mathcal{J}_T^{\pi,\mathrm{disc}}(\vec{\mathtt{u}})\big|_{\vec{\mathtt{u}}=\mathtt{u}^{\pi}_{1:K}}$.

Linearizations. Next, we recall the definition of the various linearizations.

Definition 4.2 (Open-Loop Linearized Dynamics)**.** We define the (open-loop) JL dynamic matrices about π as $\mathbf{A}^{\pi}_{\mathrm{ol}}(t) = \partial_x f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(t), \mathbf{u}^{\pi}(t))$ and $\mathbf{B}^{\pi}_{\mathrm{ol}}(t) = \partial_u f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(t), \mathbf{u}^{\pi}(t))$. We define the *open-loop* JL transition function $\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(s,t)$, defined for $t \geq s$ as the solution to $\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(s,t) = \mathbf{A}^{\pi}_{\mathrm{ol}}(s)\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(s,t)$, with initial condition $\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t,t) = \mathbf{I}$.

Definition 4.6 (Closed-Loop Linearizations)**.** We discretize the open-loop linearizations in Definition 4.2 defining $\mathtt{A}^{\pi}_{\mathrm{ol},k} = \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k+1}, t_k)$ and $\mathtt{B}^{\pi}_{\mathrm{ol},k} := \int_{s=t_k}^{t_{k+1}} \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k+1}, s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\mathrm{d}s$. We define an *discrete-time closed-loop* linearization $\mathtt{A}^{\pi}_{\mathrm{cl},k} := \mathtt{A}^{\pi}_{\mathrm{ol},k} + \mathtt{B}^{\pi}_{\mathrm{ol},k}\mathtt{K}^{\pi}_k$, and a discrete closed-loop *transition operator* is defined, for $1 \leq k_1 \leq k_2 \leq K+1$, $\mathtt{\Phi}^{\pi}_{\mathrm{cl},k_2,k_1} = \mathtt{A}^{\pi}_{\mathrm{cl},k_2-1} \cdot \mathtt{A}^{\pi}_{\mathrm{cl},k_2-2} \cdot \cdots \cdot \mathtt{A}^{\pi}_{\mathrm{cl},k_1}$, with the convention $\mathtt{\Phi}^{\pi}_{\mathrm{cl},k_1,k_1} = \mathbf{I}$. For $1 \leq k_1 < k_2 \leq K+1$, we define the closed-loop *Markov operator* $\mathtt{\Psi}^{\pi}_{\mathrm{cl},k_2,k_1} := \mathtt{\Phi}^{\pi}_{\mathrm{cl},k_2,k_1+1}\mathtt{B}^{\pi}_{\mathrm{ol},k_1}$.

We also recall the definition of stationary policies and JSPs.

Definition 2.2. We say $\mathbf{u}$ is an ϵ-FOS of a function $\mathcal{F} : \mathcal{U} \to \mathbb{R}$ if $\|\nabla_{\mathbf{u}}\mathcal{F}(\mathbf{u})\|_{\mathcal{L}_2(\mathcal{U})} \leq \epsilon$. We say π is *ϵ-stationary* if $\|\nabla J_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} := \|\nabla_{\bar{\mathbf{u}}}\mathcal{J}_T^{\pi}(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^{\pi}}\|_{\mathcal{L}_2(\mathcal{U})} \leq \epsilon$.

Definition 2.4. We say $\mathbf{u} \in \mathcal{U}$ is an ϵ-Jacobian Stationary Point (JSP) if $\mathcal{J}_T(\mathbf{u}) \leq \inf_{\bar{\mathbf{u}} \in \mathcal{U}} \mathcal{J}_T^{\mathrm{jac}}(\bar{\mathbf{u}}; \mathbf{u}) + \epsilon$.

Problem Constants. We recall the dynamics-constants κ_f, L_f, M_f defined in Assumption 4.1, $\kappa_{\mathrm{cost}}, L_{\mathrm{cost}}, M_{\mathrm{cost}}$ in Assumption 4.2, the strong-convexity parameter α in Assumption 2.1, the controllability parameters $t_{\mathrm{ctrl}}, \nu_{\mathrm{ctrl}}$ from Assumption 4.4, with $k_{\mathrm{ctrl}} := t_{\mathrm{ctrl}}/\tau$, and the Riccati parameter μ_{ric} from Assumption 4.3. Finally, we recall the feasibility radius from Condition 4.1. We also recall

Definition 4.1. We say $(x,u) \in \mathbb{R}^{d_{\mathrm{x}} \times d_{\mathrm{u}}}$ are *feasible* if $\|x\| \vee \|u\| \leq R_{\mathrm{feas}}$. We say a policy π is feasible if $(2\mathbf{x}^{\pi}(t), 2\mathbf{u}^{\pi}(t))$ are feasible for all $t \in [0,T]$.

Gradient and Cost Shorthands. Notably, we bound out the following shorthand for gradients and costs:

$$\nabla\mathcal{J}_T(\pi) := \nabla_{\mathbf{u}}\mathcal{J}_T^{\pi}(\mathbf{u})\big|_{\mathbf{u}=\mathbf{u}^{\pi}}, \quad \nabla\mathcal{J}_T^{\mathrm{disc}}(\pi) := \nabla_{\vec{\mathtt{u}}}\mathcal{J}_T^{\pi,\mathrm{disc}}(\vec{\mathtt{u}})\big|_{\vec{\mathtt{u}}=\mathtt{u}^{\pi}_{1:K}}, \quad \mathcal{J}_T^{\mathrm{disc}}(\pi) := \mathcal{J}_T^{\pi,\mathrm{disc}}(\mathtt{u}^{\pi}_{1:K}). \tag{A.1}$$

A.3. Full Statement of Main Result

The following is a slightly more general statement of Theorem 1, which implies Theorem 1 for appropriate choice of $\eta \leftarrow \frac{1}{c_1}\min\left\{\frac{1}{\sqrt{T}}, 1\right\}$, $\sigma_w \leftarrow (\sigma^2_{\mathrm{orac}}\iota(\delta)/N)^{\frac{1}{4}}$, and with the simplifications $\tau \leq 1 \leq T$.

Theorem 2. *Fix $\delta \in (0,1)$, define $\iota(\delta) := \log \frac{24T^2 n_{\mathrm{iter}} \max\{d_{\mathrm{x}}, d_{\mathrm{u}}\}}{\tau^2\delta}$ and $\mathrm{Err}_0(\delta) := \sqrt{\iota(\delta)/N}$, where N is the sample size, and suppose we select $\sigma_w = c(\sigma^2_{\mathrm{orac}}\iota(\delta)/N)^{\frac{1}{4}}$ for any $c \in [\frac{1}{\mathcal{O}_\star(1)}, \mathcal{O}_\star(1)]$. Then, there exists constants $c_1, c_2, \ldots, c_5 = \mathcal{O}_\star(1)$ depending on c such that the following holds. Suppose that*

$$\eta \leq \frac{1}{c_1}\min\left\{\frac{1}{\sqrt{T}}, 1\right\}, \quad \tau \leq \frac{1}{c_2} \quad N \geq c_3\iota(\delta)\max\left\{1, \frac{T^2}{\tau^2}, \frac{1}{\tau^4}, T, \sigma^2_{\mathrm{orac}}\frac{T^4}{\tau^2}, \frac{\sigma^2_{\mathrm{orac}}}{\tau^8}\right\}. \tag{A.2}$$

Then, with probability $1-\delta$, if Condition 4.1 and all listed Assumptions hold,

(a) For all $n \in [n_{\mathrm{iter}}]$, and $\pi' \in \{\pi^{(n)}, \tilde{\pi}^{(n)}\}$, $\mu_{\pi',\star} \leq 8\mu_{\mathrm{ric}}$ and $L_{\pi'} \leq 6\max\{1, L_f\}\mu_{\mathrm{ric}}$.

(a) For $\pi = \pi^{(n_{\mathrm{out}})}$, π is ϵ-stationary where

$$\epsilon^2 = c_4 T\left(\tau^2 + \frac{1}{\eta}\left(\frac{1}{n_{\mathrm{iter}}} + \left(\eta T^2 + \frac{T^2}{\tau^2}\right)\left(\frac{\iota(\delta)^2}{N} + \sigma_{\mathrm{orac}}\sqrt{\frac{\iota(\delta)}{N}}\right) + \sigma^2_{\mathrm{orac}}\frac{\iota(\delta)^2}{N}\right)\right).$$

(c) For $\pi = \pi^{(n_{\mathrm{out}})}$, $\mathbf{u}^{\pi}$ is an ϵ'-JSP, where $\epsilon' = c_5\frac{\epsilon^2}{\alpha}$.

A.4. Problem Parameters

In this section, we provide all definitions of various problem paramaters. The notation is extensive, but we maintain the following conventions:

1. $\mu_{(\cdot)}$ refers to upper bounds on Lyapunov operators, $\kappa_{(\cdot)}$ to upper bounds on zero-order terms (e.g. $\|f_{\mathrm{dyn}}(x,u)\|$) or magnitudes of transition operators, $M_{(\cdot)}$ to bounds on second-order derivatives, $L_{(\cdot)}$ to bounds on first-order derivatives, $B_{(\cdot)}$ to upper bounds on radii, $\tau_{(\cdot)}$ to step sizes, $\mathrm{Err}_{(\cdot)}$ to error terms.
2. $q \in \{1, 2, \infty\}$ corresponds to ℓ_q norms
3. Subscripts tay denote relevance to Taylor expansions of the dynamics.
4. Terms with have a subscript π hide dependence on L_π, $\mu_{\pi,\star}$ and κ_q for $q \in \{1, 2, \infty\}$

Remark A.1 (Reminder on Asymptotic Notation)**.** We let $\mathcal{O}_\star(x)$ denote a term which suppresses polynomial dependence on all the constants in Assumptions 4.1 and 4.2, as well as μ_{ric} in Assumption 4.3, and ν_{ctrl}, $t_0 \geq t_{\mathrm{ctrl}}$ and $e^{L_f t_0} \geq e^{L_f t_{\mathrm{ctrl}}}$, where $t_0 = \tau k_0$, and $\nu_{\mathrm{ctrl}}, t_{\mathrm{ctrl}}$ are given in Assumption 4.4. We let $\mathcal{O}_\pi(x)$ suppress all of these constants, as well as polynomials in L_π and $\mu_{\pi,\star}$.

A.4.1. Stability Constants

We begin by recalling the primary constants controlling the stability of a policy π.

Definition 4.7 (Lyapunov Stability Modulus)**.** Given a policy π, define $\Lambda^\pi_{K+1} = \mathbf{I}$, and $\Lambda^\pi_k = (\mathtt{A}^\pi_{\mathrm{cl},k})^\top \Lambda^\pi_{k+1} \mathtt{A}^\pi_{\mathrm{cl},k} + \tau \mathbf{I}$. We define $\mu_{\pi,\star} := \max_{k \in \{k_0,\dots,K+1\}} \|\Lambda^\pi_k\|$.

It is more convenient to prove bounds in terms of the following three quantity, which are defined in terms of the magnitudes of the closed-loop transition operators.

Definition A.1 (Norms of π)**.** We define the constants $\kappa_{\pi,\infty} := \max_{1 \leq j \leq k \leq K+1} \|\Phi^\pi_{\mathrm{cl},k,j}\|$, and

$$\kappa_{\pi,1} := \max_{k \in [K+1]} \tau \left(\sum_{j=1}^{k} \|\Phi^\pi_{\mathrm{cl},k,j}\| \vee \sum_{j=k}^{K+1} \|\Phi^\pi_{\mathrm{cl},j,k}\| \right)$$

$$\kappa^2_{\pi,2} := \max_{k \in [K+1]} \tau \left(\sum_{j=1}^{k} \|\Phi^\pi_{\mathrm{cl},k,j}\|^2 \vee \sum_{j=k}^{K+1} \|\Phi^\pi_{\mathrm{cl},j,k}\|^2 \right)$$

We also define the following upper bounds on these quantities:

$$\kappa_\infty(\mu, L) := \sqrt{\max\{1, 6L_f L\}\mu} \exp(t_0 L_f)$$
$$\kappa_2(\mu, L) := \max\{1, 6L_f L\}\mu \left(t_0 \exp(2t_0 L_f) + \mu\right)$$
$$\kappa_1(\mu, L) := \sqrt{\max\{1, 6L_f L\}\mu} \left(t_0 \exp(t_0 L_f) + 2\mu\right)$$

The following lemma is proven in Appendix G.4, and shows that shows that each of the above terms is $\mathcal{O}_\pi(1)$.

Lemma A.1. *Let π be any policy. Recall $t_0 = \tau k_0$ Then, as long as $\tau \leq 1/6L_f L_\pi$,*

$$\mu_{\pi,q} \leq \mu_q(\mu_{\pi,\star}, L_\pi) = \mathcal{O}_\pi(1).$$

A.4.2. Discretization Step Magnitudes

Next, we introduce various maximal discretization step sizes for which our discrete-time dynamics are sufficiently faithful to the continuous ones. The first is a general condition for the dynamics to be "close", the second is useful for closeness of solutions of Ricatti equations, the third for the discrete-time dynamics to admit useful Taylor expansions, and the fourth for discrete-time controllability. We note that the first two do not depend on π, while the second two do.

Definition A.2 (Discretization Sizes)**.** We define

$$\begin{aligned}
\tau_{\text{dyn}} &:= \frac{1}{4L_f} \\
\tau_{\text{ric}} &:= \frac{1}{4\mu_{\text{ric}}^2\left(3M_f\kappa_f\mu_{\text{ric}}L_f + 13L_f^2(1+L_f\mu_{\text{ric}})^2\right)} \\
\tau_{\text{tay},\pi} &:= \min\left\{\frac{1}{16L_fL_\pi}, \frac{1}{8\kappa_f}\right\} \leq \tau_{\text{dyn}}. \\
\tau_{\text{ctrl},\pi} &:= \frac{\nu_{\text{ctrl}}}{8L_\pi^2K_\pi^2\gamma_{\text{ctr}}^3\exp(2\gamma_{\text{ctr}})\left(\kappa_fM_f + 2L_f^2\right)}, \quad \gamma_{\text{ctr}} := \max\{1, L_ft_{\text{ctrl}}\}
\end{aligned}$$

We note that $\tau_{\text{dyn}}, \tau_{\text{ric}} = 1/\mathcal{O}_\star(1)$ and $\tau_{\text{tay},\pi}, \tau_{\text{ctrl},\pi} = 1/\mathcal{O}_\pi(1)$.

A.4.3. Taylor Expansion Constants.

We now define the relevant constants in terms of which we bound our taylor expansions.

Definition A.3 (Taylor Expansion Constants, Policy Dependent)**.** We define $L_{\text{tay},\infty,\pi} = 2L_f\kappa_{\pi,1}$, $L_{\text{tay},2,\pi} := 2L_f\kappa_{\pi,2}$, and

$$\begin{aligned}
M_{\text{tay},2,\pi} &:= 8M_f(\kappa_{\pi,\infty} + 10L_\pi^2L_f^2\kappa_{\pi,2}^2\kappa_{\pi,1}) \\
M_{\text{tay},\text{inf},\pi} &:= 8M_f(\kappa_{\pi,1} + 10L_\pi^2L_f^2\kappa_{\pi,1}^3) \\
B_{\text{tay},2,\pi} &= \min\left\{\frac{1}{\sqrt{40M_fL_\pi^2\kappa_{\pi,1}M_{\text{tay},2,\pi}}}, \frac{L_f\kappa_{\pi,2}}{2M_{\text{tay},2,\pi}}, \frac{R_{\text{feas}}}{16L_\pi L_f\kappa_{\pi,2}}\right\} \\
B_{\text{tay},\text{inf},\pi} &= \min\left\{\frac{1}{40L_\pi^2\kappa_{\pi,1}M_{\text{tay},\text{inf},\pi}}, \frac{L_f\kappa_{\pi,1}}{2M_{\text{tay},\text{inf},\pi}}, \frac{R_{\text{feas}}}{16L_\pi L_f\kappa_{\pi,1}}\right\}
\end{aligned}$$

We also define

$$\begin{aligned}
M_{\mathcal{J},\text{tay},\pi} &:= 2M_{\text{cost}}L_f^2\kappa_{\pi,2}^2(1+3L_\pi^2T)M_{\text{tay},2,\pi} + L_{\text{cost}}(1+2L_\pi T)M_{\text{tay},2,\pi} + 2L_\pi L_{\text{cost}}, \\
B_{\text{stab},\pi} &:= (\max\{6, 36L_fL_\pi\}\mu_{\pi,\star}\cdot 12TM_fL_\pi(1+L_fK_\pi)B_\infty)^{-1} \\
L_{\nabla,\pi,\infty} &:= L_{\text{cost}}(1 + \frac{3L_f}{2}\kappa_{\pi,\infty} + 3L_\pi\kappa_{\pi,1})
\end{aligned}$$

The following is a consequence of Lemma A.1.

Lemma A.2. *By Lemma A.1, $M_{\text{tay},2,\pi}, M_{\text{tay},\text{inf},\pi}, L_{\nabla,\pi,\infty}, L_{\text{tay},q,\pi} = \mathcal{O}_\pi(1)$, $B_{\text{tay},2,\pi}, B_{\text{tay},\text{inf},\pi} = 1/\mathcal{O}_\pi(1)$, $M_{\mathcal{J},\text{tay},\pi} = T\cdot\mathcal{O}_\pi(1)$, and $B_{\text{stab},\pi} = \frac{1}{T}\cdot\mathcal{O}_\pi(1)$.*

The first group of four constants arises in Taylor expansions of the dynamics, the fith in a Taylor expansion of the cost functional, and the sixth in controlling the stability of policies under changes to the input, and the last upper bounds the norm of the gradient.

A.4.4. Estimation Error Terms.

Finally, we define the following error terms which arise in the errors of the extimated nominal trajectories, Markov operators, and gradients. Note that the first term has no dependence on π, while the latter two do.

Definition A.4 (Error Terms)**.** Define $\iota(\delta) := \log\frac{24T^2n_{\text{iter}}\max\{d_{\mathsf{x}},d_{\mathsf{u}}\}}{\tau^2\delta} = \log\frac{24K^2n_{\text{iter}}d_\star}{\delta}$, where $d_\star := \max\{d_{\mathsf{x}}, d_{\mathsf{u}}\}$.

Further, define

$$\begin{aligned}
\mathrm{Err}_{\hat{x}}(\delta) &:= \sigma_{\mathrm{orac}}\sqrt{2\frac{d_\star \iota(\delta)}{N}} \\
\mathrm{Err}_{\Psi,\pi}(\delta) &:= \sqrt{\frac{\iota(\delta)}{N}}\left(\frac{2\sigma_{\mathrm{orac}}}{\sigma_w}d_\star^{3/2} + 8L_{\mathrm{tay},\infty,\pi}d_\star\right) + 4\sigma_w M_{\mathrm{tay},2,\pi}d_\star^{3/2} = \mathcal{O}_\pi(\sqrt{\frac{\iota(\delta)}{N}}(1+\frac{\sigma_{\mathrm{orac}}}{\sigma_w}+\sigma_w) \\
\mathrm{Err}_{\nabla,\pi}(\delta) &:= (L_{\mathrm{cost}}\mathrm{Err}_{\Psi,\pi}(\delta) + (1+\kappa_{\pi,\infty})M_{\mathrm{cost}}\mathrm{Err}_{\hat{x}}(\delta))\,(1+2TL_\pi).
\end{aligned}$$

We note that, in view of Lemma A.1,

By Lemmas A.1 and A.2, we have

Lemma A.3. *Define* $\mathrm{Err}_0(\delta) = \sqrt{\frac{\iota(\delta)}{N}}$. *Then,*

$$\begin{aligned}
\mathrm{Err}_{\hat{x}}(\delta) &= \sigma_{\mathrm{orac}}\sqrt{2d_\star}\mathrm{Err}_0(\delta) \le \mathcal{O}_\star\left(\mathrm{Err}_0(\delta)\right) \\
\mathrm{Err}_{\Psi}(\delta) &\le \mathcal{O}_\pi\left(\mathrm{Err}_0(\delta)(1+\frac{\sigma_{\mathrm{orac}}}{\sigma_w}) + \sigma_w\right) \\
\mathrm{Err}_{\nabla,\pi}(\delta) &\le \mathcal{O}_\pi\left(T\left(\mathrm{Err}_0(\delta)(1+\sigma_{\mathrm{orac}}+\frac{\sigma_{\mathrm{orac}}}{\sigma_w}) + \sigma_w\right)\right).
\end{aligned} \tag{A.3}$$

If we further tune $\sigma_w = c\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}$ *for any* $c \in [1/\mathcal{O}_\star(1), \mathcal{O}_\star(1)]$, *then*

$$\begin{aligned}
\mathrm{Err}_{\hat{x}}(\delta) &\le \mathcal{O}_\star\left(\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)\right) \\
\mathrm{Err}_{\Psi}(\delta) &\le \mathcal{O}_\pi\left(\mathrm{Err}_0(\delta) + \sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}\right) \\
\mathrm{Err}_{\nabla,\pi}(\delta) &\le \mathcal{O}_\pi\left(T\left(\mathrm{Err}_0(\delta) + \sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}\right)\right).
\end{aligned} \tag{A.4}$$

A.5. Gradient Discretization

We begin by stating with the precise statement of Proposition 4.2, which relates norms of gradients of the discretized objective to that of the continuous-time one. We begin with the following proposition which bounds the difference between the continuous-time gradient, and a (normalized) embedding of the discrete-time gradient into continuous-time. We define the constant

$$\kappa_\nabla := \left((1+L_f)M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}(3\kappa_f M_f + 8L_f^2 + L_f)\right) = \mathcal{O}_\star(\cdot)1 \tag{A.5}$$

Proposition A.4 (Discretization of the Gradient)**.** *Let* π *be feasible, and let* $\tilde{\nabla}\mathcal{J}_T(\pi) = \frac{1}{\tau}\tau(\nabla\mathcal{J}_T^{\mathrm{disc}}(\pi))$ *is the continuous-time inclusion of the discrete-time gradient, normalized by* τ^{-1}. *Then,*

$$\sup_{t\in[0,T)} \|\nabla\mathcal{J}_T(\pi)(t) - \tilde{\nabla}\mathcal{J}_T(\pi)(t)\| \le \tau e^{\tau L_f}\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}L_\pi\kappa_\nabla,$$

The above result is proven in Appendix I.3. By integrating, we see that $\|\nabla\mathcal{J}_T(\pi) - \tilde{\nabla}\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \le \sqrt{T}\tau e^{\tau L_f}\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}L_\pi\kappa_\nabla$, and thus the triangle inequality gives $\|\nabla\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \le \|\tilde{\nabla}\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} + \sqrt{T}\tau e^{\tau L_f}\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}L_\pi\kappa_\nabla$. We can see that for any $\vec{\mathsf{u}} = \mathsf{u}_{1:K} \in \mathsf{U}$, $\|\vec{\mathsf{u}}\|_{\ell_2}^2 = \sum_{k=1}^K \|\mathsf{u}_k\|^2 = \frac{1}{\tau}\int_0^T \|\mathsf{u}_{k(t)}\|^2 = \frac{1}{\tau}\|\mathsf{ct}(\vec{\mathsf{u}})\|_{\mathcal{L}_2(\mathcal{U})}^2$. Hence, in particular, $\|\nabla\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \le \frac{1}{\sqrt{\tau}}\|\nabla\mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2} + \sqrt{T}\tau e^{\tau L_f}\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}L_\pi\kappa_\nabla$. From this, and from using Lemma A.1 to bound $\kappa_{\pi,\infty},\kappa_{\pi,1} = \mathcal{O}_\pi(1)$, we obtain the following corollary, which is a precise statement of Proposition 4.2.

Corollary A.1. *Suppose* π *is feasible. Then, recalling* κ_∇ *from Eq. (A.5),*

$$\|\nabla\mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \le \frac{1}{\sqrt{\tau}}\|\nabla\mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2} + \sqrt{T}\tau e^{\tau L_f}\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}L_\pi\kappa_\nabla.$$

In particular, for $\tau \le 1/4L_f$,

$$\|\nabla \mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})} \le \frac{1}{\sqrt{\tau}}\|\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2} + \sqrt{T}\tau \cdot \underbrace{2\max\{\kappa_{\pi,\infty}, \kappa_{\pi,1}, 1\}L_\pi \kappa_\nabla}_{=\mathcal{O}_\pi(1)}.$$

A.6. Main Taylor Expansion Results

We now state various bounds on Taylor-expansion like terms. All the following results are proven in Appendix D. The first is a Taylor expansion of the dynamics (proof in Appendix D.1).

Proposition A.5. *Let* π *be feasible,* $\tau \le \tau_{\mathrm{tay},\pi}$ *Fix a* $\mathtt{u}_{1:K} \in \mathbf{U}$*, and define the perturbation* $\delta\mathtt{u}_{1:K} := \mathtt{u}_{1:K} - \mathtt{u}^\pi_{1:K}$*, and define*

$$B_2 := \sqrt{\tau}\|\delta\mathtt{u}_{1:K}\|_{\ell_2}, B_\infty := \max_k \|\delta\mathtt{u}_k\|.$$

Then, if $B_\infty \le R_{\mathrm{feas}}/8$*, and if for either* $q \in \{2,\infty\}$*, it holds that* $B_q \le B_{\mathrm{tay},q,\pi}$*, then*

(a) The following bounds hold for all $k \in [K+1]$

$$\|\tilde{\mathtt{x}}^\pi_k(\mathtt{u}_{1:K}]) - \mathtt{x}^\pi_k - \sum_{j=1}^{k-1}\Psi^\pi_{\mathrm{cl},k,j}\delta\mathtt{u}_j\| \le M_{\mathrm{tay},q,\pi}B_q^2, \quad \|\tilde{\mathtt{x}}^\pi_k(\mathtt{u}_{1:K}) - \mathtt{x}^\pi_k\| \le L_{\mathrm{tay},q,\pi}B_q,$$

(b) Moreover, for all $k \in [K+1]$ *and* $t \in [0,T]$*,*

$$\max\{\|\tilde{\mathtt{x}}^\pi_k(\mathtt{u}_{1:K})\|, \|\tilde{\mathtt{u}}^\pi_k(\mathtt{u}_{1:K}])\} \le \frac{3R_{\mathrm{feas}}}{4}, and \|\tilde{\mathbf{x}}^\pi(t \mid \mathtt{u}_{1:K})\| \le R_{\mathrm{feas}}.$$

Next, we provide a Taylor expansion of the discrete-time cost functional (proof in Appendix D.2).

Lemma A.6. *Consider the setting of Proposition A.5, and suppose* $B_\infty \le R_{\mathrm{feas}}/8$ *and* $B_2 \le B_{\mathrm{tay},2,\pi}$*. Then,*

$$\|\mathcal{J}_T^{\pi,\mathrm{disc}}(\delta\mathtt{u}_{1:K} + \mathtt{u}^\pi_{1:K}) - \mathcal{J}_T^{\pi,\mathrm{disc}}(\mathtt{u}^\pi_{1:K}) - \langle\delta\mathtt{u}_k, \nabla\mathcal{J}_T^{\pi,\mathrm{disc}}(\mathtt{u}^\pi_{1:K})\rangle\| \le M_{\mathcal{J},\mathrm{tay},\pi}B_2^2.$$

Next, we show sufficiently small perturbations of the nomimal input preserve stability of the dynamics (proof in Appendix D.3).

Lemma A.7. *Again consider the setting of Proposition A.5, and suppose* $B_\infty \le \min\{R_{\mathrm{feas}}/8, B_{\mathrm{tay,inf},\pi}, B_{\mathrm{stab},\pi}\}$*. Then,*

$$\mu_{\pi',\star} \le (1 + B_\infty/B_{\mathrm{stab},\pi})\mu_{\pi,\star} \le 2\mu_{\pi,\star}, \quad L_{\pi'} = L_\pi.$$

Lastly, we bound the norm of the discretized gradient (Appendix D.4).

Lemma A.8. *Let* π *be feasible, and let* $\tau \le \tau_{\mathrm{dyn}}$*. Then*

$$\max_{k\in[K]}\|(\nabla\mathcal{J}_T^{\mathrm{disc}}(\pi))_k\| \le \tau L_{\nabla,\pi,\infty}$$

A.7. Estimation Errors

In this section, we bound the various estimation errors. All the proofs are given in Appendix E. We begin with a simple condition we need for estimation of Markov parameters to go through.

Definition A.5. We say π is *estimation-friendly if* π is feasible, and if

$$\sigma_{\mathrm{orac}}\sqrt{\frac{\iota(\delta)}{2NL_\pi}} \le \sigma_w \le \frac{B_{\mathrm{tay,inf},\pi}}{2\sqrt{d_\star}}, \quad \tau \le \tau_{\mathrm{tay},\pi}$$

Our first result is recovery of the nominal trajectory and Markov operators. Recovery of the nominal trajectory follows from Gaussian concentration, and recovery of the Markov operator for the Matrix Hoeffding inequality (Tropp (2012, Theorem 1.4)) combined with the Taylor expansion of the dynamics due to Proposition A.5. The following is proven in Appendix E.1. To state the bound, we recall the estimation error terms in Definition A.4.

Proposition A.9. *Fix $\delta \in (0,1)$ and suppose that N is large enough that π is estimation friendly. Then, for any estimation-friendly* $\textsc{EstMarkov}(\pi; N, \sigma_w)$ *(Algorithm 2) returns estimates with such that, with probability $1 - \delta/2n_{\mathrm{iter}}$.*

$$\max_{1 \le j < k \le K+1} \left\| \Psi^{\pi}_{\mathrm{cl},k,j} - \hat{\Psi}_{k,j} \right\|_{\mathrm{op}} \le \mathrm{Err}_{\Psi,\pi}(\delta) \qquad \max_{k \in [K+1]} \|\hat{\mathtt{x}}_k - \mathtt{x}^{\pi}_k\| \le \mathrm{Err}_{\hat{x}}(\delta) \tag{A.6}$$

Let $\Pi_{\mathrm{alg}} := \{\pi^{(n)}, \tilde{\pi}^{(n)} : n \in [n_{\mathrm{iter}}]\}$ denote the set of policies constructed by the algorithm, and note that $\textsc{EstMarkov}$ is called once for each policy in Π_{alg}. We define the good estimation event as

$$\mathcal{E}_{\mathrm{est}}(\delta) := \bigcap_{n=1}^{\infty} (\mathcal{E}_n(\delta) \cap \tilde{\mathcal{E}}_n(\delta)), \tag{A.7}$$

$$\mathcal{E}_n(\delta) := \{\text{Eq. (A.6) holds for } \pi = \pi^{(n)} \text{ if } \pi^{(n)} \text{ is estimation friendly}\} \tag{A.8}$$

$$\tilde{\mathcal{E}}_n(\delta) := \{\text{Eq. (A.6)holds for } \tilde{\pi} = \pi^{(n)} \text{ if } \tilde{\pi}^{(n)} \text{ is estimation-friendly}\} \tag{A.9}$$

By Proposition A.9 and a union bound implies

$$\mathbb{P}[\mathcal{E}_{\mathrm{est}}(\delta)] \ge 1 - \delta.$$

We now show that on the good estimation event, the error of the gradient is bounded. The proof is Appendix E.2.

Lemma A.10 (Gradient Error)**.** *On the event $\mathcal{E}_{\mathrm{est}}(\delta)$, it holds that that if $\pi^{(n)}$ is estimation-friendly, then Algorithm 1(Line 4) produces*

$$\max_k \|\hat{\nabla}^{(n)}_k - (\nabla \mathcal{J}^{\mathrm{disc}}_T(\pi^{(n)}))_k\| \le \mathrm{Err}_{\nabla,\pi^{(n)}}(\delta).$$

We also bound the error in the recovery of the system paramters used for synthesizing the stabilizing gains. Recovery of said parameters requires first establishing controllability of the discrete-time Markov operator. We prove the following in Appendix E.3:

Proposition A.11. *Define $\gamma_{\mathrm{ctr}} := \max\{1, L_f t_{\mathrm{ctrl}}\}$, and suppose that $\tau \le \min\{\tau_{\mathrm{ctrl},\pi}, \tau_{\mathrm{dyn}}\}$. Then, for $k \ge k_{\mathrm{ctrl}} + 1$, it holds that*

$$\lambda_{\min}\left(\sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \Psi^{\pi}_{\mathrm{cl},k,j} (\Psi^{\pi}_{\mathrm{cl},k,j})^{\top} \right) \succeq \tau \cdot \frac{\nu_{\mathrm{ctrl}}}{8 L_\pi^2 \gamma_{\mathrm{ctr}}^2 \exp(2\gamma_{\mathrm{ctr}})}$$

With this result, Appendix E.4 upper bounds the estimation error for the discrete-time system matrices.

Proposition A.12. *Suppose $\mathcal{E}_{\mathrm{est}}(\delta)$ holds, fix $n \in n_{\mathrm{iter}}$, and let $\pi = \tilde{\pi}^{(n)}$. Then, suppose that $\tau \le \min\{\tau_{\mathrm{ctrl},\pi}, \tau_{\mathrm{dyn}}\}$, $k_0 \ge k_{\mathrm{ctrl}} + 2$, and*

$$\mathrm{Err}_{\Psi}(\delta) \le \tau \frac{\sqrt{\nu_{\mathrm{ctrl}}/t_{\mathrm{ctrl}}}}{2\sqrt{2} L_\pi \gamma_{\mathrm{ctr}} \exp(\gamma_{\mathrm{ctr}})}, \tag{A.10}$$

Then, on $\mathcal{E}_{\mathrm{est}}(\delta)$, if π is estimation-friendly, the estimates from the call of $\textsc{EstGains}(\pi; N, \sigma)$ *satisfy*

$$\|\hat{\mathtt{B}}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\| \vee \|\hat{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{ol},k}\| \le \frac{\mathrm{Err}_{\Psi,\pi}(\delta)}{\tau} \cdot t_0 \kappa_{\pi,\infty} L_\pi^2 \frac{192 \gamma_{\mathrm{ctr}}^3 \exp(2\gamma_{\mathrm{ctr}})}{\nu_{\mathrm{ctrl}}}.$$

A.8. Descent and Stabilization

In this section, we leverage the estimation results in the previous section to demonstrate the two key features of the algorithm: descent on the discrete-time objective, and stability after the synthesized gains. We begin with a standard first-order descent lemma, whose proof is given in Appendix H.1. This lemma also ensures, by invoking Lemma A.1, that the step size is sufficently small to control the stability of $\tilde{\pi}^{(n)}$, which uses the same gains as $\pi^{(n)}$ but has a slightly perturbed control input.

Lemma A.13 (Descent Lemma). *Suppose $\pi = \pi^{(n)}$ is estimation friendly, let $M \geq M_{\mathcal{J},\mathrm{tay},\pi}$, and suppose*

$$\eta \leq \frac{1}{4M}, \quad (\eta(L_{\nabla,\pi,\infty} + \frac{1}{\tau}\mathrm{Err}_{\nabla,\pi^{(n)}}(\delta)) + \mathrm{Err}_{\hat{x}}(\delta)) \leq \min\left\{\frac{R_{\mathrm{feas}}}{8}, B_{\mathrm{stab},\pi}, B_{\mathrm{tay,inf},\pi}, \frac{B_{\mathrm{tay},2,\pi}}{\sqrt{T}}\right\}.$$

Then, on event $\mathcal{E}_{\mathrm{est}}(\delta)$, it holds (again setting $\pi \leftarrow \pi^{(n)}$ on the right-hand side)

$$\mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi}^{(n)}) - \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)}) \leq -\frac{\eta}{2\tau}\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + T\left(\frac{\mathrm{Err}_{\nabla,\pi}(\delta)^2}{4\tau^2 M} + \mathrm{Err}_{\hat{x}}(\delta)L_{\nabla,\pi,\infty} + M\mathrm{Err}_{\hat{x}}(\delta)^2\right).$$

and that

$$L_{\tilde{\pi}^{(n)}} = L_{\pi^{(n)}}, \quad \mu_{\tilde{\pi}^{(n)},\star} \leq 2\mu_{\pi^{(n)},\star}.$$

The next step is to establish a stability guarantee for the certainty-equivalent gains synthesized. We begin with a generic guarantee, whose proof is given in Appendix G.

Proposition A.14 (Certainty Equivalence Bound). *Let $\hat{\mathtt{A}}_k^\pi$ and $\hat{\mathtt{B}}_k^\pi$ be estimates of $\mathtt{A}_{\mathrm{ol},k}^\pi$ and $\mathtt{B}_{\mathrm{ol},k}^\pi$, and let $\hat{\mathtt{K}}_k$ denote the corresponding certainty equivalence controller synthesized by Algorithm 3(Lines 7 and 10). Suppose that $\tau \leq \min\{\tau_{\mathrm{ric}}, \tau_{\mathrm{dyn}}\}$ and*

$$\max_{k\in[k_0:K]} \|\hat{\mathtt{A}}_k^\pi - \mathtt{A}_{\mathrm{ol},k}^\pi\|_{\mathrm{op}} \vee \|\hat{\mathtt{B}}_k^\pi - \mathtt{B}_{\mathrm{ol},k}^\pi\|_{\mathrm{op}} \leq \tau(2^{17}\mu_{\mathrm{ric}}^4 \max\{1, L_f^3\})^{-1}$$

Then, if $\pi' = (\mathtt{u}_{1:K}^\pi, \hat{\mathtt{K}}_{1:K})$, we have

$$\mu_{\pi',\star} \leq 4\mu_{\mathrm{ric}}, \quad L_{\pi'} \leq 6\max\{1, L_f\}\mu_{\mathrm{ric}}.$$

As a direct corollary of the above proposition and Proposition A.12, we obtain the following:

Lemma A.15. *Suppose $\mathcal{E}_{\mathrm{est}}(\delta)$ holds, fix $n \in n_{\mathrm{iter}}$, and let $\pi = \tilde{\pi}^{(n)}$. Then, suppose that $\tau \leq \min\{\tau_{\mathrm{ctrl},\pi}, \tau_{\mathrm{dyn}}\}$, π is estimation-friendly, $k_0 \geq k_{\mathrm{ctrl}} + 2$, and*

$$\mathrm{Err}_{\Psi,\pi}(\delta) \leq \tau^2\left(2^{25}\mu_{\mathrm{ric}}^4 \max\{1, L_f^3\} \cdot t_0\kappa_{\pi,\infty}L_\pi^2 \frac{\gamma_{\mathrm{ctr}}^3 \exp(2\gamma_{\mathrm{ctr}})}{\nu_{\mathrm{ctrl}}}\right)^{-1} \tag{A.11}$$

Then,

$$\mu_{\pi^{(n+1)},\star} \leq 4\mu_{\mathrm{ric}}, \quad L_{\pi^{(n+1)}} \leq 6\max\{1, L_f\}\mu_{\mathrm{ric}}.$$

Proof. One can check that, as $\mu_{\mathrm{ric}}, L_\pi \geq 1$, Eq. (A.11) implies Eq. (A.10). Thus, the lemma follows directly from Propositions A.12 and A.14, as well as noting $192 \cdot 2^{17} \leq 2^{25}$ □

A.9. Concluding the proof.

In this section, we conclude the proof. First, we define uniform upper bounds on all π-dependent parameters.

Uniform upper bounds on parameters. To begin, define

$$\bar{\mu} = 8\mu_{\mathrm{ric}}, \quad \bar{L} = 6\max\{L_f, 1\}\mu_{\mathrm{ric}}. \tag{A.12}$$

Next, for $q \in \{1, 2, \infty\}$, define $\bar{\kappa}_q := \kappa_q(\bar{\mu}, \bar{L})$ defined in Definition A.1. We define $\bar{\tau}_{\mathrm{tay}}, \bar{\tau}_{\mathrm{ctrl}}$ alogously to $\tau_{\mathrm{tay},\pi}, \tau_{\mathrm{ctrl},\pi}$ in Definition A.2 with $\kappa_{\pi,\infty}$ replaced by $\bar{\kappa}_\infty$ and L_π with $\bar{L}$. For $q \in \{2, \infty\}$, we define $\bar{M}_{\mathrm{tay},q}, \bar{M}_{\mathcal{J},\mathrm{tay}}, \bar{L}_{\mathrm{tay},q}, \bar{L}_{\nabla,\infty}, \bar{B}_{\mathrm{tay},q}, \bar{B}_{\mathrm{stab}}$ analogously to $M_{\mathrm{tay},q,\pi}, M_{\mathcal{J},\mathrm{tay},\pi}, L_{\mathrm{tay},q,\pi}, L_{\nabla,\pi,\infty}, B_{\mathrm{tay},q,\pi}B_{\mathrm{stab},\pi}$ in Definition A.3, with all occurences of $\kappa_{\pi,\infty}, \kappa_{\pi,1}, \kappa_{\pi,2}$ replaced by $\bar{\kappa}_\infty, \bar{\kappa}_1, \bar{\kappa}_2$ and all occurences of L_π replaced by $\bar{L}$. Finally,

we define $\overline{\mathrm{Err}}_\Psi, \overline{\mathrm{Err}}_\nabla$ to be analogous to $\mathrm{Err}_{\Psi,\pi}, \mathrm{Err}_{\nabla,\pi}$ but with the same above substitutions. From Lemmas A.1 and A.2, we have

$$\begin{aligned}
&\bar{\kappa}_q, \bar{M}_{\mathrm{tay},q}, \bar{L}_{\mathrm{tay},q}, \bar{L}_{\nabla,\infty}, \bar{B}_{\mathrm{tay},q} = \mathcal{O}_\star(1)\\
&\tau_{\mathrm{dyn}}, \tau_{\mathrm{ric}}, \bar{\tau}_{\mathrm{tay}}, \bar{\tau}_{\mathrm{ctrl}} = 1/\mathcal{O}_\star(1)\\
&\bar{M}_{\mathcal{J},\mathrm{tay}} = T \cdot \mathcal{O}_\star(1).\\
&\bar{B}_{\mathrm{stab}} = \frac{1}{T}\mathcal{O}_\star(1)
\end{aligned}$$

Moreover, recalling $\mathrm{Err}_0(\delta) := \sqrt{\iota(\delta)/N}$, and setting $\sigma_w = c\sqrt{\mathrm{Err}_0\sigma_{\mathrm{orac}}}$ for any $c \in [1/\mathcal{O}_\star(1), \mathcal{O}_\star(1)]$, Lemma A.3 gives

$$\begin{aligned}
\mathrm{Err}_{\hat{x}}(\delta) &= \mathcal{O}_\star(\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta))\\
\mathrm{Err}_{\Psi,\pi}(\delta) &= \mathcal{O}_\star(\mathrm{Err}_0(\delta) + \sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)})\\
\mathrm{Err}_{\nabla,\pi}(\delta) &= \mathcal{O}_\star(T(\mathrm{Err}_0(\delta) + \sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)})).
\end{aligned} \tag{A.13}$$

Statement of Main Guarantee, Explicit Constants. We begin by stating our main guarantee, first with explicit constants. We then translate into a $\mathcal{O}_\star(1)$ notation. To begin, define the following descent error term:

$$\overline{\mathrm{Err}}_{\mathrm{dec}}(\delta) := T\left(\frac{\overline{\mathrm{Err}}_\nabla(\delta)^2}{4\tau^2\bar{M}_{\mathcal{J},\mathrm{tay}}} + \mathrm{Err}_{\hat{x}}(\delta)\bar{L}_{\nabla,\infty} + \bar{M}_{\mathcal{J},\mathrm{tay}}\mathrm{Err}_{\hat{x}}(\delta)^2\right) \tag{A.14}$$

And note that for $\sigma_w = c\sqrt{\mathrm{Err}_0\sigma_{\mathrm{orac}}}$ for $c \in [1/\mathcal{O}_\star(1), \mathcal{O}_\star(1)]$ (using numerous simplifications, such as $T/\tau \geq 1$)

$$\overline{\mathrm{Err}}_{\mathrm{dec}}(\delta) := \mathcal{O}_\star(1) \cdot \left(\frac{T^3}{\tau^2}\left(\mathrm{Err}_0(\delta)^2 + \sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)\right) + T\sigma_{\mathrm{orac}}^2\mathrm{Err}_0(\delta)^2\right).$$

Theorem 3. *Fix $\delta \in (0,1)$, and suppose that $\eta \leq \frac{1}{4\bar{M}_{\mathcal{J},\mathrm{tay}}}$, $k_0 \geq k_{\mathrm{ctrl}} + 2$, and suppose*

$$\sigma_{\mathrm{orac}}\sqrt{\frac{\iota(\delta)}{2N\bar{L}}} \leq \sigma_w \leq \frac{\bar{B}_{\mathrm{tay},\infty}}{2\sqrt{d_\star}}, \tag{A.15a}$$

$$(\eta(\bar{L}_{\nabla,\infty} + \frac{1}{\tau}\overline{\mathrm{Err}}_\nabla(\delta)) + \mathrm{Err}_{\hat{x}}(\delta)) \leq \min\left\{\frac{R_{\mathrm{feas}}}{8}, \bar{B}_{\mathrm{stab}}, \bar{B}_{\mathrm{tay},\infty}, \frac{\bar{B}_{\mathrm{tay},2}}{\sqrt{T}}\right\} \tag{A.15b}$$

$$\overline{\mathrm{Err}}_\Psi(\delta) \leq \tau^2\left(2^{25}\mu_{\mathrm{ric}}^4\max\{1, L_f^3\} \cdot t_{\mathrm{ctrl}}\kappa_{\pi,\infty}\bar{L}^2\frac{\gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}})}{\nu_{\mathrm{ctrl}}}\right)^{-1} = \frac{\tau^2}{\mathcal{O}_\star(1)} \tag{A.15c}$$

$$\tau \leq \min\{\bar{\tau}_{\mathrm{tay}}, \bar{\tau}_{\mathrm{ctrl}}, \tau_{\mathrm{ric}}\} = \frac{1}{\mathcal{O}_\star(1)} \tag{A.15d}$$

Then, for $\pi = \pi^{(n_{\mathrm{out}})}$ returned by Algorithm 1 satisfies all four properties with probability $1-\delta$:

(a) $\mu_{\pi,\star} \leq 8\mu_{\mathrm{ric}}$ and $L_\pi \leq 6\max\{1,\kappa_f\}\mu_{\mathrm{ric}} = \bar{L}$. In fact, for all $n \in [n_{\mathrm{iter}}]$, and $\pi' \in \{\pi^{(n)}, \tilde{\pi}^{(n)}\}$, $\mu_{\pi',\star} \leq \bar{\mu} = 8\mu_{\mathrm{ric}}$ and $L_{\pi'} \leq \bar{L} = 6\max\{1, L_f\}\mu_{\mathrm{ric}}$.

(b) The discrete-time stabilized gradient is bounded by

$$\begin{aligned}
\tau\|\mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2}^2 &\leq 2T\overline{\mathrm{Err}}_\nabla(\delta)^2 + \frac{2}{\eta}\left(\frac{2(1+T)\kappa_{\mathrm{cost}}}{n_{\mathrm{iter}}} + \overline{\mathrm{Err}}_{\mathrm{dec}}(\delta)\right)\\
&= \frac{1}{\eta}\mathcal{O}_\star(1) \cdot \left(\frac{T}{n_{\mathrm{iter}}} + (\eta T^3 + \frac{T^3}{\tau^2})\left(\mathrm{Err}_0(\delta)^2 + \sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)\right) + T\sigma_{\mathrm{orac}}^2\mathrm{Err}_0(\delta)^2\right),
\end{aligned}$$

where the last line holds when $\sigma_w = c\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}$ for some $c \in [\frac{1}{\mathcal{O}_\star(1)}, \mathcal{O}_\star(1)]$.

(c) Recall $\kappa_\nabla := \Big((1+L_f)M_{\mathrm{cost}}(1+\kappa_f)+L_{\mathrm{cost}}(3\kappa_f M_f+8L_f^2+L_f)\Big) = \mathcal{O}_\star(1)$ *from Eq. (A.5). Then* π *is* ϵ*-stationary for*

$$\begin{aligned}\epsilon^2 &= 4T\overline{\mathrm{Err}}_\nabla(\delta)^2+\frac{4}{\eta}\left(\frac{2(1+T)\kappa_{\mathrm{cost}}}{n_{\mathrm{iter}}}+\overline{\mathrm{Err}}_{\mathrm{dec}}(\delta)\right)+4T\tau^2\left(\max\{\bar\kappa_\infty,\bar 1_\infty,1\}\bar L\kappa_\nabla\right)^2\\ &= \mathcal{O}_\star(1)\cdot T\left(\tau^2+\frac{1}{\eta}\left(\frac{1}{n_{\mathrm{iter}}}+(\eta T^2+\frac{T^2}{\tau^2})\left(\mathrm{Err}_0(\delta)^2+\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)\right)+\sigma_{\mathrm{orac}}^2\mathrm{Err}_0(\delta)^2\right)\right).\end{aligned}$$

where the last line holds when $\sigma_w = c\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}$ *for some* $c\in[\frac{1}{\mathcal{O}_\star(1)},\mathcal{O}_\star(1)]$.

(d) $\mathbf{u}^\pi$ *is an* ϵ'-JSP*, where* $\epsilon' = 64\epsilon^2\bar L^2/\alpha = \mathcal{O}_\star(1)\cdot\frac{\epsilon^2}{\alpha}$.

We prove Theorem 3 from the above results in Appendix A.9.2 just below. Appendix A.9.1 below translates the above theorem into Theorem 2 which uses $\mathcal{O}_\star(\cdot)$ notation.

A.9.1. Translating Theorem 3 into Theorem 2

Proof. It suffices to translate the conditions Eqs. (A.15a) to (A.15d) into $\mathcal{O}_\star(\cdot)$ notation. Again, recall $\mathrm{Err}_0(\delta) = \sqrt{\iota(\delta)/N}$, and take $\sigma_w = c\sqrt{\mathrm{Err}_0(\delta)\sigma_{\mathrm{orac}}}$ for $c\in[1/\mathcal{O}_\star(1),\mathcal{O}_\star(1)]$. Then, Eq. (A.15a) holds for $\mathrm{Err}_0(\delta)\le 1/c_1$, where $c_1=\mathcal{O}_\star(1)$. Next, to make Eq. (A.15b) hold, it suffices that

$$\max\left\{(\eta\bar L_{\nabla,\infty},\frac{\eta}{\tau}\overline{\mathrm{Err}}_\nabla(\delta),\mathrm{Err}_{\hat x}(\delta)\right\}\le\frac{1}{3}\min\left\{\frac{R_{\mathrm{feas}}}{8},\bar B_{\mathrm{stab}},\bar B_{\mathrm{tay},\infty},\frac{\bar B_{\mathrm{tay},2}}{\sqrt{T}}\right\},$$

The term $\eta\bar L_{\nabla,\infty}$ is sufficiently bounded where $\eta\le\frac{1}{c_2\sqrt{T}}$ for $c_2=\mathcal{O}_\star(1)$. Recalling $\mathrm{Err}_{\nabla,\pi}(\delta)\le\mathcal{O}_\star(T(\mathrm{Err}_0(\delta)+\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}))$ from Eq. (A.13), and that $\eta\le\frac{1}{c_2\sqrt{T}}$, it is enough that $(\mathrm{Err}_0(\delta)+\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)})\le\frac{c_2\tau}{c_3T}$ for $c_3=\mathcal{O}_\star(1)$. Finally $\mathrm{Err}_{\hat x}(\delta)$ is bounded for $\mathrm{Err}_{\hat x}(\delta)=\mathrm{Err}_0(\delta)\le 1/c_4\sqrt{T}$, where $c_4=\mathcal{O}_\star(1)$. Collecting these conditions, we have that for $c_1,c_2,c_3,c_4=\mathcal{O}_\star(1)$, Eqs. (A.15a) and (A.15b) hold for

$$\eta\le\frac{1}{c_2\sqrt{T}},\quad \mathrm{Err}_0(\delta)+\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}\le\frac{c_2\tau}{c_3T},$$

Next,as $\mathrm{Err}_{\Psi,\pi}(\delta)=\mathcal{O}_\star(\mathrm{Err}_0(\delta)+\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)})$ from Eq. (A.13),*Eq.* (A.15c) holds as long as $\mathrm{Err}_0(\delta)+\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}\le\tau^2/c_5$ for a $c_5=\mathcal{O}_\star(1)$. Combining,

$$\eta\le\frac{1}{c_2\sqrt{T}},\quad \mathrm{Err}_0(\delta)+\sqrt{\sigma_{\mathrm{orac}}\mathrm{Err}_0(\delta)}\le\min\left\{\frac{c_2\tau}{c_3T},\frac{\tau^2}{c_5}\right\}\quad \mathrm{Err}_0(\delta)\le\min\left\{\frac{1}{c_1},\frac{1}{c_4\sqrt{T}}\right\}$$

Finally, Eq. (A.15d) requirs $\tau\le 1/c_6$, for $c_6=\mathcal{O}_\star(1)$, and that $\eta\le 1/c_7$ where $c_7=4\bar M_{\mathcal{J},\mathrm{tay}}=\mathcal{O}_\pi(1)$. By shrinking constants if necessary, this can be simplified into

$$\eta\le\min\{\frac{1}{c_7},\}\frac{1}{c_2\sqrt{T}}\},\quad \mathrm{Err}_0(\delta)\le\min\left\{\min\left\{\frac{c_2\tau}{c_3T},\frac{\tau^2}{c_5},\frac{1}{c_1},\frac{1}{c_4\sqrt{T}}\right\},\frac{1}{\sigma_{\mathrm{orac}}}\min\left\{\frac{c_2\tau}{c_3T},\frac{\tau^2}{c_5}\right\}^2\right\}.$$

And recall $\mathrm{Err}_0(\delta)=\sqrt{\iota(\delta)/N}$, this becomes

$$\tau\le\frac{1}{c_6},\quad \eta\le\frac{1}{c_2}\min\left\{1,\frac{1}{\sqrt{T}}\right\},\quad N\ge\iota(\delta)\min\left\{\min\left\{\frac{1}{c_1},\frac{c_2\tau}{c_3T},\frac{\tau^2}{c_5},\frac{1}{c_4\sqrt{T}}\right\},\frac{1}{\sigma_{\mathrm{orac}}}\min\left\{\frac{c_2\tau}{c_3T},\frac{\tau^2}{c_5}\right\}^2\right\}^{-2}.$$

By consolidating constants and relabeling $c_1,c_2=\mathcal{O}_\star(1)$ as needed, it suffices that

$$\begin{aligned}\eta\le c_1\min\left\{\frac{1}{\sqrt{T}},1\right\},\quad \tau\le\frac{1}{c_2},\quad N&\ge c_3\iota(\delta)\min\left\{\min\left\{1,\frac{\tau}{T},\tau^2,\frac{1}{\sqrt{T}}\right\},\frac{1}{\sigma_{\mathrm{orac}}}\min\left\{\frac{\tau}{T},\tau^2\right\}^2\right\}^{-2}\\ &= c_3\iota(\delta)\max\left\{1,\frac{T^2}{\tau^2},\frac{1}{\tau^4},T,\sigma_{\mathrm{orac}}^2\frac{T^4}{\tau^2},\frac{\sigma_{\mathrm{orac}}^2}{\tau^8}\right\}.\end{aligned}$$

Having shown that the above conditions suffice to ensure Theorem 3 holds, the bound follows (again replacing $\mathrm{Err}_0(\delta)$ with $\sqrt{\iota(\delta)/N}$).

□

A.9.2. Proof of Theorem 3

We shall show the following invariant. At each step n,

$$\mu_{\pi^{(n)},\star} \le \bar{\mu}/2, \quad L_{\pi^{(n)}} \le \bar{L}.. \tag{A.16}$$

Lemma G.3 shows that Eq. (A.16) holds for $n = 1$. Next, for $n \ge 1$, directly combining Lemmas A.13 and A.15 imply the following per-round guarantee.

Lemma A.16 (Per-Round Lemma)**.** *Suppose that* $\eta \le \frac{1}{4\bar{M}_{\mathcal{J},\mathrm{tay}}}$, $k_0 \ge k_{\mathrm{ctrl}} + 2$, *Then if* $\pi^{(n)}$ *satisfies Eq. (A.16) and Eqs. (A.15a) to (A.15d). Then, on* $\mathcal{E}_{\mathrm{est}}(\delta)$,

(a) $\max_k \|\hat{\nabla}_k^{(n)} - (\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)}))_k\| \le \overline{\mathrm{Err}}_\nabla(\delta)$*; thus* $\tau\|\hat{\nabla}_{1:K}^{(n)} - (\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)})\|_{\ell_2}^2 \le T\overline{\mathrm{Err}}_\nabla(\delta)^2$.

(b) The following descent guarantee holds

$$\mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi}^{(n)}) - \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)}) \le -\frac{\eta}{2\tau}\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + \overline{\mathrm{Err}}_{\mathrm{dec}}(\delta)$$

(c) $L_{\tilde{\pi}^{(n)}} = L_{\pi^{(n)}} \le \bar{L}$ *and* $\mu_{\tilde{\pi}^{(n)},\star} \le 2\mu_{\pi^{(n)},\star} \le \bar{\mu}$.

(d) $\mu_{\pi^{(n+1)},\star} \le 4\mu_{\mathrm{ric}} = \bar{L}/2, \quad L_{\pi^{(n+1)}} \le 6\max\{1, L_f\}\mu_{\mathrm{ric}} = \bar{L}.$*; that is* $\pi^{(n+1)}$ *satisfies Eq. (D.4)*

Proof. Part (a) follows from Lemma A.10, and parts (b) and (c) follow from Lemma A.13, with the necessary replacement of π-dependent terms wither $\overline{(\cdot)}$ terms. Part (b) allows us to make the same substiutions in Lemma A.15, which gives part (c). □

Proof of Theorem 3. Under the conditions of this lemma, Lemma A.16 holds. As Lemma G.3 shows that Eq. (A.16) holds for $n = 1$, induction implies Lemma G.3 holds for all $n \in [n_{\mathrm{iter}}]$ on $\mathcal{E}_{\mathrm{est}}(\delta)$, an event which occurs with probability $1 - \delta$. We now prove each part of the present theorem in sequence.

Part (a). Directly from Lemma A.16(d)

Part (b). Notice that, since $\tilde{\pi}^{(n)}$ and $\pi^{(n+1)}$ differ only in their gains, $\mathcal{J}_T^{\mathrm{disc}}(\pi^{(n+2)}) = \mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi}^{(n)})$. Therefore, summing up the descent guarantee in Lemma A.16(b), we have

$$\begin{aligned}
\mathcal{J}_T^{\mathrm{disc}}(\pi^{(n_{\mathrm{iter}}+1)}) - \mathcal{J}_T^{\mathrm{disc}}(\pi^{(1)}) &\le -\frac{\eta}{2\tau}\sum_{n=1}^{n_{\mathrm{iter}}}\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + n_{\mathrm{iter}}\overline{\mathrm{Err}}_{\mathrm{dec}}(\delta) \\
&\le -\frac{\eta}{2\tau}n_{\mathrm{iter}}\min_{n\in[n_{\mathrm{iter}}]}\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + n_{\mathrm{iter}}\overline{\mathrm{Err}}_{\mathrm{dec}}(\delta) \\
&= -\frac{\eta}{2\tau}\|\hat{\nabla}_{1:K}^{(n_{\mathrm{out}})}\|_{\ell_2}^2 + n_{\mathrm{iter}}\overline{\mathrm{Err}}_{\mathrm{dec}}(\delta)
\end{aligned}$$

where we recall that our algorithm selects to output $\pi^{(n_{\mathrm{out}})}$, where n_{out} minimizes $\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2$. Recall that $\mathcal{J}_T^{\pi,\mathrm{disc}}(\pi) = \mathcal{J}_T^{\pi,\mathrm{disc}}(\mathtt{u}_{1:K}^\pi)$ where $\mathcal{J}_T^{\pi,\mathrm{disc}}(\vec{\mathtt{u}}) := V(\tilde{\mathtt{x}}_{K+1}^\pi(\vec{\mathtt{u}})) + \tau\sum_{k=1}^K Q(\tilde{\mathtt{x}}_k^\pi(\vec{\mathtt{u}}), \tilde{\mathtt{u}}_k^\pi(\vec{\mathtt{u}}), t_k)$. Hence, for all feasible π, Assumption 4.2 implies $0 \le \mathcal{J}_T^{\pi,\mathrm{disc}}(\pi) \le \kappa_{\mathrm{cost}}(1 + \tau K) = (1+T)\kappa_{\mathrm{cost}}$. By Condition 4.1, $\pi^{(n+1)}$ and $\pi^{(1)}$ are by feasible, and thus $\mathcal{J}_T^{\mathrm{disc}}(\pi^{(n+1)}) - \mathcal{J}_T^{\mathrm{disc}}(\pi^{(1)}) \ge -(1+T)\kappa_{\mathrm{cost}}$. Therefore, by rearranging the previous display,

$$\tau\|\hat{\nabla}_{1:K}^{(n_{\mathrm{out}})}\|_{\ell_2}^2 \le \frac{1}{\eta}\left(\frac{2(1+T)\kappa_{\mathrm{cost}}}{n_{\mathrm{iter}}} + \overline{\mathrm{Err}}_{\mathrm{dec}}(\delta)\right).$$

By Lemma A.16(a), and AM-GM imply then

$$\tau\|\mathcal{J}_T^{\mathrm{disc}}(\pi^{(n_{\mathrm{out}})})\|_{\ell_2}^2 \le 2T\overline{\mathrm{Err}}_\nabla(\delta)^2 + \frac{2}{\eta}\left(\frac{2(1+T)\kappa_{\mathrm{cost}}}{n_{\mathrm{iter}}} + \overline{\mathrm{Err}}_{\mathrm{dec}}(\delta)\right).$$

Part (c). Note that $\tau \le \bar{\tau}_{\mathrm{tay}}$ implies $\tau \le 1/4L_f$. From Corollary A.1, and for $\kappa_\nabla = \mathcal{O}_\star(1)$ as in Eq. (A.5), the following holds for any feasible π:

$$\begin{aligned}\|\nabla \mathcal{J}_T(\pi)\|_{\mathcal{L}_2(\mathcal{U})}^2 &\le \left(\frac{1}{\sqrt{\tau}}\|\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2} + \sqrt{T}\tau \cdot 2\max\{\kappa_{\pi,\infty}, \kappa_{\pi,1}, 1\} L_\pi \kappa_\nabla\right)^2 \\ &\le \frac{2}{\tau}\|\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi)\|_{\ell_2}^2 + 4T\tau^2 \left(\max\{\kappa_{\pi,\infty}, \kappa_{\pi,1}, 1\} L_\pi \kappa_\nabla\right)^2.\end{aligned}$$

Apply the above with $\pi = \pi^{(n_{\mathrm{out}})}$ gives part c, and upper bound $\kappa_{\pi,\infty}, \kappa_{\pi,1}, L_\pi$ by $\bar{\kappa}_\infty, \bar{\kappa}_1, \bar{L}$ concludes.

Part (d). This follows directly from Proposition 4.1, noting that $L_\pi \le \bar{L}$ for $\pi = \pi^{(n_{\mathrm{out}})}$, and that $\bar{\tau}_{\mathrm{tay}} = \frac{1}{16\bar{L}L_f}$, so that the step-size condition of Proposition 4.1 is met. □

B. Discussion and Extensions

B.1. Separation between and Open-Loop and Closed-Loop Gradients

In this section, we provided an illustrative example as to why a approximation JSP is more natural than canonical stationary points. Fix an $\epsilon \in (0, 1]$, and consider the system with dynamic map

$$f_\epsilon(x, u) = 2x + u - \epsilon.$$

Let $\mathbf{x}_\epsilon(t \mid \mathbf{u})$ denote the scalar trajectory with

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}_\epsilon(t \mid \mathbf{u}) = f_\epsilon(\mathbf{x}_\epsilon(t \mid \mathbf{u}), \mathbf{u}(t)), \quad \mathbf{x}_\epsilon(0 \mid \mathbf{u}) = \epsilon.$$

Then, $\mathbf{x}_\epsilon(t \mid \mathbf{0}) = \epsilon$ for all t. We can now consider the following planning objective

$$\mathcal{J}_{T,\epsilon}(\mathbf{u}) = \frac{1}{2}\int_0^T \left(\mathbf{x}_\epsilon(t \mid \mathbf{u})^2 + \mathbf{u}(t)^2\right) \mathrm{d}t. \tag{B.1}$$

Since the dynamics f_ϵ are affine, we find that

$$\mathbf{u} \text{ is an } \epsilon'\text{-JSP of } \mathcal{J}_{T,\epsilon} \quad \Longleftrightarrow \quad \mathbf{u} \le \inf_{\mathbf{u}'} \mathcal{J}_{T,\epsilon}(\mathbf{u}') + \epsilon'.$$

In particular, as $\mathcal{J}_{T,\epsilon}(\mathbf{0}) = \frac{1}{2}T\epsilon^2$, and as $\mathcal{J}_{T,\epsilon} \ge 0$,

$$\mathbf{u} = 0 \text{ is an } \frac{T\epsilon^2}{2}\text{-JSP of } \mathcal{J}_{T,\epsilon}. \tag{B.2}$$

However, we show that the magnitude of the gradient at $\mathbf{u} = 0$ is much larger. We compute the following shortly below.

Lemma B.1. *For $T \ge 1$, we have $\|\nabla \mathcal{J}_{T,\epsilon}(\mathbf{0})(t)\|_{\mathcal{L}_2(\mathcal{U})} \ge \sqrt{T}\epsilon e^T/4\sqrt{2}$.*

Thus, the magnitude of the gradient (through open-loop dynamics) is exponentially larger than the suboptimality of the cost. This suggests that gradients through open-loop dynamics are poor proxy for global optimality, motivating instead the JSP. Moreover, one can easily compute that if π has inputs $\mathtt{u}_k^\pi = 0$ and stabilizing gains $\mathtt{K}_k = -3$, then for sufficiently small step sizes, the gradients of $\mathbf{J}_T^\pi(\mathbf{u})\big|_{\mathbf{u}=0}$ scale only as $c\epsilon\sqrt{T}$ for a universal $c > 0$, and do not depend exponentially on the horizon.

Proof of Lemma B.1. We have from Lemma C.5 that

$$\nabla \mathcal{J}_{T,\epsilon}(\mathbf{0})(t) = \int_{s=t}^T \underbrace{\mathbf{x}_\epsilon(s \mid \mathbf{0})}_{=\epsilon} \cdot \Phi(s, t)\mathbf{B}(t),$$

where $\Phi(s,t)$ solves the ODE $\Phi(t,t) = 1$ and $\frac{\mathrm{d}}{\mathrm{d}s}\Phi(s,t) = 2\Phi(s,t)$. Thus, $\Phi(s,t) = \exp(2(t-s))$. Moreover, $\mathbf{B}(t) = 1$. Hence,

$$\begin{aligned}\nabla\mathcal{J}_{T,\epsilon}(\mathbf{0})(t) &= \epsilon\int_{s=t}^{T}\exp(2(t-s)) \\ &= \epsilon\frac{1}{2}(e^{2(T-t)}) - 1)\end{aligned}$$

Hence, for $t \le T/2$ and $T \ge 1$,

$$|\nabla\mathcal{J}_{T,\epsilon}(\mathbf{0})(t)| \ge \epsilon\frac{1}{2}(e^T - 1) \ge \epsilon\frac{1}{2}(e^T - 1) \ge \frac{\epsilon e^T}{4}.$$

Hence,

$$\|\nabla\mathcal{J}_{T,\epsilon}(\mathbf{u})(t)\|^2_{\mathcal{L}_2(\mathcal{U})} \ge \frac{T}{2}\left(\frac{\epsilon e^T}{4}\right)^2,$$

so $\|\nabla\mathcal{J}_{T,\epsilon}(\mathbf{u})(t)\|_{\mathcal{L}_2(\mathcal{U})} \ge \sqrt{T}\epsilon e^T/4\sqrt{2}$. □

B.2. Global Stability Guarantees of JSPs and Consequences of (Westenbroek et al., 2021)

(Westenbroek et al., 2021) demonstrate that, for a certain class of nonlinear systems whose Jacobian Linearizations satisfy various favorable properties, an ϵ-FOS point $\mathbf{u}$ of the objective $\mathcal{J}_T$ corresponds to a trajectory which converges exponentially to a desired equilibirum. Examining their proof, the first step follows from Westenbroek et al. (2021, Lemma 2), which establishes that $\mathbf{u}$ is an $\epsilon' = \epsilon^2/2\alpha$-JSP, and it is this property (rather than the ϵ-FOS) that is used throughout the rest of the proof. Hence, their result extends from FOSs to JSPs. Hence, the *local* optimization guarantees established in this work imply, via Westenbroek et al. (2021, Theorem 1), exponentially stabilizing *global* behavior.

B.3. Projections to ensure boundedness.

Let us describe one way to ensure the feasibility condition, Definition 4.7. Suppose that f_{dyn} has the following stability property, which can be thought of as the state-output anologue of BIBO stability, and is common in the control literature (Jadbabaie & Hauser, 2001). For example, we may consider the following assumption.

Assumption B.1. There exists some function $\phi_{\mathrm{sys}} : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ such that that, if $\|\mathbf{u}(t)\| \le R$ for all $t \in [0,T]$, then $\|\mathbf{x}(t \mid \mathbf{u})\| \le \phi_{\mathrm{sys}}(R,T)$ for all $t \in [0,T]$.

Next, fix a bound $R_u > 0$, and set

$$R_{\mathrm{feas}} := 2\max\{R_u, \phi_{\mathrm{sys}}(R_u,T)\}$$

Then, it follows that for any policy for which

$$\mathtt{u}_k^\pi \le R_u \tag{B.3}$$

for all $k \in [K]$ is feasible in the sense of Definition 4.1. We therefore modify Algorithm 1,Line 5 to the *projected gradient step*

$$\mathtt{u}_{1:K}^{(n+1)} \leftarrow \mathsf{Proj}_{(\mathcal{B}_{d_{\mathtt{u}}}(R_u))^K}\left[\mathtt{orac}_{\pi^{(n)},u}(\tilde{\mathtt{u}}_{1:K}^{(n)})\right], \quad \text{where again } \tilde{\mathtt{u}}_k^{(n)} := \mathtt{u}_k^{(n)} - \frac{\eta}{\tau}\hat{\nabla}_k^{(n)} - \mathtt{K}_k^{\pi^{(n)}}\hat{\mathtt{x}}_k,$$

where we let $\mathsf{Proj}_{(\mathcal{B}_{d_{\mathtt{u}}}(R_u))^K}$ denote the orthogonal-projection on the K-fold project of $d_{\mathtt{u}}$-dimensional balls of Euclidean radius R_u, $\mathcal{B}_{d_{\mathtt{u}}}(R_u)$. This projection is explicitly given by

$$\left(\mathsf{Proj}_{(\mathcal{B}_{d_{\mathtt{u}}}(R_u))^K}[\mathtt{u}_{1:K})]\right)_k = \mathtt{u}_k \cdot \min\left\{1, \frac{R_u}{\|\mathtt{u}_k\|}\right\},$$

here using the convention that when $\mathtt{u}_k = 0$, the above evaluates to 0. In this case, our algorithm converges (up to gradient estimation error) to a stationary-point of the projected gradient descent algorithm (see, e.g. the note https://damek.github.io/teaching/orie6300/lec22.pdf for details). We leave the control-theoretic interpretation of such stationary points to future work.

B.4. Extensions to include Process Noise

As explained in Section 2, Oracle 2.1 only adds observation noise but not process noice. Process noise somewhat complicates the analysis, because then our method will only learn the Jacobians dynamics up to a noise floor determined by the process noise. However, by generalization our Taylor expansion of the dynamics (e.g. Proposition A.5), we can show that as the process noise magnitude decreases, we would achieve better and better accuray, recovering the noiseless case in the limit. In addition, process noise may warrant greater algorithmic modifications: for example we may want to incorporate higher-order Taylor expansions of the dynamics (not just the Jacobian linearization), or more sophisticated gradient updates (i.e. `iLQG` ((Todorov & Li, 2005))) better tuned to handle process noise.

B.5. Discussion of the $\exp(L_f t_0)$ dependence.

There are two sources of the exponential dependence on $t_0 = \tau k_0$ that arises in our analysis. First, we translate open-loop controllability (Assumption 4.4) to closed-loop controllbility needed for recovery of system matrices, in an argument based on (Chen & Hazan, 2021), and which incurs dependent on $\exp(L_f t_{\mathrm{ctrl}}) \le \exp(L_f t_0)$. Second, we only consider a stability modulus (Definition 4.7) for a Lyapunov equation terminating at $k = k_0$, because we do not estimate $\mathtt{A}^{\pi}_{\mathrm{ol},k}, \mathtt{B}^{\pi}_{\mathrm{ol},k}$, and therefore cannot synthesize the system gains, for $k \le k_0$. This means that (see Lemma A.1) that many natural bounds on the discretized transition operators $\|\Phi^{\pi}_{\mathrm{cl},k,j}\|$ scale as $\mathrm{poly}(\mu_{\pi,\star}, \exp(t_0 L_f)$, yielding exponential dependence on $t_0 L_f$.

C. Jacobian Linearizations

C.1. Preliminaries

Recall $\mathcal{U}$ denotes the space of continuous-time inputs $\mathbf{u} : [0, T] \to \mathbb{R}^{d_{\mathbf{u}}}$, and U continuous-time inputs $\vec{\mathtt{u}} \in (\mathbb{R}^{d_{\mathbf{u}}})^K$

C.1.1. Exact Trajectories

We recall definitions of various trajectories.

Definition C.1 (Open-Loop Trajectories and Nomimal Trajectories)**.** For a $\mathbf{u} \in \mathcal{U}$, we define $\mathbf{x}(t \mid \mathbf{u})$ as the curve given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t \mid \mathbf{u}) = f_{\mathrm{dyn}}(\mathbf{x}(t \mid \mathbf{u}), \mathbf{u}(t)), \quad \mathbf{x}(0 \mid \mathbf{u}) = \xi_{\mathrm{init}}.$$

For a policy $\pi = (\mathtt{u}^{\pi}_{1:K}, \mathtt{K}^{\pi}_{1:K})$, we define $\mathbf{u}^{\pi} = \mathsf{ct}(\mathtt{u}^{\pi}_{1:K})$, $\mathbf{x}^{\pi}(t) = \mathbf{x}(t \mid \mathbf{u}^{\pi})$, and $\mathtt{x}^{\pi}_k = \mathbf{x}^{\pi}(t_k)$.

Similarly, we present a summary of the definition of various stabilized trajectories, consistent with Definitions 2.1 and 4.4.

Definition C.2 (Stabilized Trajectories)**.** For $\bar{\mathbf{u}} \in \mathcal{U}$ and a policy π, we define continuous-time perturbations of the dynamics with feedback

$$\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) := \mathbf{x}(t \mid \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}), \quad \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) := \bar{\mathbf{u}}(t) + \mathtt{K}^{\pi}_{k(t)}(\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t_{k(t)} \mid \bar{\mathbf{u}})) - \mathtt{x}^{\pi}_{k(t)}),$$

there specialization to discrete-time inputs $\vec{\mathtt{u}} \in \mathsf{U}$

$$\tilde{\mathbf{x}}^{\pi}(t \mid \vec{\mathtt{u}}) := \tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \mathsf{ct}(\bar{\mathbf{u}})), \quad \tilde{\mathbf{u}}^{\pi}(t \mid \vec{\mathtt{u}}) := \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathsf{ct}(\bar{\mathbf{u}})),$$

and their discrete samplings

$$\tilde{\mathtt{x}}^{\pi}_k(\vec{\mathtt{u}}) := \tilde{\mathbf{x}}^{\pi}(t_k \mid \vec{\mathtt{u}}), \quad \tilde{\mathtt{u}}^{\pi}_k(\vec{\mathtt{u}}) := \tilde{\mathbf{u}}^{\pi}(t_k \mid \vec{\mathtt{u}})$$

C.1.2. Trajectory Linearizations

Definition C.3 (Open-Loop Jacobian Linearizations of Trajectories)**.** We define the continuous-time Jacobian linearizations

$$\begin{aligned} \mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}) &:= \mathbf{x}(t \mid \mathbf{u}) + \langle \nabla_{\mathbf{u}} \mathbf{x}(t \mid \mathbf{u})\big|_{\mathbf{u}=\mathbf{u}^{\pi}}, \bar{\mathbf{u}} - \mathbf{u}^{\pi} \rangle_{\mathcal{L}_2(\mathcal{U})} \\ \delta\mathbf{x}^{\mathrm{jac}}(t \mid \mathbf{u}; \bar{\mathbf{u}}) &:= \mathbf{x}^{\mathrm{jac}}(t \mid \mathbf{u}; \bar{\mathbf{u}}) - \mathbf{x}(t \mid \mathbf{u}) \end{aligned} \tag{C.1}$$

Definition C.4 (Closed-Loop Jacobian Linearizations of Trajectories, Discrete-Time)**.**

$$\begin{aligned}\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) &:= \tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \mathbf{u}) + \langle \nabla_{\mathbf{u}} \tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \mathbf{u})\big|_{\mathbf{u}=\mathbf{u}^\pi}, \bar{\mathbf{u}} - \mathbf{u}^\pi \rangle_{\mathcal{L}_2(\mathcal{U})} \\ \tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) &:= \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathbf{u}) + \langle \nabla_{\mathbf{u}} \tilde{\mathbf{u}}^{\pi,\mathrm{ct}}(t \mid \mathbf{u})\big|_{\mathbf{u}=\mathbf{u}^\pi}, \bar{\mathbf{u}} - \mathbf{u}^\pi \rangle_{\mathcal{L}_2(\mathcal{U})}\end{aligned} \tag{C.2}$$

We further define the linearized differences

$$\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) := \tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \mathbf{x}^\pi(t), \quad \delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) := \tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \mathbf{u}^\pi(t) \tag{C.3}$$

Definition C.5 (Jacobian Linearization, with gains, dicrete Time)**.** Given $\vec{\mathtt{u}} \in \mathsf{U}$, we define

$$\begin{aligned}\tilde{\mathtt{x}}^{\pi,\mathrm{jac}}_k(\vec{\mathtt{u}}) &:= \tilde{\mathtt{x}}^\pi_k(\vec{\mathtt{u}}) + \langle \nabla_{\mathbf{u}} \tilde{\mathtt{x}}^\pi_k(t \mid \vec{\mathtt{u}})\big|_{\vec{\mathtt{u}}=\mathtt{u}^\pi_{1:K}}, \bar{\mathbf{u}} - \mathtt{u}^\pi_{1:K} \rangle \\ \tilde{\mathtt{u}}^{\pi,\mathrm{jac}}_k(\vec{\mathtt{u}}) &:= \tilde{\mathtt{u}}^\pi_k(\vec{\mathtt{u}}) + \langle \nabla_{\vec{\mathtt{u}}} \tilde{\mathtt{u}}^\pi_k(t \mid \mathbf{u})\big|_{\vec{\mathtt{u}}=\mathtt{u}^\pi_{1:K}}, \vec{\mathtt{u}} - \mathtt{u}^\pi_{1:K} \rangle_{\mathcal{L}_2(\mathcal{U})}\end{aligned} \tag{C.4}$$

We further define the linearized differences

$$\delta\tilde{\mathtt{x}}^{\pi,\mathrm{jac}}_k(\vec{\mathtt{u}}) := \tilde{\mathtt{x}}^{\pi,\mathrm{jac}}_k(\vec{\mathtt{u}}) - \mathtt{x}^\pi_k, \quad \delta\tilde{\mathtt{u}}^{\pi,\mathrm{jac}}_k(\vec{\mathtt{u}}) := \tilde{\mathtt{u}}^{\pi,\mathrm{jac}}_k(\vec{\mathtt{u}}) - \mathtt{u}^\pi_k \tag{C.5}$$

C.1.3. Jacobian Linearizated Dynamics

We now recall the definitions of various linearizations, consistent with Definition 4.6.

Definition C.6 (Open-Loop, On-Policy Linearized Dynamics)**.** We define the open-loop, on-policy linearization around a policy π via

$$\mathbf{A}^\pi_{\mathrm{ol}}(t) = \partial_x f_{\mathrm{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t)), \quad \mathbf{B}^\pi_{\mathrm{ol}}(t) = \partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t)).$$

Definition C.7 (Open-Loop, On-Policy Linearized Transition, Markov Operators, and Discrete-Dynamics)**.** We define the linearized transition function $\mathbf{\Phi}^\pi_{\mathrm{ol}}(s,t)$ defined for $s > t$ as the solution to $\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{\Phi}^\pi_{\mathrm{ol}}(s,t) = \mathbf{A}^\pi_{\mathrm{ol}}(s)\mathbf{\Phi}^\pi_{\mathrm{ol}}(s,t)$, with initial condition $\mathbf{\Phi}^\pi_{\mathrm{ol}}(t,t) = \mathbf{I}$. We *discretize* the open-loop transition function by define

$$\mathtt{A}^\pi_{\mathrm{ol},k} = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, t_k), \quad \mathtt{B}^\pi_{\mathrm{ol},k} := \int_{s=t_k}^{t_{k+1}} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathrm{d}s.$$

Definition C.8 (Closed-Loop Jacobian Linearization, Discrete-Time)**.** We define a *discrete-time closed-loop* linearization

$$\mathtt{A}^\pi_{\mathrm{cl},k} := \mathtt{A}^\pi_{\mathrm{ol},k} + \mathtt{B}^\pi_{\mathrm{ol},k}\mathtt{K}^\pi_k = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, t) + \int_{s=t}^{t_{k+1}} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathtt{K}^\pi_k,$$

and a discrete closed-loop *transition operator* is defined, for $1 \le k_1 \le k_2 \le K+1$, $\mathtt{\Phi}^\pi_{\mathrm{cl},k_2,k_1} = \mathtt{A}^\pi_{\mathrm{cl},k_2-1} \cdot \mathtt{A}^\pi_{\mathrm{cl},k_2-2} \cdots \mathtt{A}^\pi_{\mathrm{cl},k_1}$, with the convention $\mathtt{\Phi}^\pi_{\mathrm{cl},k_1,k_1} = \mathbf{I}$. Finally, we define the closed-loop *markov operator* via $\mathtt{\Psi}^\pi_{\mathrm{cl},k_2,k_1} := \mathtt{\Phi}^\pi_{\mathrm{cl},k_2,k_1+1}\mathtt{B}^\pi_{\mathrm{ol},k_1}$ for $1 \le k_1 < k_2 \le K+1$.

Definition C.9 (Closed-Loop Jacobian Linearizations, Continuous-Time)**.** We define

$$\mathbf{\Phi}^\pi_{\mathrm{cl}}(s,t) := \begin{cases} \mathbf{\Phi}^\pi_{\mathrm{ol}}(s,t) & s,t \in \mathcal{I}_k \\ \tilde{\mathbf{\Phi}}^\pi_{\mathrm{cl}}(s, t_{k_2}) \cdot \mathtt{\Phi}^\pi_{\mathrm{cl},k_2,k_1} \cdot \mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k_1+1}, t) & t \in \mathcal{I}_{k_1}, s \in \mathcal{I}_{k_2}, k_2 > k_1, \end{cases}$$

where above, we define

$$\tilde{\mathbf{\Phi}}^\pi_{\mathrm{cl}}(s, t_k) = \mathbf{\Phi}^\pi_{\mathrm{ol}}(s, t_k) + (\int_{s'=t_k}^{s} \mathbf{\Phi}^\pi_{\mathrm{ol}}(s, s')\mathbf{B}^\pi_{\mathrm{ol}}(s')\mathrm{d}s)\mathtt{K}_k.$$

Lastly, we define

$$\mathbf{\Psi}^\pi_{\mathrm{cl}}(s,t) = \mathbf{\Phi}^\pi_{\mathrm{cl}}(s,t)\mathbf{B}(t).$$

C.2. Characterizations of the Jacobian Linearizations

In this section we provide characterizations of the Jacobian Linearizations of the open-loop and closed-loop trajectories.

Lemma C.1 (Implicit Characterization of the linearizations in open-loop)**.** *Given* $\mathbf{u}, \bar{\mathbf{u}} \in \mathcal{U}$*, define* $\delta\bar{\mathbf{u}}(t) = \bar{\mathbf{u}}(t) - \mathbf{u}(t)$*. Then,*

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}; \mathbf{u}) = \mathbf{A}(t \mid \mathbf{u})\delta\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}; \mathbf{u}) + \mathbf{B}(t \mid \mathbf{u})\delta\bar{\mathbf{u}}(t)$$

with initial condition $\delta\mathbf{x}^{\mathrm{jac}}(0 \mid \bar{\mathbf{u}}) = 0$*, where*

$$\mathbf{A}(t \mid \mathbf{u}) = \partial_x f_{\mathrm{dyn}}(x,u)\big|_{x=\mathbf{x}(t|\mathbf{u}),u=\mathbf{u}(t)} \quad \textit{and} \quad \mathbf{B}(t \mid \mathbf{u}) = \partial_u f_{\mathrm{dyn}}(x,u)\big|_{x=\mathbf{x}(t|\mathbf{u}),u=\mathbf{u}(t)}.$$

Proof. The result follows directly from Lemma C.8 and the definition of $\delta\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}; \mathbf{u})$. □

Lemma C.2 (Implicit Characterization of the linearizations in closed-loop)**.** *Given a policy* π *and* $\bar{\mathbf{u}} \in \mathcal{U}$*, set* $\delta\bar{\mathbf{u}}^{\pi}(t) = \mathbf{u}^{\pi}(t) - \mathbf{u}(t)$ *. Then, recalling* $\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) = \tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \mathbf{x}^{\pi}(t)$*,*

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) &= \mathbf{A}^{\pi}_{\mathrm{ol}}(t)\delta\tilde{\mathbf{x}}^{\pi,\mathrm{ct}}(t \mid \bar{\mathbf{u}}) + \mathbf{B}^{\pi}_{\mathrm{ol}}(t)\delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t)\\ \delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}} &:= \delta\bar{\mathbf{u}}^{\pi}(t) + \mathtt{K}^{\pi}_{k(t)}\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t_{k(t)} \mid \bar{\mathbf{u}}),\end{aligned}$$

with initial condition $\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(0 \mid \bar{\mathbf{u}}) = 0$*.*

Proof. The result follows directly from Lemma C.8 and the definitions of $\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})$ and the construction of the perturbed input $\delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}$. □

Lemma C.3 (Explicit Characterizations of Linearizations, Continuous-Time)**.** *For a policy* π*, we have:*

$$\begin{aligned}\delta\mathbf{x}^{\mathrm{jac}}(t \mid \mathbf{u}^{\pi} + \delta\mathbf{u}; \mathbf{u}^{\pi}) &= \int_{s=0}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(t)\delta\mathbf{u}(s)\mathrm{d}s.\\ \delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \mathbf{u}^{\pi} + \delta\mathbf{u}) &= \int_{s=0}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\mathrm{d}s.\end{aligned}$$

Proof. The first condition follows directly from the characterization of the evolution of $\delta\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}; \mathbf{u})$ and Lemma C.8. For the second condition, we will directly argue that the proposed formula satisfied the differential equation in Lemma C.2. By the Leibniz integral rule we have:

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\left(\int_{s=0}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\mathrm{d}s\right) &= \int_{s=0}^{t} \tfrac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\right)\mathrm{d}s + \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t)\delta\mathbf{u}(t)\\ &= \int_{s=0}^{t} \tfrac{\mathrm{d}}{\mathrm{d}t}\left(\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\right)\mathrm{d}s + \mathbf{B}^{\pi}_{\mathrm{ol}}(t)\delta\mathbf{u}(t).\end{aligned}$$

Using the expression for $\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)$ in Definition C.9, we can similarly calculate:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s) = \mathbf{A}^{\pi}_{\mathrm{ol}}(t)\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s) + \mathbf{B}^{\pi}_{\mathrm{ol}}(t)\mathtt{K}_k\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k(t)},s).$$

Together the precding quatinities demonstrate that:

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}\left(\int_{s=0}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\mathrm{d}s\right) &= \mathbf{A}^{\pi}_{\mathrm{ol}}(t)\cdot\left(\int_{s=0}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\mathrm{d}s\right) + \mathbf{B}^{\pi}_{\mathrm{ol}}(t)\delta\mathbf{u}(t)\\ &\quad + \mathtt{K}^{\pi}_{k(t)}\cdot\left(\int_{s=0}^{t_{k(t)}} \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\delta\mathbf{u}(s)\mathrm{d}s\right),\end{aligned}$$

which demonstrates the proposed solutions satisfies the desired differential equation. □

Lemma C.4 (Explicit Characterizations of Linearizations, Discrete-Time)**.** *For a policy π, and perturbation $\delta \mathtt{u}_{1:K} \in \mathbf{U}$,*

$$\delta \tilde{\mathtt{x}}_k^{\pi,\mathrm{jac}}(\mathtt{u}_{1:K}^\pi + \delta \mathtt{u}_{1:K}) = \sum_{j=1}^{k-1} \Psi_{\mathrm{cl},k,j}^\pi \delta \mathtt{u}_j = \sum_{j=1}^{k-1} \Phi_{\mathrm{cl},k,j+1}^\pi \mathtt{B}_{\mathrm{ol},j}^\pi \delta \mathtt{u}_j.$$

Proof. The proof follows directly from Definition C.9, Definition C.8 and Lemma C.3. □

C.3. Gradient Computations

Lemma C.5 (Computation of Continuous-Time Gradient, Open-Loop)**.** *Fix $\xi = \xi^\pi$. Define $Q_{(\cdot)}^\pi(t) := \partial_{(\cdot)} Q(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t), t)$. Then,*

$$\nabla \mathcal{J}_T(\mathbf{u})(t)\big|_{\mathbf{u}=\mathbf{u}^\pi} = Q_u^\pi(t) + \int_{s=t}^T \mathbf{\Phi}_{\mathrm{ol}}^\pi(s,t) \mathbf{B}_{\mathrm{ol}}^\pi(t) \mathrm{d}s.$$

and

$$\langle \nabla \mathcal{J}_T(\mathbf{u})(t)\big|_{\mathbf{u}=\mathbf{u}^\pi}, \delta \mathbf{u} \rangle = \int_0^T (\langle Q_u^\pi(t), \delta \mathbf{u}(t)) + \langle Q_x^\pi(t), \delta \mathbf{x}^{\mathrm{jac}}(t \mid \mathbf{u} + \delta \mathbf{u}) \rangle) \mathrm{d}t.$$

Proof. For a given perturbation $\delta \mathbf{u}$, by the chain rule we have:

$$\mathrm{D}\mathcal{J}_T(\mathbf{u})[\delta \mathbf{u}] = \int_0^T (\langle Q_u^\pi(t), \delta \mathbf{u}(t)) + \langle Q_x^\pi(t), \delta \mathbf{x}^{\mathrm{jac}}(t \mid \mathbf{u} + \delta \mathbf{u}) \rangle) \mathrm{d}t + \langle \partial_x V(\mathbf{x}^\pi(T)), \delta \mathbf{x}^{\mathrm{jac}}(T \mid \mathbf{u} + \delta \mathbf{u}) \rangle$$

Because $\delta \mathbf{u}$ is arbitrary, an application of Lemma C.3 demonstrates the desired results. □

Lemma C.6 (Computation of Continuous-Time Gradient, Closed-Loop)**.** *Fix $\xi = \xi^\pi$. Define $Q_{(\cdot)}^\pi(t) := \partial_{(\cdot)} Q(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t), t)$. Then,*

$$\begin{aligned}\nabla \mathcal{J}_T(\pi)(t) := \nabla_{\bar{\mathbf{u}}} \mathcal{J}_T^\pi(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^\pi}(t) = Q_u^\pi(t) + \mathbf{\Psi}_{\mathrm{cl}}^\pi(T,t)^\top (\partial_x V(\mathbf{x}^\pi(T))) \\ + \int_{s=t}^T \mathbf{\Psi}_{\mathrm{cl}}^\pi(s,t)^\top Q_x^\pi(s) \mathrm{d}s + \int_{s=t_{k(t)+1}}^T \mathbf{\Psi}_{\mathrm{cl}}^\pi(t_{k(s)},t)^\top \mathtt{K}_{k(s)}{}^\top Q_u^\pi(s) \mathrm{d}s.\end{aligned}$$

Proof. The proof follows the steps of Lemma C.5, but replaces the open-loop state and input perturbations with the appropriate closed-loop perturbations, as defined in Lemma C.3 and calculated in Lemma C.2. □

Similarly, we can compute the gradient of the discrete-time objective. Its proof is analogous to the previous two.

Lemma C.7 (Computation of Discrete-Time Gradient)**.**

$$\begin{aligned}(\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi))_k = \tau Q_u(\mathtt{x}_k^\pi, \mathtt{u}_k^\pi, t_k) + (\Psi_{\mathrm{cl},K+1,k}^\pi)^\top V_x(\mathtt{x}_{K+1}^\pi) \\ + \tau \sum_{j=k+1}^K (\Psi_{\mathrm{cl},j,k}^\pi)^\top (Q_x(\mathtt{x}_j^\pi, \mathtt{u}_j^\pi, t_j) + (\mathtt{K}_j^\pi)^\top Q_u(\mathtt{x}_j^\pi, \mathtt{u}_j^\pi, t_j))\end{aligned}$$

Moreorever, defining the shorthand $\delta \tilde{\mathtt{x}}_k^{\mathrm{jac}} = \delta \tilde{\mathtt{x}}_k^{\pi,\mathrm{jac}}(\mathtt{u}_{1:K}^\pi + \delta \mathtt{u}_{1:K})$,

$$\begin{aligned}&\langle \delta \mathtt{u}_{1:K}, \nabla \mathcal{J}_T^{\pi,\mathrm{disc}}(\pi) \rangle \\ &= \langle \partial_x V(\mathtt{x}_{K+1}^\pi), \delta \tilde{\mathtt{x}}_k^{\mathrm{jac}} \rangle + \tau \sum_{k=1}^K \langle \partial_x Q(\mathtt{x}_k^\pi, \mathtt{u}_k^\pi, t_k), \delta \tilde{\mathtt{x}}_k^{\mathrm{jac}} \rangle + \langle \partial_u Q(\mathtt{x}_k^\pi, \mathtt{u}_k^\pi, t_k), \delta \mathtt{u}_k + \mathtt{K}_k \delta \tilde{\mathtt{x}}_k^{\mathrm{jac}} \rangle.\end{aligned}$$

C.4. Technical Tools

The first supportive lemma is a standard result from variational calculus, and characterizes how the solution to the controlled differential equation changes under perturbations to the input. Note that this result does not depend on how the input is generated, namely, whether the perturbation is generated in open-loop or closed-loop. Concretely, the statement of the following result is equivalent to Theorem 5.6.9 from (Polak, 2012).

Lemma C.8 (State Variation of Controlled CT Systems)**.** *For each nominal input $\mathbf{u} \in \mathcal{U}$ and perturbation $\delta\mathbf{u} \in \mathcal{U}$ we have:*

$$\langle \nabla_{\mathbf{u}} \mathbf{x}(t \mid \mathbf{u}), \delta\mathbf{u} \rangle = \delta x(t), \tag{C.6}$$

where the curve $\delta x(\cdot)$ satisfies:

$$\frac{d}{dt}\delta x(t) = \mathbf{A}(t \mid \mathbf{u})\delta x(t) + \mathbf{B}(t \mid \mathbf{u})\delta\mathbf{u}(t), \tag{C.7}$$

where we recall that:

$$\mathbf{A}(t \mid \mathbf{u}) = \partial_x f_{\mathrm{dyn}}(x,u)\big|_{x=\mathbf{x}(t|\mathbf{u}),u=\mathbf{u}(t)} \quad \textit{and} \quad \mathbf{B}(t \mid \mathbf{u}) = \partial_u f_{\mathrm{dyn}}(x,u)\big|_{x=\mathbf{x}(t|\mathbf{u}),u=\mathbf{u}(t)}.$$

Moreover, we have

$$\delta x(t) = \int_{s=0}^{t} \mathbf{\Phi}(t,s)\mathbf{B}(s \mid \mathbf{u})\delta\mathbf{u}(s),$$

where the transition operator satisfies $\frac{d}{dt}\mathbf{\Phi}(t,s) = \mathbf{A}(t \mid \mathbf{u})\mathbf{\Phi}$ and $\mathbf{\Phi}(s,s) = \mathbf{I}$.

The following result is equivalent to Lemma 5.6.2 from (Polak, 2012).

Lemma C.9 (Picard Lemma)**.** *Consider two dynamical $\frac{\mathrm{d}}{\mathrm{d}t}y(t) = \phi(y(t),t)$, $i \in \{1,2\}$, and suppose that $y \mapsto \phi(y,s)$ is L-Lipschitz fo each s fixed. Let $z(t)$ be any other absolutely continuous curve. Then,*

$$\|y(t) - z(t)\| \leq \exp(tL) \cdot \left(\|y(0) - z(0)\| + \int_{s=0}^{t} \|\frac{\mathrm{d}}{\mathrm{d}s}z(s) - f(z(s),s)\| \right).$$

Lemma C.10 (Solution to Afine ODEs)**.** *Consider an affine ODE given by $\mathbf{y}(0) = y_0$, $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}(t) = \mathbf{A}(t)\mathbf{y}(t) + \mathbf{B}(t)u$. Then,*

$$y(t) = \Phi(t,0)y_0 + \left(\int \Phi(t,s)\mathbf{B}(s)\mathrm{d}s \right) u,$$

where $\Phi(t,s)$ solves the ODE $\Phi(s,s) = \mathbf{I}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\Phi(t,s) = \mathbf{A}(t)\Phi(t,s)$.

D. Taylor Expansions of the Dynamics

D.1. Proof of Proposition A.5

Recall $\delta \mathtt{u}_k = \mathtt{u}_k - \mathtt{u}_k^\pi$, and define $\mathbf{u} = \mathtt{ct}(\mathtt{u}_{1:K})$ and $\delta\mathbf{u} := \mathtt{ct}(\mathtt{u}^\pi_{1:K})$. We define shorthand for relevant continuous curves and their discretizations:

$$\begin{aligned} \mathbf{y}(t) &= \tilde{\mathbf{x}}^\pi(t \mid \mathbf{u}^\pi + \delta\mathbf{u}) = \tilde{\mathbf{x}}^\pi(t \mid \mathbf{u}^\pi) \\ \mathtt{y}_k &= \mathbf{y}(t_k) = \tilde{\mathtt{x}}^\pi_k\left(\mathtt{u}^\pi_{1:K} + \delta\mathtt{u}_{1:K}\right) = \tilde{\mathtt{x}}^\pi_k\left(\mathtt{u}_{1:K}\right) \\ \mathbf{y}^{\mathtt{jac}}(t) &= \tilde{\mathbf{x}}^{\pi,\mathtt{jac}}(t \mid \mathbf{u}^\pi + \delta\mathbf{u}), \quad \mathtt{y}^{\mathtt{jac}}_k = \mathbf{y}^{\mathtt{jac}}(t_k) \end{aligned} \tag{D.1}$$

We also define their differences from the nominal as

$$\delta\mathbf{y}(t) = \mathbf{y}(t) - \mathbf{x}^\pi(t), \quad \delta\mathbf{y}^{\mathtt{jac}}(t) = \mathbf{y}^{\mathtt{jac}}(t) - \mathbf{x}^\pi(t), \quad \delta\mathtt{y}_k = \mathtt{y}_k - \mathtt{x}^\pi_k, \quad \delta\mathtt{y}^{\mathtt{jac}}_k = \mathtt{y}^{\mathtt{jac}}_k - \mathtt{x}^\pi_k.$$

And the Jacobian error

$$\mathbf{e}^{\mathtt{jac}}(t) := \mathbf{y}(t) - \mathbf{y}^{\mathtt{jac}}(t), \quad \mathtt{e}^{\mathtt{jac}}_k := \mathbf{e}^{\mathtt{jac}}(t_k) = \mathtt{y}_k - \mathtt{y}^{\mathtt{jac}}_k.$$

The main challenge is recursively controlling $\|\mathtt{e}^{\mathtt{jac}}_k\|$. We begin with a computation which is immediate from Definition C.1 (the first equality) and (the second equality):

Lemma D.1 (Curve Computations)**.** *For $t \in \mathcal{I}_k$,*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}(t) &= f_{\mathrm{dyn}}(\mathbf{y}(t), \mathtt{u}^\pi_k + \delta\mathtt{u}_k + \mathtt{K}^\pi_k(\delta\mathtt{y}_k)) \\ \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}^{\mathtt{jac}}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}^\pi(t) + \mathbf{A}^\pi_{\mathrm{ol}}(t)\delta\mathbf{y}^{\mathtt{jac}}(t) + \mathbf{B}^\pi_{\mathrm{ol}}(t)\left(\delta\mathtt{u}_k + \mathtt{K}_k(\delta\mathtt{y}^{\mathtt{jac}}_k)\right) \end{aligned}$$

Computing the Jacobian Linearization. The first step of the proof is a computation of the Jacobian linearization and a bound on its magnitude.

Lemma D.2 (Computation of Jacobian Linearization)**.**

$$\mathtt{y}^{\mathtt{jac}}_k = \sum_{j=1}^{k-1} \mathtt{\Phi}^\pi_{\mathrm{cl},k,j+1}\mathtt{B}^\pi_{\mathrm{ol},j}\delta\mathtt{u}_j = \sum_{j=1}^{k-1} \mathtt{\Psi}^\pi_{\mathrm{cl},k,j}\delta\mathtt{u}_j \tag{D.2}$$

Therefore,

$$\max_{k\in[K+1]} \|\mathtt{y}^{\mathtt{jac}}_k\| \le L_{\mathrm{ol}} L_f \min\{B_2\kappa_{\pi,2}, B_\infty\kappa_{\pi,1}\} \tag{D.3}$$

Proof. Eq. (D.2) follows from Lemma C.6. Eq. (D.3) follows from Cauchy Schwartz/Holder's inequality, and the bound $\|\mathtt{B}^\pi_{\mathrm{ol},j}\| \le \tau L_{\mathrm{ol}} L_f$ due to Lemma I.3. □

Recursion on proximity to Jacobian linearization. Next, we argue that the true dynamics $\mathbf{y}(t)$ remain close to $\mathbf{y}^{\mathtt{jac}}(t)$. We establish a recursion under the following invariant:

$$\begin{aligned} &\|\mathtt{y}^{\mathtt{jac}}_k\| \vee \|\mathtt{y}_k\| \vee \|\mathtt{u}^\pi_k + \delta\mathtt{u}_k + \mathtt{K}^\pi_k(\delta\mathtt{y}_k)\| \vee \|\mathtt{u}^\pi_k + \delta\mathtt{u}_k + \mathtt{K}^\pi_k(\delta\mathtt{y}^{\mathtt{jac}}_k)\| \le \frac{3}{4}R_{\mathrm{feas}} \\ &\tau \le \min\left\{\frac{1}{16L_fL_\pi}, \frac{1}{8\kappa_f}\right\} \end{aligned} \tag{D.4}$$

We prove the following recursion:

Lemma D.3 (Recursion on Error of Linearization)**.** *Suppose Eq. (D.4) holds. Let $\tilde{\mathbf{\Phi}}_{\mathrm{cl},\pi}(t,t_k) = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t,t_k) + (\int_{s=t_k}^t \mathbf{\Phi}^\pi_{\mathrm{ol}}(t,s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathrm{d}s)\mathtt{K}^\pi_k$. Then, the following bound holds:*

$$\sup_{t\in\mathcal{K}} \|\mathbf{e}^{\mathtt{jac}}(t) - \tilde{\mathbf{\Phi}}_{\mathrm{cl},\pi}(t,t_k)\mathtt{e}^{\mathtt{jac}}_k\| \le M_f\tau\left(4\|\delta\mathtt{u}_k\|^2 + 20L_\pi^2\|\mathtt{e}^{\mathtt{jac}}_k\|^2 + 20L_\pi^2\|\mathtt{y}^{\mathtt{jac}}_k - \mathtt{y}_k\|^2\right).$$

In particular, we have

$$\|\mathtt{e}^{\mathtt{jac}}_{k+1} - \mathtt{A}^\pi_{\mathrm{cl},k}\mathtt{e}^{\mathtt{jac}}_k\| \le M_f\tau\left(4\|\delta\mathtt{u}_k\|^2 + 20L_\pi^2\|\mathtt{e}^{\mathtt{jac}}_k\|^2 + 20L_\pi^2\|\mathtt{y}^{\mathtt{jac}}_k - \mathtt{y}_k\|^2\right).$$

To do prove Lemma D.3, we introduce another family of curves $\check{\mathbf{y}}_k^{\mathtt{jac}}(t)$, defined for $t \geq t_k$, which begin at $\mathbf{y}(t_k)$ but evolve according to the Jacobian linearization:

$$\frac{\mathrm{d}}{\mathrm{d}t}\check{\mathbf{y}}_k^{\mathtt{jac}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}^\pi(t) + \mathbf{A}_{\mathrm{ol}}^\pi(t)\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t) + \mathbf{B}_{\mathrm{ol}}^\pi(t)\left(\delta\mathtt{u}_k + \mathtt{K}_k(\delta\mathtt{y}_k)\right)$$
$$\check{\mathbf{y}}_k^{\mathtt{jac}}(t_k) = \mathbf{y}(t_k) = \mathtt{y}_k, \quad \delta\mathtt{y}_k^{\mathtt{jac}}(t) = \mathtt{y}_k^{\mathtt{jac}}(t) - \mathbf{x}^\pi(t).$$

We begin by establishing feasibility of all relevant continuous-time curves on the interval $\mathcal{I}_k$.

Lemma D.4. *Suppose that Eq. (D.4) holds. Then, for all $t \in \mathcal{I}_k$,*

$$\|\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \vee \|\mathbf{y}(t)\| \vee \|\mathbf{y}^{\mathtt{jac}}(t)\| \leq R_{\mathrm{feas}}.$$

Proof. Let us start with $\mathbf{y}(t)$. Define the shorthand $\bar{\mathtt{u}}_k := \mathtt{u}_k^\pi + \delta\mathtt{u}_k + \mathtt{K}_k^\pi(\delta\mathtt{y}_k)$, so that $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}(t) = f_{\mathrm{dyn}}(\mathbf{y}(t), \mathtt{u}_k)$. Under Eq. (D.4), $\|\bar{\mathtt{u}}_k\| \leq \frac{3}{4}R_{\mathrm{feas}}$. At $t = t_k$, $\|\mathbf{y}(t)\| \leq \frac{3}{4}R_{\mathrm{feas}}$. Moreover, if at a given t, $\|\mathbf{y}(t)\| \leq R_{\mathrm{feas}}$, then $(\mathbf{y}(t), \mathtt{u}_k)$ is feasible, so $\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{y}(t)\| = \|f_{\mathrm{dyn}}(\mathbf{y}(t), \bar{\mathtt{u}}_k)\| \leq \kappa_f$. Thus, letting $t_\star := \sup\{t \in \mathcal{I}_K : \|\mathbf{y}(t)\| \leq R_{\mathrm{feas}}\}$, we see that if $\tau \leq \frac{R_{\mathrm{feas}}}{4\kappa_f}$, $t_\star = t_{k+1}$.

The arguments for $\check{\mathbf{y}}_k^{\mathtt{jac}}(t)$ and $\mathbf{y}^{\mathtt{jac}}(t)$ are similar: if, say, $\|\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \leq R_{\mathrm{feas}}$ for a given $t \in \mathcal{I}_k$, then by

$$\begin{aligned}\|\frac{\mathrm{d}}{\mathrm{d}t}\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| &= \|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}^\pi(t) + \mathbf{A}_{\mathrm{ol}}^\pi(t)\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t) + \mathbf{B}_{\mathrm{ol}}^\pi(t)\mathtt{u}_k\| \\ &\leq \kappa_f + L_f(\|\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| + \|\mathtt{u}_k\|) \\ &\leq \kappa_f + 2R_{\mathrm{feas}}L_f.\end{aligned}$$

where above we use Eq. (D.4) to bound $\|\mathtt{u}_k\| \leq R_{\mathrm{feas}}$, feasibility of π. As $\|\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \leq \frac{3}{4}R_{\mathrm{feas}}$, integrating (specifically, again considering $t_\star := \sup\{t \in \mathcal{I}_K : \|\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \leq R_{\mathrm{feas}}\}$) shows that as long as $\tau(\kappa_f + 2L_fR_{\mathrm{feas}}) \leq R_{\mathrm{feas}}/4$, $\|\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \leq R_{\mathrm{feas}}$ for all $t \in \mathcal{I}_k$. For this, it suffices that $\tau \leq \min\{\frac{1}{16L_f}, 1/8\kappa_f\}$. □

We continue with a crude bound on the difference $\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t) = \check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{x}^\pi(t)$.

Lemma D.5. *Suppose Eq. (D.4) holds all $t \in \mathcal{I}_k$*

$$\|\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \leq (\tau L_{\mathrm{ol}}L_f\|\delta\mathtt{u}_k\| + L_{\mathrm{ol}}(1 + \tau L_\pi L_f)\|\delta\mathtt{y}_k\|)$$

Similarly,

$$\|\delta\mathbf{y}(t)\| \leq (\tau L_{\mathrm{ol}}L_f\|\delta\mathtt{u}_k\| + L_{\mathrm{ol}}(1 + \tau L_\pi L_f)\|\delta\mathtt{y}_k\|).$$

Proof. Then, Picard's Lemma (Lemma C.9), feasibiliy of π and Assumption 4.1 implies that, for any $t \in \mathcal{I}_k$

$$\|\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(t)\| \leq \exp((t - t_k)L_f)\epsilon_1$$

where

$$\begin{aligned}\epsilon_1 &:= \|\delta\mathtt{y}_k\| + \int_{s=t_k}^{t} \|\mathbf{B}_{\mathrm{ol}}^\pi(t)\left(\delta\mathtt{u}_k + \mathtt{K}_k(\delta\mathtt{y}_k)\right)\|\mathrm{d}s \\ &\leq \|\delta\mathtt{y}_k\| + \int_{s=t_k}^{t_{k+1}} L_f(\|\delta\mathtt{u}_k\| + \|\mathtt{K}_k^\pi\delta\mathtt{y}_k\|) && \text{(Assumption 4.1)} \\ &\leq \tau(L_f\|\delta\mathtt{u}_k\|\|\delta\mathtt{u}_k\| + L_\pi L_f\|\delta\mathtt{y}_k\|), && \text{(Definition 4.7)}\end{aligned}$$

Bounding $\exp((t - t_k)L_f) \leq \exp(\tau L_f) = L_{\mathrm{ol}}$ concludes the first part. The second part follows from a similar argument, using Lipschitzness of f_{dyn} in accordance with Assumption 4.1, and the feasibility of $(\mathbf{y}(t), \mathtt{u}_k^\pi + \delta\mathtt{u}_k + \mathtt{K}_k^\pi(\delta\mathtt{y}_k^{\mathtt{jac}}))$ for $t \in \mathcal{I}_k$, as ensured by Eq. (D.4) and Lemma D.4. □

We are now ready to prove Lemma D.3.

Proof of Lemma D.3. Observe that, for $t \in \mathcal{I}_k$

$$\frac{\mathrm{d}}{\mathrm{d}t}(\check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}^{\mathtt{jac}}(t)) = \mathbf{A}_{\mathrm{ol}}^{\pi}(t)(\check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}^{\mathtt{jac}}(t)) + \mathbf{B}_{\mathrm{ol}}^{\pi}(t)\mathtt{K}_k^{\pi}(\check{\mathbf{y}}_k^{\mathtt{jac}}(t_k) - \mathbf{y}^{\mathtt{jac}}(t_k))$$

By solving the affine ODE define $\check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}^{\mathtt{jac}}(t)$ (applying Lemma C.10), and recalling all various defintions,

$$\check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}^{\mathtt{jac}}(t) = \tilde{\mathbf{\Phi}}_{\mathrm{cl},\pi}(t, t_k)(\mathbf{y}(t_k) - \mathbf{y}^{\mathtt{jac}}(t_k)). \tag{D.5}$$

We now bound $\check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}(t)$. By applying Picard's Lemma (Lemma C.9) and Assumption 4.1 with $L_{\mathrm{ol}} = \exp(\tau L_f)$ to control the Lipschitz constant contribution, and using the agreement of initial conditions $\check{\mathbf{y}}_k^{\mathtt{jac}}(t_k) = \mathbf{y}(t_k)$,

$$\|\check{\mathbf{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}(t)\| \le L_{\mathrm{ol}} \int_{s=t_k}^{t} \|\Delta(s)\| \mathrm{d}s. \tag{D.6}$$

where

$$\Delta(s) = f_{\mathrm{dyn}}(\check{\mathbf{y}}_k^{\mathtt{jac}}(s), \mathtt{u}_k^{\pi} + \delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k) - \frac{\mathrm{d}}{\mathrm{d}s}\check{\mathbf{y}}_k^{\mathtt{jac}}(s).$$

By a Taylor expansion, we have

$$\begin{aligned}\Delta(s) = f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi}) - \frac{\mathrm{d}}{\mathrm{d}s}\check{\mathbf{y}}_k^{\mathtt{jac}}(s) + \partial_x f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi})\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(s) + \partial_u f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi})(\delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k) \\ + \frac{1}{2}\text{remainder},\end{aligned}$$

where we bound

$$\begin{aligned}\|\text{remainder}(s)\| \le \sup_{\alpha\in[0,1]} \|\nabla^2 f_{\mathrm{dyn}}(\alpha\mathbf{x}^{\pi}(s) + (1-\alpha)\mathtt{y}_k^{\mathtt{jac}}(s), \alpha\mathtt{u}_k^{\pi} + (1-\alpha)(\mathtt{u}_k^{\pi} + \delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k))\| \\ \cdot \left(\|\delta\mathtt{y}_k^{\mathtt{jac}}(s)\|^2 + \|\delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k\|^2\right).\end{aligned}$$

From by feasibility of π, $\|\mathtt{u}_k^{\pi}\| \vee \|\mathbf{x}^{\pi}(s)\| \le R_{\mathrm{feas}}$. Moreover, $\|\mathtt{u}_k^{\pi} + \delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k\| \le R_{\mathrm{feas}}$ by Eq. (D.4) and $\|\mathtt{y}_k^{\mathtt{jac}}(s)\| \le R_{\mathrm{feas}}$ by Lemma D.4. Thus, for $\alpha \in [0,1]$,

$$\|\alpha\mathbf{x}^{\pi}(s) + (1-\alpha)\mathtt{y}_k^{\mathtt{jac}}(s)\| \vee \|\alpha\mathtt{u}_k^{\pi} + (1-\alpha)(\mathtt{u}_k^{\pi} + \delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k)\| \le R_{\mathrm{feas}}.$$

Hence, as $\|\nabla^2 f_{\mathrm{dyn}}(x,u)\| \le M_f$ for feasible (x,u), Assumption 4.1 implies

$$\begin{aligned}\|\text{remainder}(s)\| &\le M_f\left(\|\delta\mathbf{y}^{\mathtt{jac}}(s)\|^2 + 2\|\delta\mathtt{u}_k\|^2 + 2L_{\pi}^2\|\delta\mathtt{y}_k\|^2\right) && \text{(AM-GM and Definition 4.7)} \\ &\le M_f\left(\|(L_{\mathrm{ol}}L_f\tau\|\delta\mathtt{u}_k\| + L_{\mathrm{ol}}(1+\tau L_{\pi}L_f)\|\delta\mathtt{y}_k\|)\|^2 + 2\|\delta\mathtt{u}_k\|^2 + 2L_{\pi}^2\|\delta\mathtt{y}_k\|^2\right) && \text{(ALemma D.5)} \\ &\le 2M_f\left((\tau^2 L_{\mathrm{ol}}^2 L_f^2\|\delta\mathtt{u}_k\|^2) + L_{\mathrm{ol}}^2(1+\tau L_f L_{\pi})^2\|\delta\mathtt{y}_k\|^2 + \|\delta\mathtt{u}_k\|^2 + L_{\pi}^2\|\delta\mathtt{y}_k\|^2\right) && \text{(AM-GM)} \\ &= 2M_f\left((1+\tau^2 L_{\mathrm{ol}}^2 L_f^2)\|\delta\mathtt{u}_k\|^2) + (L_{\pi}^2 + L_{\mathrm{ol}}^2(1+\tau L_f L_{\pi})^2)\|\delta\mathtt{y}_k\|^2\right).\end{aligned}$$

Finally, we conclude by noting that

$$\begin{aligned}&f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi}) \\ &\quad + \partial_x f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi})\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(s) + \partial_u f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi})(\delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}^{\pi}(s) + \mathbf{A}_{\mathrm{ol}}^{\pi}(s)\delta\check{\mathbf{y}}_k^{\mathtt{jac}}(s) + \mathbf{B}_{\mathrm{ol}}^{\pi}(s)f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(s), \mathtt{u}_k^{\pi})(\delta\mathtt{u}_k + \mathtt{K}_k^{\pi}\delta\mathtt{y}_k) = \frac{\mathrm{d}}{\mathrm{d}s}\check{\mathbf{y}}_k^{\mathtt{jac}}(s),\end{aligned}$$

so that

$$\begin{aligned}\|\Delta(s)\| = \frac{1}{2}\|\text{remainder}(s)\| &\le M_f\left((1+\tau^2 L_{\mathrm{ol}}^2 L_f^2)\|\delta\mathtt{u}_k\|^2) + (L_{\pi}^2 + L_{\mathrm{ol}}^2(1+\tau L_f L_{\pi})^2)\|\delta\mathtt{y}_k\|^2\right) \\ &\le M_f\left((1+2\tau^2 L_f^2)\|\delta\mathtt{u}_k\|^2) + (L_{\pi}^2 + 2(1+\tau L_f L_{\pi})^2)\|\delta\mathtt{y}_k\|^2\right) \\ &\qquad (\tau \le 1/4L_f, \text{ so } L_{\mathrm{ol}}^2 \le 2) \\ &\le M_f\left(2\|\delta\mathtt{u}_k\|^2) + (L_{\pi}^2 + 4)\|\delta\mathtt{y}_k\|^2\right), \quad \text{(again } \tau \le 1/4L_f, \text{ and when as } \tau \le 4/L_f L_{\pi}) \\ &\le M_f\left(2\|\delta\mathtt{u}_k\|^2) + 5L_{\pi}^2\|\delta\mathtt{y}_k\|^2\right) && (L_{\pi} \ge 1)\end{aligned}$$

where in the last line, we use $\tau \le 1/4L_f$, so $L_{\mathrm{ol}}^2 \le 2$. Hence, from Eq. (D.6), for all $t \in \mathcal{I}_k$,

$$\begin{aligned}\|\check{\mathtt{y}}_k^{\mathtt{jac}}(t) - \mathbf{y}(t)\| &\le \tau L_{\mathrm{ol}} M_f M_f \left(2\|\delta \mathtt{u}_k\|^2\right) + (L_\pi^2 + 4)\|\delta \mathtt{y}_k\|^2\right). \\ &\le 2M_f\tau\left(2\|\delta \mathtt{u}_k\|^2\right) + (L_\pi^2+4)\|\delta \mathtt{y}_k\|^2\right) \\ &= M_f\tau\left(4\|\delta \mathtt{u}_k\|^2\right) + 10L_\pi^2\|\delta \mathtt{y}_k\|^2\right).\end{aligned}$$

where above we bound $L_{\mathrm{ol}} \le 2$ again. And thus, from Eq. (D.5),

$$\begin{aligned}\|\mathbf{y}(t) - \mathbf{y}^{\mathtt{jac}}(t) - \tilde{\mathbf{\Phi}}_{\mathrm{cl},\pi}(t,t_k)(\mathbf{y}(t_k) - \mathbf{y}^{\mathtt{jac}}(t_k))\| &\le M_f\tau\left(4\|\delta \mathtt{u}_k\|^2\right) + 10L_\pi^2\|\delta \mathtt{y}_k\|^2\right)., \\ &\le M_f\tau\left(4\|\delta \mathtt{u}_k\|^2\right) + 10L_\pi^2\|\|\mathtt{y}_k^{\mathtt{jac}} - \mathtt{y}_k\|^2\|\right) \\ &\le 2M_f\tau\left(4\|\delta \mathtt{u}_k\|^2 + 20L_\pi^2\|\delta \mathtt{y}_k^{\mathtt{jac}}\|^2 + 20L_\pi^2\|\mathtt{y}_k^{\mathtt{jac}} - \mathtt{y}_k\|^2\right).\end{aligned}$$

Substituing in $\mathtt{e}_k^{\mathtt{jac}} := \mathtt{y}_k - \mathtt{y}_k^{\mathtt{jac}}$ concludes. □

Solving the recursion. To upper bound the recursion, assume an inductive hypothesis that, for some R to be chosen

$$\max_{j \le k} \|\mathtt{e}_k^{\mathtt{jac}}\| \le R, \quad \text{and} \quad \forall j \le k, \ \textit{Eq.}\ (\text{D.4}) \text{ holds.} \tag{D.7}$$

Note this hypothesis is true for $k = 1$, where $\mathtt{e}_1^{\mathtt{jac}} = 0$, and all terms in Eq. (D.4) coincide with $(\mathtt{x}_1^\pi, \mathtt{u}_1^\pi)$, which is feasible. Now assume Eq. (D.7) holds. Define

$$\mathtt{v}_k := \mathtt{e}_{k+1}^{\mathtt{jac}} - \mathtt{A}_{\mathrm{cl},k}^\pi \mathtt{e}_k^{\mathtt{jac}},$$

and note that Lemma D.3, followed by our induction hypothesis, implies that for $c_1 = 4$ and $c_2 = 20L_\pi^2$,

$$\begin{aligned}\|\mathtt{v}_k\| &\le \tau M_f\left(4\|\delta \mathtt{u}_k\|^2 + 20L_\pi^2\|\mathtt{e}_k^{\mathtt{jac}}\|^2 + c_2\|\mathtt{y}_k^{\mathtt{jac}}\|^2\right) \\ &\le \tau M_f\left(4\|\delta \mathtt{u}_k\|^2 + 20L_\pi^2 R^2 + 20L_\pi^2\|\mathtt{y}_k^{\mathtt{jac}}\|^2\right).\end{aligned}$$

By unfolding the recursion for $\mathtt{v}_k := \mathtt{e}_{k+1}^{\mathtt{jac}} - \mathtt{A}_{\mathrm{cl},k}^\pi \mathtt{e}_k^{\mathtt{jac}}$, we have

$$\begin{aligned}\mathtt{e}_{k+1}^{\mathtt{jac}} &= \mathtt{v}_k + \mathtt{A}_{\mathrm{cl},k}^\pi \mathtt{e}_k^{\mathtt{jac}} \\ &= \underbrace{\mathbf{I}}_{=\mathtt{\Phi}_{\mathrm{cl},k+1,k+1}^\pi} \mathtt{v}_k + \underbrace{\mathtt{A}_{\mathrm{cl},k}^\pi}_{=\mathtt{\Phi}_{\mathrm{cl},k+1,k}^\pi} \mathtt{v}_{k-1} + \underbrace{\mathtt{A}_{\mathrm{cl},k}^\pi \mathtt{A}_{\mathrm{cl},k-1}^\pi}_{=\mathtt{\Phi}_{\mathrm{cl},k+1,k-1}^\pi} \mathtt{e}_{k-1}^{\mathtt{jac}} \\ &= \sum_{j=1}^{k+1} \mathtt{\Phi}_{\mathrm{cl},k+1,j}^\pi \mathtt{v}_k + \underbrace{\mathtt{\Phi}_{\mathrm{cl},k+1,1}^\pi \mathtt{e}_1^{\mathtt{jac}}}_{=0}.\end{aligned}$$

Thus, under our inductive hypothesis, recalling $B_2 := \tau\|\delta \mathtt{u}_{1:K}\|_{\ell_2}^2$, and $B_\infty := \tau \max_k \|\delta \mathtt{u}_k\|^2$,

$$\begin{aligned}\|\mathtt{e}_{k+1}^{\mathtt{jac}}\| &\le M_f \sum_{j=1}^{k+1} \|\mathtt{\Phi}_{\mathrm{cl},k+1,j}^\pi\|\tau(c_1\|\delta \mathtt{u}_k\|^2 + c_2 R^2 + c_2\|\mathtt{y}_k^{\mathtt{jac}}\|^2) \\ &\le 4M_f \min\{\kappa_{\pi,\infty}B_2^2, \kappa_{\pi,1}B_\infty^2\} + 20M_f L_\pi^2 \kappa_{\pi,1}(R^2 + \max_k \|\mathtt{y}_k^{\mathtt{jac}}\|^2) \\ &\le 4M_f \min\{\kappa_{\pi,\infty}B_2^2, \kappa_{\pi,1}B_\infty^2\} + 20M_f L_\pi^2 \kappa_{\pi,1}(R^2 + L_{\mathrm{ol}}^2 L_f^2 \min\{B_2\kappa_{\pi,2}, B_\infty \kappa_{\pi,1}\}^2) \qquad \text{(Lemma D.2)} \\ &\le 4M_f \min\{B_2^2(\kappa_{\pi,\infty} + 5L_\pi^2 L_{\mathrm{ol}}^2 L_f^2 \kappa_{\pi,2}^2 \kappa_{\pi,1}), B_\infty^2(\kappa_{\pi,1} + 5L_\pi^2 L_{\mathrm{ol}}^2 L_f^2 \kappa_{\pi,1}^3)\} + 20M_f L_\pi^2 \kappa_{\pi,1} R^2. \\ &\le 4M_f \min\{B_2^2(\kappa_{\pi,\infty} + 10L_\pi^2 L_f^2 \kappa_{\pi,2}^2 \kappa_{\pi,1}), B_\infty^2(\kappa_{\pi,1} + 10L_\pi^2 L_f^2 \kappa_{\pi,1}^3)\} + 20M_f L_\pi^2 \kappa_{\pi,1} R^2,\end{aligned}$$

where in the last step we use $\tau \le 1/4L_f$ to bound $L_{\mathrm{ol}}^2 \le 2$. Hence, if we select

$$\begin{aligned}R &= 8M_f \min\{B_2^2(\kappa_{\pi,\infty} + 10L_\pi^2 L_f^2 \kappa_{\pi,2}^2 \kappa_{\pi,1}), B_\infty^2(\kappa_{\pi,1} + 10L_\pi^2 L_f^2 \kappa_{\pi,1}^3)\} \\ &:= \min\{B_2^2 M_{\mathrm{tay},2,\pi}, B_\infty^2 M_{\mathrm{tay},\mathrm{inf},\pi}\},\end{aligned}$$

where we recall

$$M_{\mathrm{tay},2,\pi} := 8M_f(\kappa_{\pi,\infty} + 10L_\pi^2 L_f^2 \kappa_{\pi,2}^2 \kappa_{\pi,1})$$
$$M_{\mathrm{tay},\mathrm{inf},\pi} := 8M_f(\kappa_{\pi,1} + 10L_\pi^2 L_f^2 \kappa_{\pi,1}^3,$$

we get

$$\|\mathtt{e}_{k+1}^{\mathtt{jac}}\| \le \frac{R}{2} + 20M_f L_\pi^2 \kappa_{\pi,1} R^2.$$

Thus, if $R \le \frac{1}{40M_f L_\pi^2 \kappa_{\pi,1}}$, it holds that

$$\|\mathtt{e}_{k+1}^{\mathtt{jac}}\| \le \min\{B_2^2 M_{\mathrm{tay},2,\pi}, B_\infty^2 M_{\mathrm{tay},\mathrm{inf},\pi}\}.$$

Lastly, for the condition $R \le \frac{1}{40M_f L_\pi^2 \kappa_{\pi,1}}$ to hold, it suffices

$$B_2^2 \le \frac{1}{40M_f L_\pi^2 \kappa_{\pi,1} M_{\mathrm{tay},2,\pi}}, \quad \text{or } B_\infty^2 \le \frac{1}{40L_\pi^2 \kappa_{\pi,1} M_{\mathrm{tay},\mathrm{inf},\pi}} \tag{D.8}$$

Notince that these conditions are met for $B_q^2 \le B_{\mathrm{tay},q,\pi}^2$. Moreover,

$$\begin{aligned}
\|\mathtt{y}_{k+1} - \mathtt{x}_{k+1}^\pi\| \vee \|\mathtt{y}_{k+1}^{\mathtt{jac}} - \mathtt{x}_{k+1}^\pi\| &\le \|\mathtt{y}_{k+1} - \mathtt{y}_{k+1}^{\mathtt{jac}}\| + \|\mathtt{y}_{k+1}^{\mathtt{jac}} - \mathtt{x}_{k+1}^\pi \\
&= \|\mathtt{e}_{k+1}^{\mathtt{jac}}\| + \|\sum_{j=1}^{k} \Phi_{\mathrm{cl},k+1,j+1}^\pi \mathtt{B}_{\mathrm{ol},j}^\pi \delta\mathtt{u}_j\| \\
&\le \min_{q\in\{2,\infty\}} M_{\mathrm{tay},q,\pi} B_q^2 + \|\sum_{j=1}^{k-1} \Phi_{\mathrm{cl},k+1,j+1}^\pi \mathtt{B}_{\mathrm{ol},j}^\pi \delta\mathtt{u}_j\| \\
&\le \min_{q\in\{2,\infty\}} M_{\mathrm{tay},q,\pi} B_q^2 + L_{\mathrm{ol}} L_f \min\{B_2\kappa_{\pi,2}, B_\infty \kappa_{\pi,1}\} \qquad \text{(Lemma D.3)} \\
&\le \min\{B_2(L_{\mathrm{ol}} L_f \kappa_{\pi,2} + M_{\mathrm{tay},2,\pi} B_2), B_\infty(L_{\mathrm{ol}} L_f \kappa_{\pi,1} + M_{\mathrm{tay},\mathrm{inf},\pi} B_\infty)\} \\
&\le \min\{B_2(1.5L_f \kappa_{\pi,2} + M_{\mathrm{tay},2,\pi} B_2), B_\infty(1.5L_f \kappa_{\pi,1} + M_{\mathrm{tay},\mathrm{inf},\pi} B_\infty)\} \\
&\qquad (L_{\mathrm{ol}} \le \exp(1/4) \le 1.5)
\end{aligned}$$

Hence, $B_2 \le \frac{1}{2} L_f \kappa_{\pi,2} / M_{\mathrm{tay},2,\pi}$ implies $\|\mathtt{y}_{k+1} - \mathtt{x}_{k+1}^\pi\| \vee \|\mathtt{y}_{k+1}^{\mathtt{jac}} - \mathtt{x}_{k+1}^\pi\| \le 2L_f \kappa_{\pi,2} B_2 = L_{\mathrm{tay},2,\pi} B_2$, and $B_\infty \le \frac{1}{2} L_f \kappa_{\pi,1} / M_{\mathrm{tay},\mathrm{inf},\pi}$ implies $\|\mathtt{y}_{k+1} - \mathtt{x}_{k+1}^\pi\| \vee \|\mathtt{y}_{k+1}^{\mathtt{jac}} - \mathtt{x}_{k+1}^\pi\| \le 2L_f \kappa_{\pi,1} B_\infty = L_{\mathrm{tay},\infty,\pi} B_2$. Combinining these conditions with Eq. (D.8) implies we require $B_q \le B_{\mathrm{tay},q,\pi}$, where

$$B_{\mathrm{tay},2,\pi} = \min\left\{\frac{1}{\sqrt{40M_f L_\pi^2 \kappa_{\pi,1} M_{\mathrm{tay},2,\pi}}}, \frac{L_f \kappa_{\pi,2}}{2M_{\mathrm{tay},2,\pi}}\right\}$$
$$B_{\mathrm{tay},\mathrm{inf},\pi} \le \min\left\{\frac{1}{40L_\pi^2 \kappa_{\pi,1} M_{\mathrm{tay},\mathrm{inf},\pi}}, \frac{L_f \kappa_{\pi,1}}{2M_{\mathrm{tay},\mathrm{inf},\pi}}\right\}$$

Lastly, we need to check that the feasibility invariant Eq. (D.4) for $k = k+1$ is maintained. Under the above conditions, it was shown that

$$\|\delta\mathtt{y}_{k+1}\| \vee \|\delta\mathtt{y}_{k+1}^{\mathtt{jac}}\| = \|\mathtt{y}_{k+1} - \mathtt{x}_{k+1}^\pi\| \vee \|\mathtt{y}_{k+1}^{\mathtt{jac}} - \mathtt{x}_{k+1}^\pi\| \le L_{\mathrm{tay},q,\pi} B_q.$$

Hence, if $B_q \le \frac{R_{\mathrm{feas}}}{8L_\pi L_{\mathrm{tay},q,\pi}} \le \frac{R_{\mathrm{feas}}}{4L_{\mathrm{tay},q,\pi}}$ for either $q \in \{2,\infty\}$,

$$\|\mathtt{y}_{k+1}^{\mathtt{jac}}\| \vee \|\mathtt{y}_{k+1}\| \le \|\mathtt{x}_{k+1}^\pi\| + L_{\mathrm{tay},q,\pi} B_q \le \frac{R_{\mathrm{feas}}}{2} + L_{\mathrm{tay},q,\pi} B_2 \le \frac{3}{4} R_{\mathrm{feas}}$$

Moreover, if $B_q \leq \frac{R_{\text{feas}}}{8L_\pi L_{\text{tay},q,\pi}}$ for either $q \in \{2,\infty\}$, and $B_\infty \leq \frac{R_{\text{feas}}}{8}$,

$$\begin{aligned}
&\|\mathtt{u}^\pi_{k+1} + \delta\mathtt{u}_{k+1} + \mathtt{K}^\pi_{k+1}(\delta\mathtt{y}_{k+1})\| \vee \|\mathtt{u}^\pi_{k+1} + \delta\mathtt{u}_{k+1} + \mathtt{K}^\pi_{k+1}(\delta\mathtt{y}^{\mathtt{jac}}_{k+1})\| \\
&\leq \|\mathtt{u}^\pi_{k+1}\| + \|\delta\mathtt{u}_{k+1}\|\|\mathtt{K}^\pi_{k+1}\|(\|\delta\mathtt{y}_{k+1}\| \vee \|\delta\mathtt{y}^{\mathtt{jac}}_{k+1}\|) \\
&\leq \frac{R_{\text{feas}}}{2} + \|\delta\mathtt{u}_{k+1}\| + L_\pi L_{\text{tay},q,\pi} B_2 \leq \frac{R_{\text{feas}}}{2} + B_\infty + L_\pi L_{\text{tay},q,\pi} B_2 \leq \frac{3R_{\text{feas}}}{4}.
\end{aligned}$$

This concludes the demonstration of Eq. (D.4) for $k = k+1$. Collecting our conditions, and recalling $L_{\text{tay},2,\pi} = 2L_f\kappa_{\pi,2}$, and $L_{\text{tay},\infty,\pi} = 2L_f\kappa_{\pi,1}$ we have show that if we take $B_q \leq B_{\text{tay},q,\pi}$, where

$$\begin{aligned}
B_{\text{tay},2,\pi} &= \min\left\{\frac{1}{\sqrt{40M_f L_\pi^2 \kappa_{\pi,1} M_{\text{tay},2,\pi}}}, \frac{L_f\kappa_{\pi,2}}{2M_{\text{tay},2,\pi}}, \frac{R_{\text{feas}}}{16L_\pi L_f \kappa_{\pi,2}}\right\} \\
B_{\text{tay},\text{inf},\pi} &= \min\left\{\frac{1}{40L_\pi^2 \kappa_{\pi,1} M_{\text{tay},\text{inf},\pi}}, \frac{L_f\kappa_{\pi,1}}{2M_{\text{tay},\text{inf},\pi}}, \frac{R_{\text{feas}}}{16L_\pi L_f \kappa_{\pi,1}}\right\}
\end{aligned}$$

and $B_\infty \leq \frac{R_{\text{feas}}}{8}$, then

$$\|\mathtt{y}_{k+1} - \mathtt{x}^\pi_{k+1}\| \vee \|\mathtt{y}^{\mathtt{jac}}_{k+1} - \mathtt{x}^\pi_{k+1}\| \leq L_{\text{tay},q,\pi} B_2,$$

and

$$\|\mathtt{e}^{\mathtt{jac}}_{k+1}\| = \|\mathtt{y}_{k+1} - \mathtt{y}^{\mathtt{jac}}_{k+1}\| \leq M_{\text{tay},q,\pi} B_q^2.$$

In addition, we have show that $\|\mathtt{y}_{k+1}\| \vee \|\mathtt{u}^\pi_{k+1} + \delta\mathtt{u}_{k+1} + \mathtt{K}^\pi_{k+1}(\delta\mathtt{y}_{k+1})\| \leq \frac{3}{4}R_{\text{feas}}$. This concludes the induction.

Substituing in $\mathtt{y}_k = \tilde{\mathtt{x}}^\pi_k(\mathtt{u}_{1:K})$ and using the computation of $\mathtt{y}^{\mathtt{jac}}_{k+1}$ in Lemma D.2 concludes the proof of the pertubation bounds. Moreover, the fact that Eq. (D.4) holds for all k, and consequently the conclusion of Lemma D.4 establishes the norm bounds $\|\tilde{\mathtt{x}}^\pi_k(\mathtt{u}_{1:K})\| \vee \|\tilde{\mathtt{u}}^\pi_k(\mathtt{u}_{1:K})\| \leq \frac{3R_{\text{feas}}}{4}$, and $\|\tilde{\mathbf{x}}^\pi(t \mid \mathtt{u}_{1:K})\| \leq R_{\text{feas}}$.

□

D.2. Taylor Expansion of the Cost (Lemma A.6)

Proof. Recall the definition

$$\mathcal{J}_T^{\pi,\text{disc}}(\vec{\mathtt{u}}) := V(\tilde{\mathtt{x}}^\pi_{K+1}(\vec{\mathtt{u}})) + \tau\sum_{k=1}^K Q(\tilde{\mathtt{x}}^\pi_k(\vec{\mathtt{u}}), \tilde{\mathtt{u}}^\pi_k(\vec{\mathtt{u}}), t_k).$$

Define the shorthand

$$\begin{aligned}
\delta\tilde{\mathtt{x}}_k &:= \tilde{\mathtt{x}}^\pi_k(\delta\mathtt{u}_{1:K} + \mathtt{u}^\pi_{1:K}) - \mathtt{x}^\pi_k \\
\delta\tilde{\mathtt{x}}^{\text{jac}}_k &:= \tilde{\mathtt{x}}^{\pi,\text{jac}}_k(\delta\mathtt{u}_{1:K} + \mathtt{u}^\pi_{1:K})
\end{aligned}$$

$$\begin{aligned}
&\mathcal{J}_T^{\pi,\text{disc}}(\vec{\mathtt{u}}) - \mathcal{J}_T^{\pi,\text{disc}}(\delta\mathtt{u}_{1:K} + \mathtt{u}^\pi_{1:K}) \\
&:= V(\mathtt{x}^\pi_{K+1} + \delta\tilde{\mathtt{x}}_{K+1}) - V(\mathtt{x}^\pi_{K+1}) + \tau\sum_{k=1}^K Q(\mathtt{x}^\pi_k + \delta\tilde{\mathtt{x}}_k, \delta\mathtt{u}_k + \mathtt{K}^\pi_k\delta\tilde{\mathtt{x}}_k + \mathtt{u}^\pi_k, t_k) - Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k)
\end{aligned}$$

Notice that Proposition A.5 and feasibility of π implies that

$$\|\mathtt{x}^\pi_k + \delta\tilde{\mathtt{x}}_k\|\|\mathtt{x}^\pi_k\| \vee \|\delta\mathtt{u}_k + \mathtt{K}^\pi_k\mathtt{e}_k\| \vee \|\mathtt{u}^\pi_k\| \leq R_{\text{feas}} \tag{D.9}$$

Hence a Taylor expansion and Assumption 4.2 imply

$$\begin{aligned}
&|V(\mathtt{x}^\pi_{K+1} + \delta\tilde{\mathtt{x}}_{K+1}) - V(\mathtt{x}^\pi_{K+1}) - \langle \partial_x V(\mathtt{x}^\pi_{K+1}), \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle| \\
&\le \frac{1}{2}\sup_{\alpha\in[0,1]} \|\partial_{xx} V(\mathtt{x}^\pi_{K+1} + \alpha\delta\tilde{\mathtt{x}}_{K+1})\|\|\delta\tilde{\mathtt{x}}_{K+1}\|^2 + |\langle \partial_x V(\mathtt{x}^\pi_{K+1}), \delta\tilde{\mathtt{x}}_{K+1} - \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_{K+1}\rangle| \\
&\le \frac{M_{\mathrm{cost}}}{2}\|\delta\tilde{\mathtt{x}}_{K+1}\|^2 + L_{\mathrm{cost}}\|\delta\tilde{\mathtt{x}}_{K+1} - \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_{K+1}\|.
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\left|Q(\mathtt{x}^\pi_k + \delta\tilde{\mathtt{x}}_k, \delta\mathtt{u}_k + \mathtt{K}^\pi_k\delta\tilde{\mathtt{x}}_k, t_k) - Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k) - \langle \partial_x Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k), \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle - \langle \partial_u Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k), \delta\mathtt{u}_k + \mathtt{K}_k\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle\right| \\
&\le \left|\langle \partial_x Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k), \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle| + |\langle \partial_u Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k), \mathtt{K}_k\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle\right| \\
&\quad + \frac{1}{2}M_{\mathrm{cost}}\left(\|\delta\tilde{\mathtt{x}}_k\|^2 + \|\delta\mathtt{u}_k + \mathtt{K}_k\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\|^2\right) \\
&\le L_{\mathrm{cost}}(1+L_\pi)\|\delta\tilde{\mathtt{x}}_k - \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\| + \frac{1}{2}M_{\mathrm{cost}}\left((1+2L_\pi^2)\|\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\|^2 + 2\|\delta\mathtt{u}_k\|^2\right) \\
&\le 2L_\pi L_{\mathrm{cost}}\|\delta\tilde{\mathtt{x}}_k - \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\| + \frac{1}{2}M_{\mathrm{cost}}\left(3L_\pi^2\|\mathtt{e}_k\|^2 + 2\|\delta\mathtt{u}_k\|^2\right). \qquad (L_\pi \ge 1)
\end{aligned}$$

Then, from Lemma C.7,

$$\begin{aligned}
&\langle \delta\mathtt{u}_{1:K}, \nabla\mathcal{J}^{\pi,\mathrm{disc}}_T(\mathbf{u})\rangle \\
&= \langle \partial_x V(\mathtt{x}^\pi_{K+1}), \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle + \tau\sum_{k=1}^K \langle \partial_x Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k), \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle + \langle \partial_u Q(\mathtt{x}^\pi_k, \mathtt{u}^\pi_k, t_k), \delta\mathtt{u}_k + \mathtt{K}_k\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\rangle.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\|\mathcal{J}^{\pi,\mathrm{disc}}_T(\delta\mathtt{u}_{1:K} + \mathtt{u}^\pi_{1:K}) - \mathcal{J}^{\pi,\mathrm{disc}}_T(\mathtt{u}^\pi_{1:K}) - \langle \delta\mathtt{u}_{1:K}, \nabla\mathcal{J}^{\pi,\mathrm{disc}}_T(\pi)\rangle\| \\
&\le \frac{M_{\mathrm{cost}}}{2}\|\delta\tilde{\mathtt{x}}_{K+1}\|^2 + L_{\mathrm{cost}}\|\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_{K+1} - \delta\tilde{\mathtt{x}}_{K+1}\| + \tau\sum_{k=1}^K 2L_\pi L_{\mathrm{cost}}\|\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k - \delta\tilde{\mathtt{x}}_k\| + \frac{1}{2}M_{\mathrm{cost}}\left(3L_\pi^2\|\delta\tilde{\mathtt{x}}_k\|^2 + 2\|\delta\mathtt{u}_k\|^2\right) \\
&\le \frac{M_{\mathrm{cost}}}{2}(1+3L_\pi^2 T)\max_{k\in[K+1]}\|\delta\tilde{\mathtt{x}}_k\|^2 + L_{\mathrm{cost}}(1+2L_\pi T)\max_{k\in[K+1]}\|\delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k - \delta\tilde{\mathtt{x}}_k\| + L_\pi L_{\mathrm{cost}}\underbrace{\tau\sum_{t=1}^T\|\delta\mathtt{u}_k\|^2}_{=B_2} \\
&= \frac{M_{\mathrm{cost}}}{2}(1+3L_\pi^2 T)\max_{k\in[K+1]}\|\delta\tilde{\mathtt{x}}_k\|^2 + L_{\mathrm{cost}}(1+2L_\pi T)\max_{k\in[K+1]}\|\delta\tilde{\mathtt{x}}_k - \delta\tilde{\mathtt{x}}^{\mathrm{jac}}_k\| + 2L_\pi L_{\mathrm{cost}}B_2^2.
\end{aligned}$$

From Proposition A.5,

$$\begin{aligned}
\max_{k\in[K+1]}\|\delta\tilde{\mathtt{x}}_k - \delta\tilde{\mathtt{x}}_k\| &\le M_{\mathrm{tay},2,\pi}B_2^2 \\
\max_{k\in[K+1]}\|\mathtt{e}_k\|^2 &\le 4L_f^2\kappa_{\pi,2}^2 B_2^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\|\mathcal{J}^{\pi,\mathrm{disc}}_T(\delta\mathtt{u}_{1:K} + \mathtt{u}^\pi_{1:K}) - \mathcal{J}^{\pi,\mathrm{disc}}_T(\mathtt{u}^\pi_{1:K}) - \langle \delta\mathtt{u}_{1:K}, \nabla\mathcal{J}^{\pi,\mathrm{disc}}_T(\pi)\rangle\| \\
&\le \underbrace{(2M_{\mathrm{cost}}L_f^2\kappa_{\pi,2}^2(1+3L_\pi^2 T)M_{\mathrm{tay},2,\pi} + L_{\mathrm{cost}}(1+2L_\pi T)M_{\mathrm{tay},2,\pi} + 2L_\pi L_{\mathrm{cost}})}_{:=M_{\mathcal{J},\mathrm{tay},\pi}}B_2^2.
\end{aligned}$$

□

D.3. Proof of Lemma A.7

We begin with the following lemma, which we show shortly below.

Lemma D.6. *Consider the setting of Proposition A.5, with $B_\infty \le \min\{B_{\mathrm{tay,inf},\pi}, R_{\mathrm{feas}}/8\}$. Let π' denote the policy with gains $\mathtt{K}_k^\pi$ and inputs $\mathtt{u}_k^{\pi'} = \mathtt{u}_k = \mathtt{u}_k^\pi + \delta\mathtt{u}_k$. Then,*

$$\tau \sum_{k=1}^{K} \|\mathtt{A}_{\mathrm{cl},k}^{\pi'} - \mathtt{A}_{\mathrm{cl},k}^{\pi}\| \le 12TM_f L_\pi (1 + L_f K_\pi) B_\infty.$$

With this lemma, we turn to the proof of Lemma A.7. Notice that, as π and π' have the same gains, $L_\pi = L_{\pi'}$. Therefore, following the proof of Lemma A.1 (see, specifically, the proof of Claim G.7), we have that for $\|\mathtt{A}_{\mathrm{cl},k}^{\pi'} - \mathbf{I}\|$ and $\|\mathtt{A}_{\mathrm{cl},k}^{\pi} - \mathbf{I}\|$ are both at most $\kappa := 3\tau L_f L_\pi$ for $\tau \le 1/6L_f L_\pi$.

Let us now construct an interpolating curves $\mathtt{X}_k(s)$ with $\mathtt{X}_k(0) = \mathtt{A}_{\mathrm{cl},k}^{\pi}$ and $\mathtt{X}_k(1) = \mathtt{A}_{\mathrm{cl},k}^{\pi'}$, and define the interpolating Lyapunov function

$$\mathtt{\Lambda}_{K+1}(s) = \mathbf{I}, \quad \mathtt{\Lambda}_k(s) = \mathtt{X}_k(s)^\top \mathtt{\Lambda}_{k+1} \mathtt{X}_k(s) + \tau \mathbf{I},$$

Define

$$\begin{aligned}
\Delta &= 3 \sup_{s\in[0,1]} \max\{1, 2\kappa\} \sum_{k=k_0}^{K} \|\mathtt{X}_k'(s)\| \\
&= \max\{1, 2\kappa\} \sum_{k=k_0}^{K} \|\mathtt{A}_{\mathrm{cl},k}^{\pi'} - \mathtt{A}_{\mathrm{cl},k}^{\pi}\| \\
&= \max\{3, 18L_f L_\pi\} \sum_{k=k_0}^{K} \|\mathtt{A}_{\mathrm{cl},k}^{\pi'} - \mathtt{A}_{\mathrm{cl},k}^{\pi}\|.
\end{aligned}$$

and recall the the shorthand $\|\mathtt{\Lambda}_{k_0:K+1}(s)\|_{\mathrm{max,op}} := \max_{k\in[k_0:K+1]} \|\mathtt{\Lambda}_k(s)\|$. Then, as long as $\|\mathtt{\Lambda}_{k_0:K+1}(0)\|_{\mathrm{max,op}}\Delta < 1$, Lemma F.13 (re-indexing to terminate the backward recursion at k_0 instead of 1) implies

$$\|\mathtt{\Lambda}_{k_0:K+1}(1)\|_{\mathrm{max,op}} \le (1 - \|\mathtt{\Lambda}_{k_0:K+1}(0)\|_{\mathrm{max,op}}\Delta)^{-1} \|\mathtt{\Lambda}_{k_0:K+1}(0)\|_{\mathrm{max,op}}.$$

We see that $\|\mathtt{\Lambda}_{k_0:K+1}(0)\|_{\mathrm{max,op}} = \|\mathtt{\Lambda}_{k_0:K+1}^{\pi}\|_{\mathrm{max,op}} = K_{\pi,\star}$, and $\|\mathtt{\Lambda}_{k_0:K+1}(1)\|_{\mathrm{max,op}} = \|\mathtt{\Lambda}_{k_0:K+1}^{\pi'}\|_{\mathrm{max,op}} = \mu_{\pi',\star}$. Thus, combining with the inequality $(1-x)^{-1} \le 1 + 2x$ for $x \in [0, 1/2]$, we have that as long as $\Delta\mu_{\pi,\star} \le 1/2$,

$$\mu_{\pi',\star} \le (1 + 2\Delta\mu_{\pi,\star})\mu_{\pi,\star}.$$

Lastly, we can bound

$$\begin{aligned}
2\Delta\mu_{\pi,\star} &= \max\{6, 36L_f L_\pi\}\mu_{\pi,\star} \sum_{k=k_0}^{K} \|\mathtt{A}_{\mathrm{cl},k}^{\pi'} - \mathtt{A}_{\mathrm{cl},k}^{\pi}\| \\
&\le \underbrace{\max\{6, 36L_f L_\pi\}\mu_{\pi,\star} \cdot 12TM_f L_\pi (1 + L_f K_\pi) B_\infty}_{:=1/B_{\mathrm{stab},\pi}} B_2. \qquad \text{(Lemma D.6)}
\end{aligned}$$

In sum, for $B_\infty \le B_{\mathrm{stab},\pi}$, we have $\mu_{\pi',\star} \le (1 + B_\infty/B_{\mathrm{stab},\pi})\mu_{\pi,\star}$, which concludes the proof.

Proof of Lemma D.6. Due to Proposition A.5, and the fact that $\mathbf{x}^{\pi'}(t) = \tilde{\mathbf{x}}^\pi(t \mid \mathtt{u}_{1:K})$ and $\mathbf{u}^{\pi'}(t) = \tilde{\mathtt{u}}_{k(t)}^\pi(\mathtt{u}_{1:K})$, we have that

$$\forall t \in [0, T], \quad \|\mathbf{x}^{\pi'}(t)\| \vee \|\mathbf{u}^{\pi'}(t)\| \le R_{\mathrm{feas}}. \tag{D.10}$$

Moreover each initial condition ξ with norm $\|\xi\| = 1$, we have that from Lemma C.10 and the definitions of $\mathtt{A}^{\pi'}_{\mathrm{cl},k}, \mathtt{A}^{\pi}_{\mathrm{cl},k}$ from Definition C.8 that

$$(\mathtt{A}^{\pi'}_{\mathrm{cl},k} - \mathtt{A}^{\pi}_{\mathrm{cl},k})\xi = \mathbf{z}_2(\tau) - \mathbf{z}_1(\tau),$$

where

$$\mathbf{z}_2(0) = \mathbf{z}_1(0) = \xi,$$

and where $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{z}_2(t) = \mathbf{A}^{\pi'}_{\mathrm{ol}}(t_k + t)\mathbf{z}_2(t) + \mathbf{B}^{\pi'}_{\mathrm{ol}}(t_k + t)\mathtt{K}_k\xi$, and where $\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{z}_1(t) = \mathbf{A}^{\pi}_{\mathrm{ol}}(t_k + t)\mathbf{z}_2(t) + \mathbf{B}^{\pi}_{\mathrm{ol}}(t_k + t)\mathtt{K}_k\xi$.

By the Picard Lemma, Lemma C.9, and by bounding $\|\mathbf{A}^{\pi'}_{\mathrm{ol}}(t)\| \le L_f$ by Eq. (D.10) and Assumption 4.1, it follows that

$$\begin{aligned}\|(\mathtt{A}^{\pi'}_{\mathrm{cl},k} - \mathtt{A}^{\pi}_{\mathrm{cl},k})\xi\| &\le \exp(\tau L_f)\int_0^\tau (\|\mathbf{A}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{A}^{\pi}_{\mathrm{ol}}(t_k + t)\|\|\mathbf{z}_1(t)\| + \|\mathbf{B}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{B}^{\pi}_{\mathrm{ol}}(t_k + t)\|\|\mathtt{K}_k\|\|\xi\|)\mathrm{d}t \\ &\le \exp(\tau L_f)\int_0^\tau (\|\mathbf{A}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{A}^{\pi}_{\mathrm{ol}}(t_k + t)\|\|\mathbf{z}_1(t)\| + \|\mathbf{B}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{B}^{\pi}_{\mathrm{ol}}(t_k + t)\|L_\pi)\mathrm{d}t\end{aligned}$$

Set $L_{\mathrm{ol}} = \exp(\tau L_f)$. Following the computation in , we can bound $\sup_{t\in[0,\tau}\|\mathbf{z}_1(t)\| \le \|\mathtt{A}^{\pi}_{\mathrm{cl},k}\| = \|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k+1,k}\| \le 5/3$ provided $\tau \le 1/4L_fL_\pi$ (recall we assume $L_\pi \ge 1$). Hence,

$$\|(\mathtt{A}^{\pi'}_{\mathrm{cl},k} - \mathtt{A}^{\pi}_{\mathrm{cl},k})\xi\| \le L_{\mathrm{ol}}\int_0^\tau (\frac{5}{3}\|\mathbf{A}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{A}^{\pi}_{\mathrm{ol}}(t_k + t)\| + \|\mathbf{B}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{B}^{\pi}_{\mathrm{ol}}(t_k + t)\|L_\pi)\mathrm{d}t$$

Finally, by the smoothness on the dynamics Assumption 4.1 and invoking Eq. (D.10) and feasibility of π to ensure all relevant (x, u) pairs are feasible, we have

$$\begin{aligned}\|\mathbf{A}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{A}^{\pi}_{\mathrm{ol}}(t_k + t)\| &= \|\partial_x f_{\mathrm{dyn}}(\mathbf{x}^{\pi'}(t), \mathbf{u}^{\pi'}(t)) - \partial_x f_{\mathrm{dyn}}(\mathbf{x}^{\pi}(t), \mathbf{u}^{\pi}(t))\| \\ &\le M_f\left(\|\mathbf{x}^{\pi'}(t) - \mathbf{x}^{\pi}(t)\| + \|\mathbf{u}^{\pi'}(t) - \mathbf{u}^{\pi}(t))\|\right) \\ &\le M_f\left(\|\mathbf{x}^{\pi'}(t) - \mathbf{x}^{\pi}(t)\| + \|\delta\mathtt{u}_k\|\right).\end{aligned}$$

Applying a similar bound to the term $\|\mathbf{B}^{\pi'}_{\mathrm{ol}}(t_k + t) - \mathbf{B}^{\pi}_{\mathrm{ol}}(t_k + t)\|$, we conclude

$$\begin{aligned}\|\mathtt{A}^{\pi'}_{\mathrm{cl},k} - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| &\le \sup_{\xi:\|\xi\|} = \|(\mathtt{A}^{\pi'}_{\mathrm{cl},k} - \mathtt{A}^{\pi}_{\mathrm{cl},k})\xi\| \\ &\le L_{\mathrm{ol}}M_f(\frac{5}{3} + L_\pi)\int_0^\tau \left(\|\mathbf{x}^{\pi'}(t) - \mathbf{x}^{\pi}(t)\| + \|\delta\mathtt{u}_k\|\right)\mathrm{d}t \\ &\le \tau L_{\mathrm{ol}}M_f(\frac{5}{3} + L_\pi)\left((1 + \tau L_{\mathrm{ol}}L_f)\|\delta\mathtt{u}_k\| + L_{\mathrm{ol}}(1 + \tau L_\pi L_f)\|\mathbf{x}^{\pi'}(t_k) - \mathbf{x}^{\pi}(t_k)\right) && \text{(Lemma D.5)} \\ &\le \tau L_{\mathrm{ol}}M_f(\frac{5}{3} + L_\pi)\left(\frac{5L_{\mathrm{ol}}}{4}\|\delta\mathtt{u}_k\| + \frac{5}{4}L_{\mathrm{ol}}\|\mathbf{x}^{\pi'}(t_k) - \mathbf{x}^{\pi}(t_k)\right) && (L_{\mathrm{ol}} \ge 1, \tau \le 1/4L_fL_\pi \le 1/4L_f) \\ &= \tau\frac{5L^2_{\mathrm{ol}}}{4}M_f(\frac{5}{3} + L_\pi)\left(\|\delta\mathtt{u}_k\| + \|\mathbf{x}^{\pi'}(t_k) - \mathbf{x}^{\pi}(t_k)\|\right) \\ &= \tau M_f \cdot \frac{5e^{1/2}}{4} \cdot \frac{8}{3} \cdot L_\pi\left(\|\delta\mathtt{u}_k\| + \|\mathbf{x}^{\pi'}(t_k) - \mathbf{x}^{\pi}(t_k)\|\right) && (L_\pi \ge 1, \tau L_f \le 1/4) \\ &\le 6\tau M_f L_\pi\left(\|\delta\mathtt{u}_k\| + \|\mathbf{x}^{\pi'}(t_k) - \mathbf{x}^{\pi}(t_k)\|\right) \\ &\le 6\tau M_f L_\pi\left(\|\delta\mathtt{u}_k\| + 2L_f\kappa_{\pi,1}B_\infty\right). && \text{(Proposition A.5)}\end{aligned}$$

Summing the bound, and using $K\tau = T$, we have

$$\begin{aligned}\sum_{k=1}^K \|\mathtt{A}^{\pi'}_{\mathrm{cl},k} - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| &\le 6M_fL_\pi\left(\tau\sum_{k=1}^K\|\delta\mathtt{u}_k\| + 2TL_f\kappa_{\pi,1}B_\infty\right) \\ &\le 6M_fL_\pi\left(K\tau B_\infty + 2TL_f\kappa_{\pi,1}B_\infty\right) \\ &\le 12TM_fL_\pi(1 + L_f\kappa_{\pi,1})B_\infty.\end{aligned}$$

□

D.4. Proof of Lemma A.8

Proof. Using Condition F.3 and and $1 \vee \|\mathtt{K}_k\| \le L_\pi$,

$$\begin{aligned}\|(\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi))_k\| &\le \tau \|Q_u(\mathtt{x}_k^\pi, \mathtt{u}_k^\pi, t_k)^2\|^2 + \|\Psi_{\mathrm{cl},K+1,k}^\pi)^\top V_x(\mathtt{x}_{K+1}^\pi\| \\ &\quad + \left\| \tau \sum_{j=k+1}^{K} (\Psi_{\mathrm{cl},j,k}^\pi)^\top (Q_x(\mathtt{x}_j^\pi, \mathtt{u}_j^\pi, t_j) + (\mathtt{K}_j^\pi)^\top Q_u(\mathtt{x}_j^\pi, \mathtt{u}_j^\pi, t_j)) \right\| \\ &\le \tau L_{\mathrm{cost}} + \tau \|\Psi_{\mathrm{cl},K+1,k}^\pi\| L_{\mathrm{cost}}^2 + 2L_\pi^2 L_{\mathrm{cost}}^2 \left(\tau \sum_{j=k+1}^{K} \|\Psi_{\mathrm{cl},j,k}^\pi\| \right)^2.\end{aligned}$$

Using $\tau \le 1/4L_f$ and Lemma I.3, we can bound $\|\Psi_{\mathrm{cl},j,k}^\pi\| = \|\Phi_{\mathrm{cl},j,k+1}^\pi \mathtt{B}_{\mathrm{ol},k}^\pi\| \le \tau L_f \exp(\tau L_f) \|\Phi_{\mathrm{cl},j,k+1}^\pi\| \le \tau L_f \exp(1/4) \|\Phi_{\mathrm{cl},j,k+1}^\pi\|$. As $\exp(1/2) \le 3/4$, we conclude that

$$\begin{aligned}\|(\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi))_k\| &\le \tau L_{\mathrm{cost}} + \frac{3}{2}\tau^2 L_f L_{\mathrm{cost}} + 3\tau L_\pi L_{\mathrm{cost}} \left(\tau \sum_{j=k+1}^{K} \|\Phi_{\mathrm{cl},j,k}^\pi\| \right)^2. \\ &\le \tau L_{\mathrm{cost}} + \frac{3}{2}\tau L_f L_{\mathrm{cost}} \|\Phi_{\mathrm{cl},K+1,k+1}^\pi\| + 3L_\pi L_{\mathrm{cost}} \kappa_{\pi,1}.\end{aligned}$$

Using $\|\Phi_{\mathrm{cl},K+1,k+1}^\pi\| \le \kappa_{\pi,\infty}$gives the boudn $\|(\nabla \mathcal{J}_T^{\mathrm{disc}}(\pi))_k\| \le \tau L_{\mathrm{cost}}(1 + \frac{3L_f}{2}\kappa_{\pi,\infty} + 3L_\pi \kappa_{\pi,1}) =: L_{\nabla,\pi,\infty}$. □

E. Estimation Proofs

E.1. Estimation of Markov Parameters: Proof of Proposition A.9

We begin with two standard concentration inequalities.

Lemma E.1. *Let $(y_i, x_i, w_i)_{i=1}^n$ be an independent sequence of triples of random vectors in with $y_i, x_i \in \mathbb{R}^d$, $w \in \mathbb{R}^{d'}$ and suppose that $y_i \mid x_i, w_i \sim \mathcal{N}(x_i, \sigma^2 \mathbf{I}_d)$ and $\max_i \|w_i\| \le R$. Then,*

$$\mathbb{P}\left[\left\| \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i) w_i \right\| \le R\sigma \sqrt{2 \cdot \frac{d \log 5 + \log((d'+1)/\delta)}{N}} \right] \ge 1 - \delta.$$

Proof of Lemma E.1. By a standard covering argument (see, e.g. Vershynin (2018, Chapter 4)), there exists a finite covering $\mathcal{T}$ of unit vectors $z \in \mathbb{R}^d$ such that (a) $\log |\mathcal{T}| \le d \log 5$ and (b), for all vectors $v \in \mathbb{R}^d$,

$$\|v\| \le 2 \sup_{z \in \mathcal{T}} \langle v, z \rangle.$$

Hence,

$$\left\| \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i) w_i \right\|_{\mathrm{op}} \le 2 \sup_{z \in \mathcal{T}} \left\| \frac{1}{N} \sum_{i=1}^{N} \langle z, y_i - x_i \rangle \cdot w_i \right\| = 2\sigma \sup_{z \in \mathcal{T}} \left\| \frac{1}{N} \sum_{i=1}^{N} \xi_i(z) \cdot w_i \right\|, \tag{E.1}$$

where above we define $\xi_i(z) := \sigma^{-1} \langle z, y_i - x_i \rangle$. Notice that $\xi_i(z) \mid w_i$ are standard Normal random variables. Thus, by standard Gaussian concentration (e.g. Boucheron et al. (2013, Chapter 2)),

$$\mathbb{P}\left[\left\| \frac{1}{N} \sum_{i=1}^{N} \xi_i(z) w_i \right\| \le R \sqrt{2 \frac{\log((d'+1)/\delta)}{N}} \right] \ge 1 - \delta.$$

Hence, union bounding over $z \in \mathcal{T}$, bounding $|\mathcal{T}| \le d \log 5$, Eq. (E.1) implies the desired bound.

$$\mathbb{P}\left[\left\| \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i) w_i \right\| \le R\sigma \sqrt{2 \frac{d \log 5 + \log((d'+1)/\delta)}{N}} \right] \ge 1 - \delta.$$

□

Lemma E.2 (Assymetric Matrix Hoeffding)**.** *Let $X_1, \dots, X_n$ be an independent sequence of matrices in $\mathbb{R}^{d_1 \times d_2}$ with $\|X_i\| \le R$. Then,*

$$\mathbb{P}\left[\frac{1}{N}\left\|\sum_i X_i - \mathbb{E}[X_i]\right\| \le 4R\sqrt{\frac{\log(\frac{d_1+d_2}{\delta})}{N}}\right] \ge 1-\delta.$$

Proof of Lemma E.2. By recentering $X_i \leftarrow X_i - \mathbb{E}[X_i]$, we may assume $\mathbb{E}[X_i] = 0$ and $\|X_i\| \le 2R$. Define the Hermitian dilation

$$Y_i = \begin{bmatrix} 0 & X_i \\ X_i^\top & 0 \end{bmatrix}.$$

Then

$$Y_i^2 = \begin{bmatrix} X_i X_i^\top & 0 \\ 0 & X_i^\top X_i \end{bmatrix} \preceq \|X_i\|^2 \mathbf{I}_{d_1+d_i} \le 4R^2 \mathbf{I}_{d_1+d_2}$$

Applying standard Matrix Hoeffding Tropp (2012, Theorem 1.4) for Hermitian matrices to the Y_i's yields

$$\mathbb{P}\left[\left\|\sum_i Y_i\right\| \ \ge \ t\right] \ \le \ (d_1+d_2)e^{-\frac{t^2}{32NR^2}}.$$

Hence, by rearranging,

$$\mathbb{P}\left[\left\|\sum_i Y_i\right\| \le 4R\sqrt{2N\log(\frac{d_1+d_2}{\delta})}\right] \ge 1-\delta$$

As $\|\sum_i Y_i\| = \sqrt{2}\,\|\sum_i X_i\|$, we conclude

$$\mathbb{P}\left[\frac{1}{N}\left\|\sum_i X_i\right\| \le 4R\sqrt{\frac{\log(\frac{d_1+d_2}{\delta})}{N}}\right] \ge 1-\delta.$$

□

We now turn to concluding the proof of Proposition A.9. We begin with a claim which bounds $\max_k \|\hat{\mathtt{x}}_k - \mathtt{x}_k^\pi\|$. Throughout, $d_\star := \max\{d_{\mathsf{x}}, d_{\mathsf{u}}\}$.

Claim E.1. *With probability at least $1-\delta/3$, the following bound holds*

$$\max_k \|\hat{\mathtt{x}}_k - \mathtt{x}_k^\pi\| \le \sigma_{\mathrm{orac}}\sqrt{2\frac{d_\star \log 5 + \log(6(K+1)/\delta)}{N}} \le \mathrm{Err}_{\hat{x}}(\delta).$$

Proof. The result follows directly from Lemma E.1, with $w_i = 1 \in \mathbb{R}$ for each i. □

Proof of Proposition A.9. Throughout, suppose the event of Claim E.1 holds. We also note that

$$\|\mathtt{w}_j^{(i)}\| \le \sigma_w\sqrt{d_{\mathsf{u}}} \le \sigma_w\sqrt{d_\star} \text{ a.s..} \tag{E.2}$$

This covers the first inequality of the proposition. To bound the error on the transition operator, tet us fix indices j, k; we perform a union bound at the end of the proof. For each perturbation sampled perturbation $\mathtt{w}_{1:K}^{(i)}$, define a perturbed control input

$$\check{\mathtt{u}}_k^{(i)} = \mathtt{u}_k^\pi + \mathtt{w}_k^{(i)} + \mathtt{K}_k^\pi(\mathtt{x}_k^\pi - \hat{\mathtt{x}}_k), \quad \tilde{\mathtt{u}}_k^i = \mathtt{u}_k^\pi + \mathtt{w}_k^{(i)} - \mathtt{K}_k^\pi\hat{\mathtt{x}}_k.$$

and observe that

$$\mathtt{x}^{\pi}_{\mathrm{orac},k}\left(\tilde{\mathtt{u}}^{(i)}_{1:K}\right) = \tilde{\mathtt{x}}_k\left(\tilde{\mathtt{u}}^{(i)}_{1:K}\right), \quad \forall k \in [K]. \tag{E.3}$$

Hence, we have that $\mathtt{y}_k$ defined in Line 7 satisfies

$$\mathtt{y}^{(i)}_k \sim \mathcal{N}(\tilde{\mathtt{x}}_k\left(\tilde{\mathtt{u}}^{(i)}_{1:K}\right), \sigma^2_{\mathrm{orac}}\mathbf{I}_{d_{\mathtt{x}}}).$$

Now, define the terms

$$\mathtt{z}^{(i)}_k := \mathtt{y}^{(i)}_k - \mathtt{x}^{\pi}_k$$

Lastly, let $\mathbb{E}_{\mathrm{orac}}[\cdot]$ denote expectations with respect to the Gaussian noise of the oracle, (conditioning on $\mathtt{w}^{(i)}_{1:K}$) while $\mathbb{E}[\cdot]$ denotes total expectation. We argue an error bound on

$$\begin{aligned}\left\|\frac{\sigma_w^{-2}}{N}\sum_{i=1}^{N}\mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j - \Psi^{\pi}_{\mathrm{cl},k,j}\right\|_{\mathrm{op}} \le \underbrace{\frac{\sigma_w^{-2}}{N}\left\|\sum_{i=1}^{N}\mathbb{E}_{\mathrm{orac}}[\mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j] - \mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j\right\|_{\mathrm{op}}}_{=\mathrm{Term}_1} \\ + \underbrace{\frac{\sigma_w^{-2}}{N}\left\|\sum_{i=1}^{N}\mathbb{E}[\mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j] - \mathbb{E}_{\mathrm{orac}}[\mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j]\right\|_{\mathrm{op}}}_{=\mathrm{Term}_2} \\ + \underbrace{\left\|\frac{\sigma_w^{-2}}{N}\sum_{i=1}^{N}\mathbb{E}[\mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j] - \Psi^{\pi}_{\mathrm{cl},k,j}\right\|_{\mathrm{op}}}_{=\mathrm{Term}_3},\end{aligned}$$

which essentially bounds the estiation error of $\Psi^{\pi}_{\mathrm{cl},k,j}$ in the absence of observation noise.

Bounding Term_1**.** Applying Lemma E.1 with $\|\mathtt{w}^{(i)}_j\| \le \sqrt{d_\star}\sigma_w$, we have that with probability $1-\delta/3$,

$$\begin{aligned}\mathrm{Term}_1 &\le \frac{\sigma_{\mathrm{orac}}}{\sigma_w}\sqrt{2d_\star \cdot \frac{d_\star \log 5 + \log(6(d_\star+1)/\delta)}{N}} \\ &\le \frac{\sigma_{\mathrm{orac}}}{\sigma_w}\sqrt{2d_\star \cdot \frac{d_\star \log 5 + \log(12d_\star/\delta)}{N}}.\end{aligned}$$

Bounding Term_2**.** On the event of Claim E.1, then as long as $\mathrm{Err}_{\hat{x}}(\delta) \le \sigma_w\sqrt{d_\star/L_\pi}$

$$\|\check{\mathtt{u}}^{(i)}_k - \mathtt{u}^{\pi}_k\| = \|\mathtt{w}^{(i)}_k + \mathtt{K}^{\pi}_k(\mathtt{x}^{\pi}_k - \hat{\mathtt{x}}_k)\| \le \|\mathtt{w}^{(i)}_k\| + L_\pi \mathrm{Err}_{\hat{x}}(\delta) \le \sigma_w\sqrt{d_\star} + L_\pi \mathrm{Err}_{\hat{x}}(\delta) \le 2\sigma_w\sqrt{d_\star}.$$

Notice that as $\mathrm{Err}_{\hat{x}}(\delta) = \sigma_{\mathrm{orac}}\sqrt{2d_\star\iota(\delta)/N}$, $\mathrm{Err}_{\hat{x}}(\delta) \le \sigma_w\sqrt{d_\star/L_\pi}$ holds for $N \ge (\sigma_{\mathrm{orac}}/\sigma_w)^2 2L_\pi\iota(\delta)$, i.e. which holds for when π is estimation-friendly.

Moroever, when π is estimation-friendly, $B_{\mathrm{tay,inf},\pi} \ge 2\sigma_w\sqrt{d_\star}$, so the conditions of Proposition A.5 hold. Therefore,

$$\|\mathtt{z}^{(i)}_k\| \le 2\sigma_w L_{\mathrm{tay},\infty,\pi}\sqrt{d_\star}.$$

and thus

$$\|\mathtt{w}^{(i)}_j\mathtt{z}^{(i)}_k\| \le 2d_\star\sigma_w^2 L_{\mathrm{tay},\infty,\pi}.$$

Applying Lemma E.2 with $X_i \leftarrow \mathtt{z}^{(i)}_k\mathtt{w}^{(i)\top}_j$ with $R \leftarrow 2d_\star L_{\mathrm{tay},\infty,\pi}\sigma_w^2$, it holds that with probability $1-\delta/3$ that

$$\mathrm{Term}_2 \le \sigma_w^{-2}\cdot 8L_{\mathrm{tay},\infty,\pi}\sigma_w^2 d_{\mathtt{u}}\sqrt{\frac{\log(\frac{3(d_{\mathtt{u}}+d_{\mathtt{x}})}{\delta})}{N}} \le 8L_{\mathrm{tay},\infty,\pi}d_\star\sqrt{\frac{\log(6d_\star/\delta)}{N}}.$$

Bounding Term_3 . As established in the bound on Term_2, the conditions of Proposition A.5 hold, and $\|\check{\mathtt{u}}_k^{(i)} - \mathtt{u}_k^\pi\| \le 2\sigma_w\sqrt{d_\star}$.. Therefore,

$$\|\mathtt{z}_k^{(i)} - \sum_{\ell=1}^k \Psi^\pi_{\mathrm{cl},k,\ell}(\check{\mathtt{u}}_\ell^{(i)} - \mathtt{u}_\ell^\pi)\| \le 4\sigma_w^2 M_{\mathrm{tay},2,\pi} d_\star.$$

Consequently, bounding $\|\mathtt{w}_j^{(i)}\| \le \sigma_w d_\star$,

$$\frac{1}{\sigma_w^2}\|\mathtt{z}_k^{(i)}\mathtt{w}_j^{(i)} - \sum_{\ell=1}^k \Psi^\pi_{\mathrm{cl},k,\ell}(\check{\mathtt{u}}_\ell^{(i)} - \mathtt{u}_\ell^\pi)\mathtt{w}_j^{(i)}\| \le 4\sigma_w M_{\mathrm{tay},2,\pi} d_\star^{3/2}$$

and thus, by Jensen's inequality,

$$\begin{aligned}
4\sigma_w M_{\mathrm{tay},2,\pi} d_\star^{3/2} &\ge \frac{1}{N\sigma_w^2}\|\sum_{i=1}\mathbb{E}[\mathtt{z}_k^{(i)}\mathtt{w}_j^{(i)}] - \sum_{\ell=1}^k \mathbb{E}[\Psi^\pi_{\mathrm{cl},k,\ell}(\check{\mathtt{u}}_\ell^{(i)} - \mathtt{u}_\ell^\pi)\mathtt{w}_j^{(i)}]\| \\
&= \frac{1}{N\sigma_w^2}\|\sum_{i=1}\mathbb{E}[\mathtt{z}_k^{(i)}\mathtt{w}_j^{(i)}] - \sum_{\ell=1}^k \Psi^\pi_{\mathrm{cl},k,\ell}\mathbb{E}[(\mathtt{w}_k^{(i)}[\ell] + \mathtt{K}_k^\pi(\mathtt{x}_\ell^\pi - \hat{\mathtt{x}}_\ell))\mathtt{w}_j^{(i)}]\| \\
&= \frac{1}{N\sigma_w^2}\|\sum_{i=1}\mathbb{E}[\mathtt{z}_k^{(i)}\mathtt{w}_j^{(i)}] - \sigma_w^2\sum_{\ell=1}^k \Psi^\pi_{\mathrm{cl},k,\ell}\mathbb{I}_{\ell=j}]\| \\
&= \frac{1}{N\sigma_w^2}\|\sum_{i=1}\mathbb{E}[\mathtt{z}_k^{(i)}\mathtt{w}_j^{(i)}] - \sigma_w^2\Psi^\pi_{\mathrm{cl},k,j}] = \mathrm{Term}_3.
\end{aligned}$$

In sum, with probability at least $1 - 3\delta/4$, the following bound holds

$$\begin{aligned}
&\left\|\frac{\sigma_w^{-2}}{N}\sum_{i=1}^N \mathtt{z}_k^{(i)}\mathtt{w}_j^{(i)\top} - \Psi^\pi_{\mathrm{cl},k,j}\right\|_{\mathrm{op}} \\
&\le \mathrm{Term}_1 + \mathrm{Term}_2 + \mathrm{Term}_3 \\
&\le \frac{\sigma_{\mathrm{orac}}}{\sigma_w}\sqrt{2d_\star \cdot \frac{d_\star \log 5 + \log(\frac{12d_\star}{\delta})}{N}} + 8L_{\mathrm{tay},\infty,\pi} d_\star \sqrt{\frac{\log(6d_\star/\delta)}{N}} + 4\sigma_w M_{\mathrm{tay},2,\pi} d_\star^{3/2} \\
&\le \sqrt{\frac{\log\frac{12d_\star}{\delta}}{N}}\left(\frac{2\sigma_{\mathrm{orac}}}{\sigma_w} d_\star^{3/2} + 8L_{\mathrm{tay},\infty,\pi} d_\star\right) + 4\sigma_w M_{\mathrm{tay},2,\pi} d_\star^{3/2}.
\end{aligned}$$

The final bound follows from a union bound over all $\binom{K}{2} \le K^2$ pairs, and replacing δ with $\delta/2n_{\mathrm{iter}}$.

□

E.2. Error in the Gradient (Proof of Lemma A.10)

Recall the definitions

$$\begin{aligned}
(\nabla\mathcal{J}_T^{\mathrm{disc}}(\pi))_k = \tau Q_u(\mathtt{x}_k^\pi, \mathtt{u}_k^\pi, t_k) + (\Psi^\pi_{\mathrm{cl},K+1,k})^\top V_x(\mathtt{x}_{K+1}^\pi) + \\
+ \tau\sum_{j=k+1}^K (\Psi^\pi_{\mathrm{cl},j,k})^\top (Q_x(\mathtt{x}_j^\pi, \mathtt{u}_j^\pi, t_j) + (\mathtt{K}_j^\pi)^\top Q_u(\mathtt{x}_j^\pi, \mathtt{u}_j^\pi, t_j))
\end{aligned}$$

and

$$\begin{aligned}
\hat{\nabla}_k = Q_u(\bar{\mathtt{x}}_k^\pi, \mathtt{u}_k^\pi, t_j) + \hat{\Psi}_{K+1,k}^\top V_x(\bar{\mathtt{x}}_{K+1}) \\
+ \tau\sum_{j=k+1}^K \hat{\Psi}_{j,k}^\top \left(Q_x(\hat{\mathtt{x}}_j, \mathtt{u}_j^\pi, t_j) + (\mathtt{K}_j^\pi)^\top Q_u(\hat{\mathtt{x}}_j, \mathtt{u}_j^\pi, t_j)\right)
\end{aligned}$$

Using $V_x(\cdot)$ and $Q_x(\cdot), Q_u(\cdot)$ are all at most L_{cost} in magnitude, that the gradients of the cost are M_{cost}-Lipschitz, and $1 \vee \|\mathtt{K}_j^\pi\| \le L_\pi$ we can bound

$$
\begin{aligned}
&\|(\nabla \mathcal{J}_T^{\text{disc}}(\pi))_k - \hat{\nabla}_k\| \\
&\le L_{\text{cost}} \|\hat{\Psi}_{K+1,k} - \Psi^\pi_{\text{cl},K+1,k}\| + 2L_\pi L_{\text{cost}} \tau \sum_{j=k+1}^{K} \|\hat{\Psi}_{j,k} - \Psi^\pi_{\text{cl},j,k}\| \\
&\quad + M_{\text{cost}} \left(\|\mathtt{x}_k^\pi - \bar{\mathtt{x}}_k^\pi\| + \|\Psi^\pi_{\text{cl},K+1,k}\| \cdot \|\mathtt{x}^\pi_{K+1} - \hat{\mathtt{x}}_{K+1}\| + 2\tau L_\pi \sum_{j=k+1}^{K} \|\Psi^\pi_{\text{cl},j,k}\| \|\mathtt{x}_j^\pi - \hat{\mathtt{x}}_j\| \right) \\
&\le L_{\text{cost}} \text{Err}_{\Psi,\pi}(\delta)(1 + 2L_\pi \tau K) + M_{\text{cost}} \text{Err}_{\hat{x}}(\delta) \left(1 + \|\Psi^\pi_{\text{cl},K+1,k}\| + 2\tau L_\pi \sum_{j=k+1}^{K} \|\Psi^\pi_{\text{cl},j,k}\| \right) \\
&\le L_{\text{cost}} \text{Err}_{\Psi,\pi}(\delta)(1 + 2L_\pi \tau K) + M_{\text{cost}} \text{Err}_{\hat{x}}(\delta) \left(1 + \kappa_{\pi,\infty} + 2K\tau L_\pi \kappa_{\pi,\infty} \right) \\
&= L_{\text{cost}} \text{Err}_{\Psi,\pi}(\delta)(1 + 2TL_\pi) + M_{\text{cost}} \text{Err}_{\hat{x}}(\delta) \left(1 + (1 + 2TL_\pi)\kappa_{\pi,\infty} \right) \\
&\le \underbrace{(L_{\text{cost}} \text{Err}_{\Psi,\pi}(\delta) + (1 + \kappa_{\pi,\infty}) M_{\text{cost}} \text{Err}_{\hat{x}}(\delta))(1 + 2TL_\pi)}_{:= \text{Err}_{\nabla,\pi}(\delta)}
\end{aligned}
$$

□

E.3. Discrete-Time Closed-Loop Controllability (Proposition A.11)

We begin by lower bounding the continuous-time controllability Grammian under a policy π, and then turn to lower bounding its discretization. At the end of the proof, we remark upon how the bound can be refined. The first part of the argument follows (Chen & Hazan, 2021).

Equivalent characterization of controllability Gramian smallest singular value. The following is a continuous-time analogue of Chen & Hazan (2021, Lemma 15).

Lemma E.3 (Characterization of Controllability Gramian smallest singular value)**.** *Let $\mathbf{\Psi}(t,s) \in \mathbb{R}^{d_{\mathtt{x}} \times d_{\mathtt{u}}}$ be an arbitrary (locally square integrable), and set*

$$\Lambda := \int_{s=t-t_{\text{ctrl}}}^{t} \mathbf{\Psi}(t,s)\mathbf{\Psi}(t,s)^\top \mathrm{d}s.$$

Then, $\lambda_{\min}(\Lambda) \ge \nu$ if and only if for all unit vectors ξ, there exists some $\mathbf{u}_\xi(s)$ such that $\xi = \int_{s=t-t_{\text{ctrl}}}^{t} \mathbf{\Psi}(t,s)\mathbf{u}_\xi(s)$ and $\int_{s=t-t_{\text{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\|^2 \le \nu^{-1}$.

Proof of Lemma E.3. Fix any unit vector $\xi \in \mathbb{R}^n$, define. First, suppose $\int_{s=t-t_{\text{ctrl}}}^{t} \mathbf{\Psi}(t,s)\mathbf{\Psi}(t,s)^\top \mathrm{d}s \ge \nu$.

$$\mathbf{u}_\xi(s) := \mathbf{\Psi}(t,s)^\top \Lambda^{-1} \xi.$$

One can verify then that

$$\int_{s=t-t_{\text{ctrl}}}^{t} \mathbf{\Psi}(t,s)\mathbf{u}_\xi(s)\mathrm{d}s = \Lambda\Lambda^{-1}\xi = \xi \qquad \int_{s=t-t_{\text{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\|^2 \mathrm{d}s = \xi\Lambda^{-1} \cdot \Lambda \cdot \Lambda^{-1}\xi = \xi\Lambda^{-1}\xi \le \lambda_{\min}(\Lambda)^{-1}.$$

On the other hand, suppose that there exists a control $\mathbf{u}_\xi(s)$ with $\int_{s=t-t_{\text{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\|^2 \le \lambda_{\min}(\Lambda)^{-1}$ such that

$\int_{s=t-t_{\mathrm{ctrl}}}^{t} \mathbf{\Psi}(t,s)\mathbf{u}_\xi(s)\mathrm{d}s = \xi$. As ξ is a unit vector, i.e. $\xi^\top\xi = 1$,

$$\begin{aligned}
1 &= \left(\xi^\top \int_{s=t-t_{\mathrm{ctrl}}}^{t} \mathbf{\Psi}(t,s)\mathbf{u}_\xi(s)\mathrm{d}s\right)^2 \\
&= \left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \xi^\top\mathbf{\Psi}(t,s)\mathbf{u}_\xi(s)\mathrm{d}s\right) \\
&\le \left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\xi^\top\mathbf{\Psi}(t,s)\|^2\mathrm{d}s\right)\cdot\left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\|^2\mathrm{d}s\right) \\
&\le \xi^\top\Lambda\xi\cdot\left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\|^2\mathrm{d}s\right) \\
&\le \xi^\top\Lambda\xi\cdot\lambda_{\min}(\Lambda)^{-1}.
\end{aligned}$$

The bound follows. □

Lower bounding the controllability Gramian until algernative policies. This next part is the continuous-time analogue of (**?**)Lemma 16]chen2021black, establishing controllability of the closed-loop linearized system in feedback with policy π.

Lemma E.4 (Controllabiity of Closed-Loop Transitions, Continuous-Time)**.** *Recall $L_\pi \ge 1$, and $\gamma_{\mathrm{ctr}} := \max\{1, L_f t_{\mathrm{ctrl}}\}$. Then, under Assumption 4.4,*

$$\int_{s=t-t_{\mathrm{ctrl}}}^{t} \mathbf{\Psi}^\pi_{\mathrm{cl}}(t,s)\mathbf{\Psi}^\pi_{\mathrm{cl}}(t,s)^\top\mathrm{d}s \succeq \frac{\nu_{\mathrm{ctrl}}}{4L_\pi^2\gamma_{\mathrm{ctr}}^2\exp(2\gamma_{\mathrm{ctr}})}.$$

Proof of Lemma E.4. Fix any $\xi \in \mathbb{R}^{d_{\mathsf{x}}}$ of unit norm. Lemma E.3 and Assumption 4.4 guarantee the existence of an input $\mathbf{u}_\xi(s)$ for which

$$\int_{s=t-t_{\mathrm{ctrl}}}^{t} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t,s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathbf{u}_\xi(s)\mathrm{d}s = \xi, \quad \int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\|^2\mathrm{d}s \le \nu_{\mathrm{ctrl}}^{-1}.$$

Let

$$\mathbf{z}_\xi(s') = \int_{s=t-t_{\mathrm{ctrl}}}^{s'} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t,s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathbf{u}_\xi(s)\mathrm{d}s.$$

Define now the input

$$\tilde{\mathbf{u}}_\xi(s) := \mathbf{u}_\xi(s) - \mathbb{I}\{t_k(s) > t - t_{\mathrm{ctrl}}\}\mathtt{K}^\pi_{k(s)}\mathbf{z}_\xi(t_k(s)).$$

It can be directly verified (by induction on k) that

$$\forall s' \in [t-t_{\mathrm{ctrl}}, t], \quad \int_{s=t-t_{\mathrm{ctrl}}}^{s'} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t,s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathbf{u}_\xi(s)\mathrm{d}s = \int_{s=t-t_{\mathrm{ctrl}}}^{s'} \mathbf{\Psi}^\pi_{\mathrm{cl}}(t,s)\tilde{\mathbf{u}}_\xi(s)\mathrm{d}s,$$

so in particular

$$\xi = \int_{s=t-t_{\mathrm{ctrl}}}^{t} \mathbf{\Psi}^\pi_{\mathrm{cl}}(t,s)\tilde{\mathbf{u}}_\xi(s)\mathrm{d}s.$$

We may now bound

$$\begin{aligned}
\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\tilde{\mathbf{u}}_\xi(s)\|^2 &= \int_{s=t-t_{\mathrm{ctrl}}}^{t} \left(\|\mathbf{u}_\xi(s) - \mathbb{I}\{t_k(s) > t - t_{\mathrm{ctrl}}\}\mathtt{K}^\pi_{k(s)}\mathbf{z}_\xi(t_k(s))\|^2\right)\mathrm{d}s \\
&\le 2\int_{s=t-t_{\mathrm{ctrl}}}^{t} \left(\|\mathbf{u}_\xi(s)\|^2 + \|\mathtt{K}^\pi_{k(s)}\|\|\mathbf{z}_\xi(t_k(s))\|^2\right)\mathrm{d}s \\
&\le 2(\nu_{\mathrm{ctrl}}^{-1} + L_\pi^2\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{z}_\xi(t_k(s))\|^2\mathrm{d}s), \qquad \text{(E.4)}
\end{aligned}$$

We now adopt the following claim, mirroring the proof of Chen & Hazan (2021, Lemma 16).

Claim E.2. *The following bound holds:*

$$\forall s' \in [t - t_{\mathrm{ctrl}}, t], \quad \|\mathbf{z}_\xi(s')\|^2 \le t_{\mathrm{ctrl}} \nu_{\mathrm{ctrl}}^{-1} L_f^2 \exp(2 t_{\mathrm{ctrl}} L_f).$$

Proof of Claim E.2. We bound

$$\begin{aligned}
\|\mathbf{z}_\xi(s')\|^2 &= \left\| \int_{s=t-t_{\mathrm{ctrl}}}^{s'} \mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s) \mathbf{B}_{\mathrm{ol}}^\pi(s) \mathbf{u}_\xi(s) \mathrm{d}s \right\|^2 \\
&\le \left(\int_{s=t-t_{\mathrm{ctrl}}}^{s'} \|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\| \|\mathbf{B}_{\mathrm{ol}}^\pi(s)\| \|\mathbf{u}_\xi(s)\| \mathrm{d}s \right)^2 \\
&\le L_f^2 \left(\int_{s=t-t_{\mathrm{ctrl}}}^{s'} \|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\| \|\mathbf{u}_\xi(s)\| \mathrm{d}s \right)^2 && \text{(Assumption 4.1)} \\
&\le L_f^2 \left(\int_{s=t-t_{\mathrm{ctrl}}}^{s'} \|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\|^2 \mathrm{d}s \right)^2 \left(\int_{s=t-t_{\mathrm{ctrl}}}^{s'} \|\mathbf{u}_\xi(s)\| \mathrm{d}s \right)^2 && \text{(Cauchy-Schwartz)} \\
&\le L_f^2 \left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\|^2 \mathrm{d}s \right) \left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{u}_\xi(s)\| \mathrm{d}s \right) \\
&\le \nu_{\mathrm{ctrl}}^{-1} L_f^2 \left(\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\|^2 \mathrm{d}s \right).
\end{aligned}$$

By Lemma I.4, we can bound can bound $\|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\| \le \exp(L_f(t-s)) \le \exp(L_f t_{\mathrm{ctrl}})$ for $s \in [t - t_{\mathrm{ctrl}}, t]$, yielding $\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\mathbf{\Phi}_{\mathrm{ol}}^\pi(t,s)\|^2 \le t_{\mathrm{ctrl}} \exp(2 L_f t_{\mathrm{ctrl}})$. The bound claim. □

Combining Eq. (E.4) and Claim E.2,

$$\begin{aligned}
\int_{s=t-t_{\mathrm{ctrl}}}^{t} \|\tilde{\mathbf{u}}_\xi(s)\|^2 &\le 2\nu_{\mathrm{ctrl}}^{-1} \left(1 + t_{\mathrm{ctrl}}^2 L_f^2 L_\pi^2 \exp(2 L_f t_{\mathrm{ctrl}})\right) \\
&\le 2\nu_{\mathrm{ctrl}}^{-1} L_\pi^2 \left(1 + t_{\mathrm{ctrl}}^2 L_f^2 \exp(2 L_f t_{\mathrm{ctrl}})\right) && (L_\pi \ge 1) \\
&\le 2\nu_{\mathrm{ctrl}}^{-1} L_\pi^2 \left(1 + \gamma_{\mathrm{ctr}}^2 \exp(2\gamma_{\mathrm{ctr}})\right) && (\gamma_{\mathrm{ctr}} = \max\{1, t_{\mathrm{ctrl}} L_f\}) \\
&\le 4\nu_{\mathrm{ctrl}}^{-1} L_\pi^2 \gamma_{\mathrm{ctr}}^2 \exp(2\gamma_{\mathrm{ctr}}), && (\gamma_{\mathrm{ctr}} \ge 1)
\end{aligned}$$

which concludes the proof. □

Discretizing the Closed-Loop Gramian. To conclude the argument, we relate the controllability of the closed-loop Gramian in continuous-time to that in discrete-time.

Lemma E.5 (Discretization of Controllability Grammian)**.** *Suppose Assumption 4.4 holds and* $\tau \le L_f/4$*, then following holds:*

$$\begin{aligned}
&\left\| \int_{s=t_k-t_{\mathrm{ctrl}}}^{t_k} \mathbf{\Psi}_{\mathrm{cl}}^\pi(t_k,s) \mathbf{\Psi}_{\mathrm{cl}}^\pi(t_k,s)^\top \mathrm{d}s - \frac{1}{\tau} \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \Psi_{\mathrm{cl},k,j}^\pi (\Psi_{\mathrm{cl},k,j}^\pi)^\top \right\|_{\mathrm{op}} \\
&\quad \le 4\tau\gamma_{\mathrm{ctr}} \kappa_{\pi,\infty}^2 \left(\kappa_f M_f + 2L_f^2\right)
\end{aligned}$$

Proof. Recall the shorthand $L_{\mathrm{ol}} := \exp(\tau L_f)$, used in the discretization arguments in Appendix I. We can write

$$
\begin{aligned}
&\left\| \int_{s=t_k-t_{\mathrm{ctrl}}}^{t_k} \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s)\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s)^{\top}\mathrm{d}s - \tau^{-1}\sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \Psi^{\pi}_{\mathrm{cl},k,j}(\Psi^{\pi}_{\mathrm{cl},k,j})^{\top} \right\| \\
&= \left\| \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \int_{s=t_j}^{t_j+1} \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t,s)\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t,s)^{\top}\mathrm{d}s - \frac{1}{\tau}\Psi^{\pi}_{\mathrm{cl},k,j}(\frac{1}{\tau}\Psi^{\pi}_{\mathrm{cl},k,j})^{\top} \right\| \\
&\le \tau \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \max_{s\in\mathcal{I}_j} \left\| \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s)\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s)^{\top} - \frac{1}{\tau}\Psi^{\pi}_{\mathrm{cl},k,j}(\frac{1}{\tau}\Psi^{\pi}_{\mathrm{cl},k,j})^{\top} \right\| \\
&\le 2\tau \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \max_{s\in\mathcal{I}_j} \left\| \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s) - \frac{1}{\tau}\Psi^{\pi}_{\mathrm{cl},k,j} \right\| \cdot \max\left\{ \|\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s)\|, \frac{1}{\tau}\|\Psi^{\pi}_{\mathrm{cl},k,j}\| \right\} \\
&\le 2L_f L_{\mathrm{ol}}\tau\kappa_{\pi,\infty} \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \max_{s\in\mathcal{I}_j} \left\| \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_k,s) - \frac{1}{\tau}\Psi^{\pi}_{\mathrm{cl},k,j} \right\| && \text{(Lemma I.8(d))} \\
&\le 2L_f L_{\mathrm{ol}}\kappa^2_{\pi,\infty}\left(\kappa_f M_f + 2L_f^2\right) \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \tau^2 && \text{(Lemma I.8(b))} \\
&\le 2L_f L^2_{\mathrm{ol}}\kappa^2_{\pi,\infty}\left(\kappa_f M_f + 2L_f^2\right) \sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \tau^2 \\
&\le 2L_f L^2_{\mathrm{ol}}\kappa^2_{\pi,\infty}\left(\kappa_f M_f + 2L_f^2\right) k_{\mathrm{ctrl}}\tau^2 \\
&= 2\tau t_{\mathrm{ctrl}} L_f L^2_{\mathrm{ol}}\kappa^2_{\pi,\infty}\left(\kappa_f M_f + 2L_f^2\right).
\end{aligned}
$$

As $\tau \le L_f/4$, $L^2_{\mathrm{ol}} \le \exp(1/2) \le 2$, so that the above is at most $4\tau t_{\mathrm{ctrl}} L_f \kappa^2_{\pi,\infty}\left(\kappa_f M_f + 2L_f^2\right)$. Recalling $\gamma_{\mathrm{ctr}} := \max\{1, t_{\mathrm{ctrl}} L_f\}$ concludes. □

Concluding the proof.

Proof of Proposition A.11. The proof follows by combining the bounds in Lemmas E.4 and E.5. These yield

$$
\frac{1}{\tau}\lambda_{\min}\left(\sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \Psi^{\pi}_{\mathrm{cl},k,j}(\Psi^{\pi}_{\mathrm{cl},k,j})^{\top}\right) \succeq \frac{\nu_{\mathrm{ctrl}}}{4L_\pi^2\gamma_{\mathrm{ctr}}^2\exp(2\gamma_{\mathrm{ctr}})} - \kappa^2_{\pi,\infty}\cdot 4\tau\gamma_{\mathrm{ctr}}\left(\kappa_f M_f + 2L_f^2\right)
$$

Recall $\gamma_{\mathrm{ctr}} = t_{\mathrm{ctrl}} L_f$. Hence, if

$$
\tau \le \frac{\nu_{\mathrm{ctrl}}}{8L_\pi^2\kappa^2_{\pi,\infty}\gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}})\left(\kappa_f M_f + 2L_f^2\right)},
$$

it holds that

$$
\lambda_{\min}\left(\sum_{j=k-k_{\mathrm{ctrl}}}^{k-1} \Psi^{\pi}_{\mathrm{cl},k,j}(\Psi^{\pi}_{\mathrm{cl},k,j})^{\top}\right) \succeq \frac{\nu_{\mathrm{ctrl}}}{8L_\pi^2\gamma_{\mathrm{ctr}}^2\exp(2\gamma_{\mathrm{ctr}})}.
$$

□

E.4. Recovery of State-Transition Matrix (Proposition A.12)

The analysis is based on the Ho-Kalman scheme. We begin with the observation that

$$
\Psi^{\pi}_{\mathrm{cl},k,j} = \mathtt{A}^{\pi}_{\mathrm{cl},k}\Psi^{\pi}_{\mathrm{cl},k-1,j}, \quad \forall j < k.
$$

To this end, define the matrices

$$\mathcal{C}_{k|j_2,j_1} := [\Psi^{\pi}_{\mathrm{cl},k+1,j_2} \mid \Psi^{\pi}_{\mathrm{cl},k+1,j_2-1} \mid \dots \Psi^{\pi}_{\mathrm{cl},k+1,j_1}],$$

Then, we have the identity

$$\mathcal{C}_{k|k-1,j} = \mathtt{A}^{\pi}_{\mathrm{cl},k}\mathcal{C}_{k-1|k-1,j},$$

so that if $\mathrm{rank}(\mathcal{C}_{k-1|k-1,j}) = d_{\mathtt{x}}$, we have $\mathtt{A}^{\pi}_{\mathrm{cl},k} = \mathcal{C}_{k|k-1,j}\mathcal{C}^{\dagger}_{k-1|k-1,j}$, where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudoinverse. We now state and prove a more-or-less standard perturbation bound.

Lemma E.6. *Suppose* $\mathrm{rank}(\mathcal{C}_{k-1|k-1,j}) = d_{\mathtt{x}}$*, and consider any estimates* $\hat{\mathcal{C}}_{k|k-1,j}, \hat{\mathcal{C}}_{k-1|k-1,j}$ *of* $\mathcal{C}_{k|k-1,j}, \mathcal{C}_{k-1|k-1,j}$*. Define*

$$\begin{aligned}\Delta &:= \max\{\|\mathcal{C}_{k|k-1,j} - \hat{\mathcal{C}}_{k|k-1,j}\|, \|\mathcal{C}_{k-1|k-1,j} - \hat{\mathcal{C}}_{k-1|k-1,j}\|\}\\ M &:= \max\{\|\mathcal{C}_{k|k-1,j}\|, \|\mathcal{C}_{k-1|k-1,j}\|\}.\end{aligned}$$

Then, if $\Delta \le \sigma_{\min}(\mathcal{C}_{k-1|k-1,j})/2$*, the estimate* $\tilde{\mathtt{A}}_k := \hat{\mathcal{C}}_{k|k-1,j}\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}$ *satisfies*

$$\|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| \le 6\Delta M\sigma_{\min}(\mathcal{C}_{k-1|k-1,j})^{-2}.$$

Proof of Lemma E.6. Then, we have (provided $\mathrm{rank}(\hat{\mathcal{C}}_{k-1|k-1,j}) = \mathrm{rank}(\mathcal{C}_{k-1|k-1,j}) = d_{\mathtt{x}}$), we have

$$\begin{aligned}\|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| &= \|\mathcal{C}_{k|k-1,j}\mathcal{C}^{\dagger}_{k-1|k-1,j} - \hat{\mathcal{C}}_{k|k-1,j}\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}\|\\
&\le \|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}\|\|\mathcal{C}_{k|k-1,j} - \hat{\mathcal{C}}_{k|k-1,j}\| + \|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j} - \mathcal{C}^{\dagger}_{k-1|k-1,j}\|\|\mathcal{C}_{k|k-1,j}\|\\
&\le \|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}\|\|\mathcal{C}_{k|k-1,j} - \hat{\mathcal{C}}_{k|k-1,j}\|\\
&\quad + \frac{1+\sqrt{5}}{2}\|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}\| \cdot \|\mathcal{C}^{\dagger}_{k-1|k-1,j}\| \cdot \|\mathcal{C}_{k-1|k-1,j} - \hat{\mathcal{C}}_{k-1|k-1,j}\| \cdot \|\hat{\mathcal{C}}_{k|k-1,j}\|\\
&\qquad\qquad \text{(cite (Stewart, 1977), and also (Xu, 2020))}\\
&\overset{(i)}{\le} \Delta\|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}(1 + \frac{1+\sqrt{5}}{2}\|\mathcal{C}^{\dagger}_{k-1|k-1,j}\|\|\mathcal{C}_{k|k-1,j}\|)\\
&\overset{(ii)}{\le} 3\Delta M\|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}\|\mathcal{C}^{\dagger}_{k-1|k-1,j}\|,\end{aligned}$$

where in (i) we use $\Delta := \max\{\|\mathcal{C}_{k|k-1,j} - \hat{\mathcal{C}}_{k|k-1,j}\|, \|\mathcal{C}_{k-1|k-1,j} - \hat{\mathcal{C}}_{k-1|k-1,j}\|\}$, and in (ii), we use $M = \max\{\|\mathcal{C}_{k|k-1,j}\|, \|\mathcal{C}_{k-1|k-1,j}\|\}$, which admits the simplification in (ii) because $\|\mathcal{C}_{k-1|k-1,j}\|\|\mathcal{C}^{\dagger}_{k-1|k-1,j}\| \ge 1$. In particular, if $\mathrm{rank}(\mathcal{C}_{k-1|k-1,j}) = d_{\mathtt{x}}$, and $\Delta \le \sigma_{\min}(\mathcal{C}_{k-1|k-1,j})/2$, then $\|\hat{\mathcal{C}}^{\dagger}_{k-1|k-1,j}\| \le 2/\sigma_{\min}(\mathcal{C}_{k-1|k-1,j})$, and we obtain

$$\|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| \le 6\Delta M\sigma_{\min}(\mathcal{C}_{k-1|k-1,j})^{-2}.$$

□

Next, restricting our attention to $k \ge k_{\mathrm{ctrl}} + 2$, we specialize the above analysis to

$$\begin{aligned}\mathcal{C}_{k,\mathrm{in}} &:= \mathcal{C}_{k-1|k-1,k-k_0+1}, \quad \mathcal{C}_{k,\mathrm{out}} := \mathcal{C}_{k|k-1,k-k_0+1}\\
\hat{\mathcal{C}}_{k,\mathrm{in}} &:= \hat{\mathcal{C}}_{k-1|k-1,k-k_0+1}, \quad \hat{\mathcal{C}}_{k,\mathrm{out}} := \hat{\mathcal{C}}_{k|k-1,k-k_0+1}\end{aligned}$$

where $\hat{\mathcal{C}}_{(\cdot)}$ arises from the plug-in estimates

$$\hat{\mathcal{C}}_{k|j_2,j_1} := [\hat{\Psi}_{k+1,j_2} \mid \hat{\Psi}_{k+1,j_2-1} \mid \dots \hat{\Psi}_{k+1,j_1}]$$

Define further

$$\tilde{\mathtt{A}}_k := \hat{\mathcal{C}}_{k,\mathrm{out}}\hat{\mathcal{C}}^{\dagger}_{k,\mathrm{in}}, \quad \text{so that } \hat{\mathtt{A}}_k = \tilde{\mathtt{A}}_k - \hat{\mathtt{B}}_k\mathtt{K}^{\pi}_k.$$

Recall $t_0 = k_0/\tau$. We can now bound, recalling $L_{\mathrm{ol}} := \exp(\tau L_f) \le 2$ for $\tau \le L_f/4$ and $\gamma_{\mathrm{ctr}} = \max\{1, t_{\mathrm{ctrl}} L_f\} = \max\{1, \tau k_{\mathrm{ctrl}} L_f\}$,

$$\begin{aligned}
\max\{\|\mathcal{C}_{k,\mathrm{in}}\|, \|\mathcal{C}_{k,\mathrm{out}}\|\} &\le \sqrt{k_0} \max_{j<k} \|\Psi^{\pi}_{\mathrm{cl},k,j}\| \\
&\le \sqrt{k_{\mathrm{ctrl}}}\tau L_f L_{\mathrm{ol}} \kappa_{\pi,\infty} && \text{(Lemma I.8(d))} \\
&\le \sqrt{\tau t_0}\tau L_f L_{\mathrm{ol}} \kappa_{\pi,\infty} && (t_0 = k_0\tau) \\
&\le 2\kappa_{\pi,\infty}\gamma_{\mathrm{ctr}}\sqrt{\tau t_0}. && (\gamma_{\mathrm{ctr}} \ge 1)
\end{aligned}$$

Invoking Proposition A.11, we also have that provided $\tau \le \min\{\tau_{\mathrm{dyn}}, \tau_{\mathrm{ctrl},\pi}\}$, since $k_0 \ge k_{\mathrm{ctrl}} + 2$,

$$\begin{aligned}
\sigma_{\min}(\mathcal{C}_{k-1|k-1,j})^2 &= \lambda_{\min}\left(\sum_{j=k-k_0+1}^{k-1} \Psi^{\pi}_{\mathrm{cl},k-1,j}(\Psi^{\pi}_{\mathrm{cl},k-1,j})^\top\right) \\
&\ge \lambda_{\min}\left(\sum_{j=k-k_{\mathrm{ctrl}}-1}^{k-1} \Psi^{\pi}_{\mathrm{cl},k-1,j}(\Psi^{\pi}_{\mathrm{cl},k-1,j})^\top\right) \succeq \tau \cdot \frac{\nu_{\mathrm{ctrl}}}{8L_\pi^2\gamma_{\mathrm{ctr}}^2\exp(2\gamma_{\mathrm{ctr}})}.
\end{aligned}$$

Therefore, as long as

$$\Delta := \max\{\|\mathcal{C}_{k,\mathrm{in}} - \hat{\mathcal{C}}_{k,\mathrm{in}}\|, \|\mathcal{C}_{k,\mathrm{out}} - \hat{\mathcal{C}}_{k,\mathrm{out}}\|\} \le \frac{\sqrt{\tau\nu_{\mathrm{ctrl}}}}{2\sqrt{2}L_\pi\gamma_{\mathrm{ctr}}\exp(\gamma_{\mathrm{ctr}})},$$

we have

$$\|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| \le \sqrt{\frac{t_0}{\tau}}\frac{96\Delta}{\nu_{\mathrm{ctrl}}} \cdot \kappa_{\pi,\infty}L_\pi^2\gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}}).$$

Lastly, we can upper bound $\Delta \le \sqrt{k_{\mathrm{ctrl}}}\mathrm{Err}_{\Psi,\pi}(\delta) = \sqrt{t_0/\tau}\mathrm{Err}_{\Psi,\pi}(\delta)$, from which we cconlude that as long as $\mathrm{Err}_{\Psi}(\delta) \le \tau\frac{\sqrt{\nu_{\mathrm{ctrl}}/t_0}}{2\sqrt{2}L_\pi\gamma_{\mathrm{ctr}}\exp(\gamma_{\mathrm{ctr}})}$, we have

$$\|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| \le t_0\kappa_{\pi,\infty}L_\pi^2 \cdot \frac{96\mathrm{Err}_{\Psi,\pi}(\delta)}{\tau\nu_{\mathrm{ctrl}}} \cdot \gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}}).$$

Now to wrap up. We observe that $\mathtt{B}^{\pi}_{\mathrm{ol},k} = \Psi^{\pi}_{\mathrm{cl},k+1,k}$, so

$$\|\hat{\mathtt{B}}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\| = \|\Psi^{\pi}_{\mathrm{cl},k+1,k} - \hat{\Psi}_{k+1,k}\| \le \mathrm{Err}_{\Psi,\pi}(\delta).$$

Therefore,

$$\begin{aligned}
\|\hat{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{ol},k}\| &= \|\tilde{\mathtt{A}}_k - (\mathtt{A}^{\pi}_{\mathrm{cl},k} - \mathtt{B}^{\pi}_{\mathrm{ol},k}\mathtt{K}^{\pi}_k)\| \\
&\le \|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| + \|\hat{\mathtt{B}}_k\mathtt{K}^{\pi}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\mathtt{K}^{\pi}_k\| \\
&\le \|\tilde{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{cl},k}\| + \|\hat{\mathtt{B}}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\|L_\pi \\
&\le L_\pi\mathrm{Err}_{\Psi,\pi}(\delta) + t_0\kappa_{\pi,\infty}L_\pi^2 \cdot \frac{96\mathrm{Err}_{\Psi,\pi}(\delta)}{\tau\nu_{\mathrm{ctrl}}} \cdot \gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}}).
\end{aligned}$$

Lastly, we notice this upper bound on $\|\hat{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{ol},k}\|$ is larger than that on $\|\hat{\mathtt{B}}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\|$, as $L_\pi \ge 1$ by assumption, and that for $\tau \le \tau_{\mathrm{ctrl},\pi}$, $L_\pi\mathrm{Err}_{\Psi,\pi}(\delta) \le t_0\kappa_{\pi,\infty}L_\pi^2 \cdot \frac{96\mathrm{Err}_{\Psi,\pi}(\delta)}{\tau\nu_{\mathrm{ctrl}}} \cdot \gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}})$. Thus,

$$\|\hat{\mathtt{A}}_k - \mathtt{A}^{\pi}_{\mathrm{ol},k}\| \vee \|\hat{\mathtt{B}}_k - \mathtt{B}^{\pi}_{\mathrm{ol},k}\| \le \frac{\mathrm{Err}_{\Psi,\pi}(\delta)}{\tau} \cdot t_0\kappa_{\pi,\infty}L_\pi^2\gamma_{\mathrm{ctr}}^3\exp(2\gamma_{\mathrm{ctr}}) \cdot \frac{192}{\nu_{\mathrm{ctrl}}}.$$

□

F. Certainty Equivalence

In this section, we establish a general certainty equivalence bound for linear time-varying discrete-time systems; we apply this in the proof Proposition A.14 in Appendix G.1.

Let $\mathbf{\Theta}^\star := (\mathtt{A}^\star_{1:K}, \mathtt{B}^\star_{1:K})$ denote ground-truth system parameters, and let $\hat{\mathbf{\Theta}} := (\hat{\mathtt{A}}_{1:K}, \hat{\mathtt{B}}_{1:K})$ denote estimates. We work with a slightly different discretization parameterization, where the dynamics are given by $\mathtt{x}_{h+1} = \mathtt{A}_k\mathtt{x}_h + \tau\mathtt{B}_k\mathtt{u}_h$. This parametrization ensures that the norms of $\mathtt{B}_k$ scale like constants independent of τ when instantiated with $\mathtt{A}_k \leftarrow \mathtt{A}^\pi_{\mathrm{ol},k}$ and $\mathtt{B}_k \leftarrow \frac{1}{\tau}\mathtt{B}^\pi_{\mathrm{ol},k}$.

Definition F.1. Given cost matrices $\mathbf{Q}, \mathbf{R}$, step τ, and parameters $\mathbf{\Theta} = (\mathtt{A}_{1:K}, \mathtt{B}_{1:K})$, we define $\mathtt{P}^{\mathrm{opt}}_k(\mathbf{\Theta})$ as the solution to the following program

$$\begin{aligned} x^\top \mathtt{P}^{\mathrm{opt}}_k(\mathbf{\Theta})x = \min_{\mathbf{u}_{k:H}} \mathtt{x}_{K+1}^\top \mathbf{Q}\mathtt{x}_{K+1} + \tau\sum_{h=k}^{K}(\mathtt{x}_h^\top\mathbf{Q}\mathtt{x}_h + \mathtt{u}_h^\top\mathbf{Q}\mathtt{u}_h) \\ \text{s.t. } \mathtt{x}_{h+1} = \mathtt{A}_h\mathtt{x}_h + \tau\mathtt{B}_h\mathtt{u}_h, \quad \mathtt{x}_k = x. \end{aligned} \tag{F.1}$$

The closed form for $\mathtt{P}^{\mathrm{opt}}_k$ is given by the follow standard computation, modified with the reparametrized dynamics $\mathtt{x}_{h+1} = \mathtt{A}_h\mathtt{x}_h + \tau\mathtt{B}_h\mathtt{u}_h$)

Lemma F.1. *The optimal Riccati cost-to-go $\mathtt{P}^{\mathrm{opt}}_{1:K+1} = \mathtt{P}^{\mathrm{opt}}_k(\mathbf{\Theta})$ is given by the solution to the following recursion with final condition $\mathtt{P}^{\mathrm{opt}}_{K+1} = \mathbf{Q}$ and*

$$\mathtt{P}^{\mathrm{opt}}_k = \mathtt{A}_k^\top\mathtt{P}^{\mathrm{opt}}_{k+1}\mathtt{A}_k - \tau\left(\mathtt{B}_k\mathtt{P}^{\mathrm{opt}}_{k+1}\mathtt{A}_k\right)^\top(\mathbf{R} + \tau\mathtt{B}_k^\top\mathtt{P}^{\mathrm{opt}}_{k+1}\mathtt{B}_k)^{-1}\left(\mathtt{B}_k\mathtt{P}^{\mathrm{opt}}_{k+1}\mathtt{A}_k\right) + \tau\mathbf{Q}$$

Moreover, defining $\mathtt{K}^{\mathrm{opt}}_k = \mathtt{K}^{\mathrm{opt}}_k(\Theta) := -(\mathbf{R} + \tau\mathtt{B}_k^\top\mathtt{P}^{\mathrm{opt}}_k\mathtt{B}_k)^{-1}\mathtt{B}_k^\top\mathtt{P}^{\mathrm{opt}}_k\mathtt{A}_k$, the optimal control for Eq. (F.1) is given by $\mathtt{x}_k = \mathtt{K}^{\mathrm{opt}}_k\mathtt{u}_k$.

Proof. This follows by reparametrizing the standard discrete-time Ricatti update (see e.g. Anderson & Moore (2007, Section 2.4)), with $\mathtt{B}_k \leftarrow \tau\mathtt{B}_k$, $\mathbf{Q} \leftarrow \tau\mathbf{Q}$, and $\mathbf{R} \leftarrow \tau\mathbf{R}$, and simplifying dependence on τ. □

The following identity is also standard (again, consult Anderson & Moore (2007, Section 2.4), albeit again with the reparamerizations $\mathtt{B}_k \leftarrow \tau\mathtt{B}_k$, $\mathbf{Q} \leftarrow \tau\mathbf{Q}$,and $\mathbf{R} \leftarrow \tau\mathbf{R}$):

$$\mathtt{P}^{\mathrm{opt}}_k = (\mathtt{A}_k + \tau\mathtt{B}_k\mathtt{K}^{\mathrm{opt}}_k)^\top\mathtt{P}^{\mathrm{opt}}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathtt{K}^{\mathrm{opt}}_k) + \tau(\mathbf{Q} + (\mathtt{K}^{\mathrm{opt}}_k)^\top\mathbf{R}(\mathtt{K}^{\mathrm{opt}}_k)) \tag{F.2}$$

Next, we define the cost-to-go functions associated for arbitrary sequences of feedback matrices, and from the optimal feedback matrices from another instance $\mathbf{\Theta}'$.

Definition F.2 (Feedback and Certainty Equivalent Cost-to-go)**.** Given a sequence of feedback gains $\mathtt{K}_{1:K}$, we define the induced cost-to-go

$$\begin{aligned} \mathtt{P}^{\mathrm{feed}}_k(\mathbf{\Theta}; \mathtt{K}_{1:K}) := \mathtt{x}_{K+1}^\top\mathbf{Q}\mathtt{x}_{K+1} + \tau\sum_{h=k}^{K}(\mathtt{x}_h^\top\mathbf{Q}\mathtt{x}_h + \mathtt{u}_h^\top\mathbf{Q}\mathtt{u}_h) \\ \text{s.t. } \mathtt{x}_{h+1} = (\mathtt{A}_h + \tau\mathtt{B}_h\mathtt{K}_k)\mathtt{x}_h \quad \mathtt{x}_k = x. \end{aligned}$$

And define the *certainty equivalent* cost-to-go as $\mathtt{P}^{\mathrm{ce}}_k(\mathbf{\Theta}; \mathbf{\Theta}') = \mathtt{P}^{\mathrm{feed}}_k(\mathbf{\Theta}; \mathtt{K}^{\mathrm{opt}}_{1:K}(\mathbf{\Theta}'))$ as the feedback cost-to-go for $\mathbf{\Theta}$ using the optimal gains for $\mathbf{\Theta}'$.

In particular, $\mathtt{P}^{\mathrm{ce}}_k(\mathbf{\Theta}; \mathbf{\Theta}) = \mathtt{P}^{\mathrm{opt}}_k(\mathbf{\Theta})$. We now present upper bounds on $\mathtt{P}^{\mathrm{ce}}_k(\mathbf{\Theta}; \mathbf{\Theta}')$. We assume bounds on the various parameters of interest.

Condition F.1. We have that there are constants $K_B, K_A \geq 1$ such that, for all $k \in [K]$,

$$\|\mathtt{B}^\star_k\| \vee \|\hat{\mathtt{B}}_k\| \leq K_B \quad \|\mathtt{A}^\star_k\| \vee \|\hat{\mathtt{A}}_k\| \leq K_A,$$

Condition F.2. We assume that there exists $\Delta_A, \Delta_B > 0$,

$$\forall k, \quad \|\hat{\mathtt{B}}_k - \mathtt{B}_k^\star\| \le \Delta_B \text{ and } \|\tau^{-1}(\hat{\mathtt{A}}_k - \mathtt{A}_k^\star)\| \le \Delta_A.$$

Condition F.3. We assume the a normalization on the cost matrices satisfy $\mathbf{R} \succeq \mathbf{I}$, $\mathbf{Q} \succeq \mathbf{I}$ and $\|\mathbf{Q}\| \ge \|\mathbf{R}\|$. As a special case, $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = \mathbf{I}$ suffices.

Lastly, the following assumption is needed to derive an upper bound on the closed-loop transition operator.

Condition F.4. We assume that $\max_k \|\mathtt{A}_k - \mathbf{I}\| \le \tau\kappa_A$.

Theorem 4 (Main Perturbation). *Suppose Conditions F.1 to F.3 hold. Define the terms*

$$\Delta_{\mathrm{ce}} := 80C^4K_A^3K_B^3(1+\tau C K_B)(\Delta_A + \Delta_B), \quad C := \max_{k\in[K+1]} \|\mathtt{P}_k^{\mathrm{opt}}(\mathbf{\Theta})\|.$$

Then, as long as $\Delta_{\mathrm{ce}} < 1$, we have

(a)

$$\max_{k\in[K+1]} \|\mathtt{P}_k^{\mathrm{ce}}(\mathbf{\Theta}; \hat{\mathbf{\Theta}})\| \le (1-\Delta_{\mathrm{ce}})^{-1} \max_{k\in[K+1]} \|\mathtt{P}_k^{\mathrm{opt}}(\mathbf{\Theta})\|$$

(b) $\max_{k\in[K+1]} \|\mathtt{K}_k^{\mathrm{opt}}(\hat{\mathbf{\Theta}})\| \le \frac{5}{4}K_BK_AC$.

(c) Moreover, if Condition F.4 holds, then the transition operators defined as

$$\mathtt{\Phi}_{j,k}^{\mathrm{ce}} := (\mathtt{A}_{j-1} + \tau\mathtt{B}_{j-1}\mathtt{K}_{j-1}^{\mathrm{opt}}(\hat{\mathbf{\Theta}})) \cdot (\mathtt{A}_{j-2} + \tau\mathtt{B}_{j-2}\mathtt{K}_{j-2}^{\mathrm{opt}}(\hat{\mathbf{\Theta}})) \cdot \cdots \cdot (\mathtt{A}_k + \tau\mathtt{B}_k\mathtt{K}_k^{\mathrm{opt}}(\hat{\mathbf{\Theta}})),$$

with the convention $\mathtt{\Phi}_{k,k+}^{\mathrm{ce}} = \mathbf{I}$ satisfy, for all $1 \le j \le k \le K$,

$$\|\mathtt{\Phi}_{j,k}^{\mathrm{ce}}\|^2 \le 2\kappa(1-\tau\gamma)^{j-k}, \quad \textit{where } \kappa = \kappa_A + \frac{5}{4}K_B^2K_AC, \quad \gamma = \frac{1-\Delta_{\mathrm{ce}}}{C},$$

provided $\kappa \le 1/2\tau$.

The proof of the Theorem 4 is outlined in Appendix F.1, and the supporting lemmas are proved in the subsequent sections. We now use this guarantee to establish Proposition A.14.

F.1. Proof Overview of Theorem 4

Step 0: Notation & Interpolating segments. To simplify notation, introduce the maximal operator norms, such that for an H-tuple of matrices $\mathtt{X}_{1:H} = (\mathtt{X}_1, \ldots, \mathbf{X}[H])$,

$$\|\mathtt{X}_{1:H}\|_{\max,\mathrm{op}} := \max_{k\in[H]} \|\mathtt{X}_k\|_{\mathrm{op}}.$$

Let us a consider the line segment joining these the parameters

$$\mathbf{\Theta}(s) = (\mathtt{A}_{1:K}(s), \mathtt{B}_{1:K}(s)) = (1-s)\mathbf{\Theta}^\star + s\hat{\mathbf{\Theta}} \tag{F.3}$$

For fixed cost matrices $\mathbf{Q}$ and $\mathbf{R}$, let $\mathtt{P}_{1:K+1}(s)$ and $\mathtt{P}_{1:K}(s)$ denote the solution to the finite-time Riccati recursion with parameters $\mathbf{\Theta}(s)$, where here $\mathbf{Q}$ also serves as a terminal cost at step $K+1$. We let $\mathtt{P}_{1:K+1}^\star, \mathtt{K}_{1:K}^\star$ be the solution for the truth $\mathbf{\Theta}^\star$ and $\hat{\mathbf{P}}_{1:H}, \hat{\mathbf{K}}_{1:H}$ the solution to the Riccati equation with $\hat{\mathbf{\Theta}}$; i.e. the certainty equivalent solution. By construction,

$$(\mathtt{P}_{1:K+1}(0), \mathtt{K}_{1:K}(0)) = (\mathtt{P}_{1:K+1}^\star, \mathtt{K}_{1:K}^\star), \quad (\mathtt{P}_{1:K+1}(1), \mathtt{K}_{1:K}(1)) = (\hat{\mathtt{P}}_{1:K+1}, \hat{\mathtt{K}}_{1:K+1})$$

For all quantity $\mathbf{X}(s)$ paramterized by $s \in [0,1]$, adopt the shorthand $\mathbf{X}'(s) := \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{X}(s)$.

Step 1. Self-Bounding ODE Method. We use an interpolation argument to study the certainty equivalence controller. Our main tool is the following interpolation bound, which states that if the norm of the s-derivative of a quantity is bounded by the norm of the quantity its self, then that quantity is uniformly bounded on a small enough range.

Lemma F.2 (Self-Bounding ODE Method, variant of Corollary 3 in (Simchowitz & Foster, 2020))**.** *Fix dimensions $d_1, d_2 \geq 1$, let $\mathcal{V} \subset \mathbb{R}^{d_1}$, let $f : \mathcal{V} \to \mathbb{R}^{d_2}$ be a $\mathcal{C}^1$ map and let $\mathbf{v}(s) : [0,1] \to \mathcal{V}$ be a $\mathcal{C}^1$ curve defined on $[0,1]$. Finally, let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^{d_2}$ and suppose that $c > 0$ and $p \geq 1$ satisfy*

$$\|\frac{\mathrm{d}}{\mathrm{d}s} f(\mathbf{v}(s))\| \leq c \max\{\|f(\mathbf{v}(s))\|, \|f(\mathbf{v}(0))\|\}^p \quad \forall s \in [0,1]. \tag{F.4}$$

Then, if $p > 1$ and if $\alpha = c(p-1)\|f(\mathbf{v}(0))\|^{p-1}$ satisfies $\alpha < 1$, the following bound holds for all $s \in [0,1]$:

$$\|f(\mathbf{v}(s)\| \leq (1-\alpha)^{-\frac{1}{p-1}} \|f(\mathbf{v}(0))\|, \quad \|\frac{\mathrm{d}}{\mathrm{d}s} f(\mathbf{v}(s)\| \leq c(1-\alpha)^{-\frac{p}{p-1}} \|f(\mathbf{v}(0))\|^p$$

Step 2. Perturbation of $\mathtt{P}_{1:K+1}(t)$ First, we show that the Riccati-updates obey the structure of Lemma F.2.

Lemma F.3. *Suppose (for simplicity) that $\lambda_{\min}(\mathbf{Q}), \lambda_{\min}(\mathbf{R}) \geq 1$. Then, for all $s \in [0,1]$*

$$\|\mathtt{P}'_{1:K+1}(s)\|_{\mathrm{max,op}} \leq 2(\Delta_A + K_A K_B \Delta_B)\|\mathtt{P}_{1:K+1}(s)\|^3_{\mathrm{max,op}},$$

Our next result gives uniform bounds on $\mathtt{P}_{1:K+1}$ and its derivative by invoking Lemma F.2.

Lemma F.4. *Suppose (for simplicity) that $\lambda_{\min}(\mathbf{Q}), \lambda_{\min}(\mathbf{R}) \geq 1$, and that $(\Delta_A + K_A K_B \Delta_B) \leq 1/8\|\mathtt{P}^{\mathrm{opt}}_{1:K+1}(\mathbf{\Theta})\|^2_{\mathrm{max,op}}$. Then, for all $s \in [0,1]$,*

$$\begin{aligned}
\|\mathtt{P}_{1:K+1}(s)\|_{\mathrm{max,op}} &\leq 1.8\|\mathtt{P}^\star_{1:K+1}\|_{\mathrm{max,op}} \\
\|\mathtt{P}'_{1:K+1}(s)\|_{\mathrm{max,op}} &\leq 12(\Delta_A + K_A K_B \Delta_B)\|\mathtt{P}^\star_{1:K+1}\|^3_{\mathrm{max,op}}
\end{aligned}$$

As the gains $\mathtt{P}_{1:K}(s)$ are explicit function of $\mathtt{P}_{1:K+1}(s)$, we optain the following perturbation bound for the gains.

Lemma F.5. *Under the assumptions of Lemma F.4, the following holds:*

$$\|\mathtt{K}^{\mathrm{opt}}_{1:K}(\mathbf{\Theta}) - \mathtt{K}^{\mathrm{opt}}_{1:K}(\hat{\mathbf{\Theta}})\|_{\mathrm{max,op}} \leq 20C^3 K_A^3 K_B^2 (1 + \tau C K_B)(\Delta_A + \Delta_B), \quad C := \|\mathtt{P}^{\mathrm{opt}}_{1:K+1}(\mathbf{\Theta})\|$$

F.1.1. Proof of Theorem 4

Proof of part (a). Consider the curve

$$\mathtt{K}_k(s) = (1-s)\mathtt{K}^{\mathrm{opt}}_k(\mathbf{\Theta}^\star) + s\mathtt{K}^{\mathrm{opt}}_k(\hat{\mathbf{\Theta}}).$$

We then note that the curve

$$\mathtt{P}^{\mathrm{ce}}_k(s) = \mathtt{P}^{\mathrm{feed}}_k(\mathbf{\Theta}; \mathtt{K}_{1:K}(s))$$

satisfies $\mathtt{P}^{\mathrm{ce}}_k(0) = \mathtt{P}^{\mathrm{ce}}_k(\mathbf{\Theta}^\star; \mathbf{\Theta}^\star) = \mathtt{P}^{\mathrm{opt}}_k(\mathbf{\Theta}^\star) = \mathtt{P}^\star_k$ and $\mathtt{P}^{\mathrm{ce}}_k(0) = \mathtt{P}^{\mathrm{feed}}_k(\mathbf{\Theta}^\star; \mathtt{K}^{\mathrm{opt}}_k(\hat{\mathbf{\Theta}})) = \mathtt{P}^{\mathrm{ce}}_k(\mathbf{\Theta}^\star; \hat{\mathbf{\Theta}})$

By Definition F.2, we can write $\mathtt{P}^{\mathrm{ce}}_k(s) = \mathtt{\Lambda}_k(s)$, where $\mathtt{\Lambda}_k$ solve the following Lyapunov equation

$$\begin{aligned}
&\mathtt{P}^{\mathrm{ce}}_{K+1}(s) = \mathbf{Q}, \quad \mathtt{P}^{\mathrm{ce}}_k(s) = \mathtt{X}_k(s)^\top \mathtt{P}^{\mathrm{ce}}_{k+1}(s)\mathtt{X}_k(s) + \tau\mathbf{Q}(s) + \mathtt{Y}_k(s) \quad \text{where} \\
&\mathtt{X}_k(s) := \mathtt{A}_k + \tau\mathtt{B}_k\mathtt{K}_k(s) \quad \text{and} \quad \mathtt{Y}_k(s) := \tau\mathbf{K}(s)^\top\mathbf{R}\mathbf{K}(s).
\end{aligned} \tag{F.5}$$

As $\mathtt{X}'_k(s) = \tau\mathtt{B}_k\mathtt{K}'_k(s)$ and $\mathtt{Y}'_k(s) = \mathsf{Sym}(\mathbf{K}(s)^\top\mathbf{R}\mathbf{K}'(s))$, salient term from Proposition F.12 evaluates to

$$\begin{aligned}
\Delta(s) &= \max_{j\in[k]} \tau^{-1}\left(2\|\mathtt{X}_j(s)'\| + \|\mathtt{P}^{\mathrm{ce}}_{1:K+1}(s)\|^{-1}_{\mathrm{max,op}}\|\mathtt{Y}_j(s)'\|\right) \\
&= \max_{j\in[k]} \tau^{-1}\left(2\tau K_B\|\mathbf{K}'(s)\| + \tau\|\mathtt{P}^{\mathrm{ce}}_{1:K+1}(s)\|^{-1}_{\mathrm{max,op}}\|\mathtt{K}_k(s)\|\|\mathbf{R}\|\mathtt{K}'_k(s)\|\right) \\
&= \max_{j\in[k]} \left(2K_B + \|\mathtt{P}^{\mathrm{ce}}_{1:K+1}(s)\|^{-1}_{\mathrm{max,op}}\|\mathtt{K}_k(s)\|\right)\|\mathtt{K}'_k(s)\| \qquad (\mathbf{R} = \mathbf{I})
\end{aligned}$$

We further bound

$$\begin{aligned}
\|\mathtt{K}_k(s)\| &= \|(1-s)\mathtt{K}_k^{\mathrm{opt}}(\boldsymbol{\Theta}) + s\mathtt{K}_k^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\| \\
&\le \|\mathtt{K}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\| \vee \|\mathtt{K}_k^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\| \\
&\le \|(\mathbf{R} + \tau\mathtt{B}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\mathtt{B}_k)^{-1}(\mathtt{B}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\mathtt{A}_k)\| \vee \|(\mathbf{R} + \tau\hat{\mathtt{B}}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\hat{\mathbf{B}})^{-1}(\hat{\mathtt{B}}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\hat{\mathtt{A}}_k)\| \\
&\le \|\mathtt{B}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\mathtt{A}_k\| \vee \|\hat{\mathtt{B}}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\hat{\mathtt{A}}_k\| && (\mathbf{R} = \mathbf{I}) \\
&\le K_B K_A(\|\mathtt{P}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\| \vee \|\mathtt{P}_k^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\|) \\
&\le 2K_B K_A \|\mathtt{P}_{1:K+1}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\|_{\max,\mathrm{op}} && \text{(Lemma F.5)} \\
&= 2K_B K_A \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}} && \text{(definition of } \mathtt{P}_k^{\mathrm{ce}})
\end{aligned}$$

Thus,

$$\begin{aligned}
\Delta(s) &\le \left(2K_B + 2K_B K_A \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}}^{-1} \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}}\right) \max_{j\in[k]} \|\mathtt{K}_k'(s)\| && (\mathbf{R} = \mathbf{I}) \\
&\le 4K_B K_A \left(1 \vee \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}}^{-1} \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}}\right) \max_{j\in[k]} \|\mathtt{K}_k'(s)\|. && (K_A \ge 1)
\end{aligned}$$

Hence, setting $\Delta_K := \sup_{s\in[0,1]} \max_{j\in[k]} \|\mathtt{K}_k'(s)\|$, Proposition F.12 implies

$$\begin{aligned}
\mathtt{P}_k^{\mathrm{ce}}(s)' &\le \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}}^2 \Delta(s) \\
&\le 4K_B K_A \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}}^2 \left(1 \vee \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}}^{-1} \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}}\right) \max_{j\in[k]} \Delta_K \\
&\le 4K_B K_A \left(\|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}}^2 \vee \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}}^2\right) \Delta_K
\end{aligned}$$

Hence, Lemma F.2 implies that as long as $4K_B K_A \Delta_K \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}} < 1$, we have

$$\sup_{s\in[0,1]} \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(s)\|_{\max,\mathrm{op}} \le (1 - 4K_B K_A \Delta_K \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}})^{-1} \|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0)\|_{\max,\mathrm{op}}$$

Using $\mathtt{P}_{1:K+1}^{\mathrm{ce}}(0) = \mathtt{P}_{1:K+1}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)$, $\mathtt{P}_{1:K+1}^{\mathrm{ce}}(1) = \mathtt{P}_{1:K+1}^{\mathrm{ce}}(\boldsymbol{\Theta}^\star; \hat{\boldsymbol{\Theta}})$, and defining the shorthand

$$C := \|\mathtt{P}_{1:K+1}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\|,$$

we conclude that for any upper bound $\Delta \ge 4K_B K_A \Delta_K C$ satisfying $\Delta < 1$,

$$\|\mathtt{P}_{1:K+1}^{\mathrm{ce}}(\boldsymbol{\Theta}^\star; \hat{\boldsymbol{\Theta}})\|_{\max,\mathrm{op}} \le (1-\Delta)^{-1} \|\mathtt{P}_{1:K+1}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\|_{\max,\mathrm{op}}.$$

By Lemma F.5, it holds thats if $8\|\mathtt{P}_{1:K+1}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\|_{\max,\mathrm{op}}^2(\Delta_A + K_A K_B \Delta_B) < 1$, we can bound. we can take $\Delta_K \le 20C^3 K_A^3 K_B^2(1+\tau C K_B)(\Delta_A + \Delta_B)$. Hence, we can bound

$$4K_B K_A \Delta_K C \le 80C^4 K_A^4 K_B^3 (1+\tau C K_B)(\Delta_A + \Delta_B) := \Delta_{\mathrm{ce}},$$

which concludes the proof of part (a).

Proof of part (b). We bound

$$\begin{aligned}
\|\mathtt{K}_{1:K}^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\|_{\max,\mathrm{op}} &\le (\|\mathtt{K}_{1:K}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\|_{\max,\mathrm{op}} + 1/4K_B) && \text{(Lemma F.5, definition of } \Delta_{\mathrm{ce}} \text{, and using } K_B, K_A, C \ge 1) \\
&= (1/4K_B) + \|(\mathbf{R} + \tau\mathtt{B}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\mathtt{B}_k)^{-1}\mathtt{B}_k^\top \mathtt{P}_k^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\mathtt{A}_k\|_{\max,\mathrm{op}} \\
& && \text{(Definition of } \mathtt{K}_k^{\mathrm{opt}}\text{, Definition G.1)} \\
&= \frac{1}{4K_B} K_B K_A \|\mathtt{P}_{1:K+1}^{\mathrm{opt}}(\boldsymbol{\Theta}^\star)\|_{\max,\mathrm{op}} && \text{(Definition of } K_A, K_B \text{ in Condition F.1, } \mathbf{R} \succeq \mathbf{I}) \\
&= \frac{1}{4K_B} + K_B K_A C && \text{(Definition of } C) \\
&\le \frac{5}{4} K_B K_A C. && (C, K_A, K_B \ge 1)
\end{aligned}$$

Proof of part (c). We aim to bound the square of the operator norm of the following term

$$\Phi^{\mathrm{ce}}_{j,k} := (\mathtt{A}_{j-1} + \tau \mathtt{B}_{j-1}\mathtt{K}^{\mathrm{opt}}_{j-1}(\hat{\mathbf{\Theta}})) \cdot (\mathtt{A}_{j-2} + \tau \mathtt{B}_{j-2}\mathtt{K}^{\mathrm{opt}}_{j-2}(\hat{\mathbf{\Theta}})) \cdot \cdots \cdot (\mathtt{A}_{k} + \tau \mathtt{B}_{k}\mathtt{K}^{\mathrm{opt}}_{k}(\hat{\mathbf{\Theta}})).$$

Using the fact that $\mathtt{P}^{\mathrm{ce}}_k(\mathbf{\Theta}; \hat{\mathbf{\Theta}})$ solves the Lyapunov equation Eq. (F.5), it follows from Lemma F.10 that if

$$\kappa_0 := \tau^{-1} \max_k \|\mathbf{I} - (\mathtt{A}_k + \tau \mathtt{B}_k \mathtt{K}_k(1))\|_{\mathrm{op}} = \tau^{-1} \max_k \|\mathbf{I} - (\mathtt{A}_k + \tau \mathtt{B}_k \mathtt{K}^{\mathrm{opt}}_k(\hat{\mathbf{\Theta}})\|_{\mathrm{op}} \le 1/2\tau,$$

then

$$\|\Phi^{\mathrm{ce}}_{j,k}\|^2 \le \max\{1, 2\kappa_0\}(1 - \tau\gamma_0)^{j-k}, \quad \gamma_0 := \frac{1}{\|\mathtt{P}^{\mathrm{ce}}_{1:K+1}(\mathbf{\Theta}; \hat{\mathbf{\Theta}})\|_{\max,\mathrm{op}}} \tag{F.6}$$

From part (a), we can lower bound $\gamma_0 \ge \gamma := \frac{1-\Delta_{\mathrm{ce}}}{C}$. Moreover, we can bound κ_0

$$\begin{aligned}
\kappa_0 &\le \tau^{-1} \max_k \|\mathbf{I} - \mathtt{A}_k\| + \max_k \|\mathtt{B}_k\| \|\mathtt{K}^{\mathrm{opt}}_k(\hat{\mathbf{\Theta}})\| \\
&\le \kappa_A + K_B \|\mathtt{K}^{\mathrm{opt}}_{1:K}(\hat{\mathbf{\Theta}})\|_{\max,\mathrm{op}} && \text{(Conditions F.1 and F.4)} \\
&\le \kappa_A + \frac{5}{4} K_B^2 K_A C := \kappa && \text{(Theorem 4(b))}
\end{aligned}$$

As $\kappa \ge 1$, we conclude via Eq. (F.6) that

$$\|\Phi^{\mathrm{ce}}_{j,k}\|^2 \le 2\kappa(1 - \tau\gamma)^{j-k}, \quad \kappa = \kappa_A + \frac{5}{4} K_B^2 K_A C, \quad \gamma = \frac{1 - \Delta_{\mathrm{ce}}}{C}.$$

□

F.2. Proof of Lemma F.3

To apply the self-bounding ODE method, we bound $\mathbf{P}'_{1:H}(s)$ in terms of $\mathtt{P}_{1:K+1}(s)$. To prove Lemma F.3 Let us first introduce some notation. Further, for simplicity, we shall suppress s in equations and let $(\cdot)\big|_s$ to denote evaluation at s. With this convention, define the matrices

$$\Xi_k(s) := (\mathtt{A}'_k + \tau \mathbf{K}_k \mathtt{B}'_k)^\top \mathtt{P}_{k+1}(\mathtt{A}_k + \tau \mathtt{B}_k \mathbf{K}_k) \tag{F.7}$$

and define the operator

$$\mathcal{T}_{k+1}(\cdot\,; s) := \{(\mathtt{A}_k + \mathtt{B}_k \mathbf{K}_k)^\top (\cdot)(\mathtt{A}_k + \mathtt{B}_k \mathbf{K}_k)\}\big|_s, \quad \text{with the convention } \mathcal{T}_{k+1}(\cdot)\big|_s = \mathcal{T}_{k+1}(\cdot; s), \tag{F.8}$$

Lastly, we define their compositions

$$\mathcal{T}_{k+i,k} := \mathcal{T}_k(\cdot) \circ \mathcal{T}_{k+1}(\cdot) \circ \cdots \circ \mathcal{T}_{k+i}(\cdot)\big|_s,$$

with the convention $\mathcal{T}_{k;k}$ is the identity map. These operators give an expression for the derivatives $\mathbf{P}'_{k-1}(s)$.

Lemma F.6. *For all s, it holds that*

$$\mathbf{P}'_k(s) = \sum_{j=k}^{K+1} \mathcal{T}_{k;j}(\Xi_k + \Xi_k^\top)\big|_s$$

Proof. Let $\mathsf{Sym}(\mathbf{X}) = \mathbf{X} + \mathbf{X}^\top$. The Ricatti update (backwards in time) is

$$\mathbf{P}_k = \mathtt{A}_k^\top \mathtt{P}_{k+1} \mathtt{A}_k - \tau(\mathtt{A}_k^\top \mathtt{P}_{k+1} \mathtt{B}_k)(\mathbf{R} + \tau \mathtt{B}_k^\top \mathtt{P}_{k+1} \mathtt{B}_k)^{-1}(\mathtt{A}_k^\top \mathtt{P}_{k+1} \mathtt{B}_k)^\top + \tau \mathbf{Q}$$

Let $\mathsf{Sym}(\mathbf{X}) := \mathbf{X} + \mathbf{X}^\top$. Then, we compute

$$\begin{aligned}
&\mathbf{P}_k(s)' \\
&= \mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{A}_k\right) - \tau\mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k + \mathtt{A}_k^\top \mathtt{P}_{k+1}(\mathtt{B}_k')\right)(\mathbf{R} + \tau\mathbf{B}^\top \mathtt{P}_{k+1}\mathbf{B})^{-1}(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^\top\big) \\
&\quad + \tau^2(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)(\mathbf{R} + \tau\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathsf{Sym}((\mathtt{B}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k))(\mathbf{R} + \tau\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^\top \\
&\quad + \mathtt{A}_k^\top(\mathtt{P}_{k+1}')\mathtt{A}_k - \tau\mathsf{Sym}\left((\mathtt{A}_k^\top(\mathtt{P}_{k+1}')\mathtt{B}_k)(\mathbf{R} + \tau\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^\top\right) \\
&\quad + \tau^2(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)(\mathbf{R} + \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathtt{B}_k^\top \mathtt{P}_{k+1}'\mathtt{B}_k^\top)(\mathbf{R} + \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^\top \\
&= \mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{A}_k + \tau((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k + \mathtt{A}_k^\top \mathtt{P}_{k+1}(\mathtt{B}_k'))\mathbf{K}_k + \tau^2\mathbf{K}_k^\top(((\mathtt{B}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k))\mathbf{K}_k\right) \\
&\quad + \mathtt{A}_k^\top(\mathtt{P}_{k+1}')\mathtt{A}_k + \tau\mathsf{Sym}\left((\mathtt{A}_k^\top(\mathtt{P}_{k+1}')\mathtt{B}_k)\mathbf{K}_k\right) + \tau^2(\mathtt{B}_k\mathbf{K}_k)(\mathtt{P}_{k+1}')(\mathtt{B}_k\mathbf{K}_k) \\
&= \mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{A}_k + \tau((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k + \mathtt{A}_k^\top \mathtt{P}_{k+1}(\mathtt{B}_k'))\mathbf{K}_k + \tau^2\mathbf{K}_k^\top((\mathtt{B}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k))\mathbf{K}_k\right) \\
&\quad + (\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k)^\top(\mathtt{P}_{k+1}')(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k)^\top
\end{aligned}$$

where above we the fact that $\mathbf{K}_k = -(\mathbf{R} + \tau\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathtt{A}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^\top$. Noting that

$$\begin{aligned}
&\mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{A}_k + \tau((\mathtt{A}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k + \mathtt{A}_k^\top \mathtt{P}_{k+1}(\mathtt{B}_k'))\mathbf{K}_k + \tau^2\mathbf{K}_k^\top((\mathtt{B}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k))\mathbf{K}_k\right) \\
&\overset{(i)}{=} \mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k) + \tau\mathtt{A}_k^\top \mathtt{P}_{k+1}(\mathtt{B}_k')\mathbf{K}_k + \tau^2\mathbf{K}_k^\top((\mathtt{B}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k))\mathbf{K}_k\right) \\
&\overset{(ii)}{=} \mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k) + \tau\mathbf{K}_k^\top(\mathtt{B}_k')\mathtt{P}_{k+1}\mathtt{A}_k + \tau^2\mathbf{K}_k^\top((\mathtt{B}_k')^\top \mathtt{P}_{k+1}\mathtt{B}_k))\mathbf{K}_k\right) \\
&= \mathsf{Sym}\left((\mathtt{A}_k')^\top \mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k) + \tau\mathbf{K}_k^\top(\mathtt{B}_k')\mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k)\right) \\
&= \mathsf{Sym}\left((\mathtt{A}_k' + \tau\mathbf{K}_k\mathtt{B}_k')^\top \mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k)\right) := \mathsf{Sym}(\Xi_k)
\end{aligned}$$

Therefore, we have

$$\mathbf{P}_k' = \mathcal{T}_{k+1}(\mathtt{P}_{k+1}') + \mathsf{Sym}(\Xi_k)$$

Thus, unfolding the recursion, we have

$$\begin{aligned}
\mathbf{P}_k' &= \mathcal{T}_{k+1}(\mathtt{P}_{k+1}') + \mathsf{Sym}(\Xi_k) \\
&= \mathcal{T}_{k+1}(\mathcal{T}_{k+2}(\mathtt{P}_{k+1}') + \mathsf{Sym}(\Xi_{k+1})) + \mathsf{Sym}(\Xi_k) \\
&= \mathcal{T}_{k+1}(\mathcal{T}_{k+2}(\mathtt{P}_{k+1}')) + \mathcal{T}_{k+1}(\mathsf{Sym}(\Xi_{k+1})) + \mathsf{Sym}(\Xi_k) \\
&= \dots \\
&= \sum_{j=k}^{K+1} \mathcal{T}_{k;j}(\mathsf{Sym}(\Xi_{k+j})).
\end{aligned}$$

□

Using this fact, a standard Lyapunov argument gives a generic upper bound on sums of these operators.

Lemma F.7. *The operators $\mathcal{T}_{j,k}(\cdot; s)$ are matrix monotone. Hence, if $\mathtt{X}_{1:K}$ are any sequence of $\mathbb{R}^{n\times n}$ matrices,*

$$\|\sum_{j=k}^{K+1} \mathcal{T}_{j,k}(\mathtt{X}_j + \mathtt{X}_j^\top; s)\|_{\mathrm{op}} \le \frac{2\|\mathtt{P}_k\|_{\mathrm{op}} \max_{j\ge k} \|\mathtt{X}_k\|_{\mathrm{op}}}{\tau}\Big|_s$$

Consequently, by Lemma F.6,

$$\|\mathtt{P}_k'(s)\|_{\mathrm{op}} \le \frac{2\|\mathtt{P}_k\|_{\mathrm{op}} \max_{j\ge k} \|\Xi_k\|_{\mathrm{op}}}{\tau}\Big|_s. \tag{F.9}$$

Proof. This is a direct consequence of rewriting $\mathtt{P}_k$ as in Eq. (F.2), applying Lemma F.10(a) with $\mathtt{X}_k = \mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k$, and upper bounding $\|\mathtt{X}_j + \mathtt{X}_j^\top\| \le \|\mathtt{X}_j\|_{\mathrm{op}}$. □

Finally, let us upper bound the norm of the matrices Ξ_k

Lemma F.8.

$$\tau^{-1}\|\Xi_k(s)\| \le (\Delta_A + \Delta_B K_A K_B)\,\|\mathtt{P}_{1:K+1}\|^2_{\max,\mathrm{op}}.$$

Proof of Lemma F.8. Recall

$$\Xi_k := (\mathtt{A}'_k + \tau\mathbf{K}_k\mathtt{B}'_k)^\top \mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k),$$

$$\begin{aligned}
\|\mathbf{K}_k\| &= \|(\mathbf{R} + \tau\mathtt{B}_k^\top\mathtt{P}_{k+1}\mathtt{B}_k)^{-1}\mathtt{B}_k^\top\mathtt{P}_{k+1}\mathtt{A}_k\| \\
&= \lambda_{\min}(\mathbf{R})^{-1}\|\mathtt{B}_k^\top\mathtt{P}_{k+1}\mathtt{A}_k\| \\
&\le \lambda_{\min}(\mathbf{R})^{-1}\|\mathtt{B}_k\|\|\mathtt{A}_k\|\|\mathtt{P}_{k+1}\| \\
&\le \lambda_{\min}(\mathbf{R})^{-1}K_A K_B\|\mathtt{P}_{k+1}\| \\
&\le K_A K_B\|\mathtt{P}_{k+1}\| \qquad \text{(F.10)}
\end{aligned}$$

Next,

$$\begin{aligned}
&\tau^{-1}\|(\mathtt{A}'_k + \tau\mathbf{K}_k\mathtt{B}'_k)^\top\mathtt{P}_{k+1}(\mathtt{A}_k + \tau\mathtt{B}_k\mathbf{K}_k)\| \\
&\le \tau^{-1}\left(\|\mathtt{A}'_k\| + \tau\|\mathtt{B}'_k\|\|\mathbf{K}_k\|\right)\|\mathtt{P}_{k+1}^{\frac12}\|\|\mathtt{P}_{k+1}^{\frac12}(\mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k)\| \\
&\overset{(i)}{\le} \tau^{-1}\left(\|\mathtt{A}'_k\| + \tau\|\mathtt{B}'_k\|\|\mathbf{K}_k\|\right)\|\mathtt{P}_{k+1}^{\frac12}\|\|\mathtt{P}_k^{\frac12}\| \\
&\le \tau^{-1}\left(\|\mathtt{A}'_k\| + \tau\|\mathtt{B}'_k\|K_A K_B\|\mathtt{P}_{k+1}\|\right)\|\mathtt{P}_{k+1}^{\frac12}\|\|\mathtt{P}_k^{\frac12}\| \\
&\le \tau^{-1}\left(\|\mathtt{A}'_k\| + \tau\|\mathtt{B}'_k\|K_A K_B\|\mathtt{P}_{1:K+1}\|_{\max,\mathrm{op}}\right)\|\mathtt{P}_{1:K+1}\|_{\max,\mathrm{op}} \\
&\le \tau^{-1}\left(\|\mathtt{A}'_k\| + \tau\|\mathtt{B}'_k\|K_A K_B\right)\|\mathtt{P}_{1:K+1}\|^2_{\max,\mathrm{op}} \qquad (\|\mathtt{P}_{1:K+1}\|_{\max,\mathrm{op}} \ge \|\mathbf{Q}\| \ge 1) \\
&\le (\Delta_A + \Delta_B K_A K_B)\,\|\mathtt{P}_{1:K+1}\|^2_{\max,\mathrm{op}},
\end{aligned}$$

where in (i) we use Eq. (F.2), which under the present notation gives

$$\mathtt{P}_k = (\mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k)\mathtt{P}_{k+1}(\mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k) + \tau\mathbf{Q},$$

and since $\mathtt{P}_k \succeq \tau\mathbf{Q}$,

$$\|\mathtt{P}_{k+1}^{\frac12}(\mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k)\|^2 = \|(\mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k)\mathtt{P}_{k+1}(\mathtt{A}_k + \mathtt{B}_k\mathtt{K}_k)\| = \|\mathtt{P}_k - \tau\mathbf{Q}\| \le \|\mathtt{P}_k\|.$$

□

Proof of Lemma F.3. From Eq. (F.9) in Lemma F.7, followed by Lemma F.8, we have for $k \in [K]$ that

$$\begin{aligned}
\|\mathtt{P}'_k(s)\|_{\mathrm{op}} &\le \tau^{-1}2\|\mathbf{P}_k\|_{\mathrm{op}}\max_{j\ge k}\|\mathbf{\Xi}_k\|_{\mathrm{op}}\big|_s \\
&\le 2\,(\Delta_A + \Delta_B K_A K_B)\,\|\mathtt{P}_{1:K+1}(0)\|^3_{\max,\mathrm{op}}.
\end{aligned}$$

□

F.3. Proof of Lemma F.4

Let us apply the Lemma F.3 with $\mathbf{v}(s) = \mathbf{\Theta}(s) = (\mathbf{A}_{1:K}(s), \mathbf{B}_{1:K}(s))$ as in Eq. (F.3) and f as the mapping from $(\mathbf{A}_{1:K}, \mathbf{B}_{1:K}) \to \mathtt{P}_{1:K+1}$. This map is algebraic and thus $\mathcal{C}^1$, and $\mathbf{v}(s)$ is also $\mathcal{C}^1$ as it is linear. Finally, take $\|\cdot\|$ to be $\|\cdot\|_{\max,\mathrm{op}}$, take $g(z) = cz^p$, where $p = 3$ and $c = 2(\Delta_A + K_A K_B\Delta_B)$. The corresponding α in Lemma F.3 is $\alpha = c(p-1)\|f(\mathbf{v}(0))\|^{p-1} = 2(\Delta_A + K_A K_B\Delta_B)\|\mathtt{P}_{1:K+1}\|^2$, then, if $\alpha \le 1/4$, i.e. $(\Delta_A + K_A K_B\Delta_B) \le /8\|\mathtt{P}_{1:K+1}(0)\|^2$, we have by Lemma F.3 that

$$\|\mathtt{P}_{1:K}(s)\|_{\max,\mathrm{op}} \le (1-\alpha)^{-\frac{1}{p-1}}\|\mathbf{P}^\star_{1:K}\|_{\max,\mathrm{op}} \le (4/3)^2\|\mathbf{P}^\star_{1:K}\|_{\max,\mathrm{op}} \le 1.8\|\mathbf{P}^\star_{1:K}(0)\|_{\max,\mathrm{op}}.$$

and

$$\|\mathtt{P}_{1:K}(s)'\|_{\max,\mathrm{op}} \le 2(4/3)^6(\Delta_A + K_A K_B \Delta_B)\|\mathbf{P}^\star_{1:K}\|^3_{\max,\mathrm{op}} \le 12(\Delta_A + K_A K_B \Delta_B)\|\mathbf{P}^\star_{1:K}(0)\|^3_{\max,\mathrm{op}}$$

F.4. Perturbation on the gains (Lemma F.5)

Proof. Observe that

$$\mathbf{K}_k = -(\mathbf{R} + \tau \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{A}_k$$

Therefore,

$$\mathbf{K}_k' = (\mathbf{R} + \tau \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1} \cdot (\mathbf{R} + \tau \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)' \cdot \underbrace{(\mathbf{R} + \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{A}_k}_{=\mathbf{K}_k} \tag{F.11}$$

$$- (\mathbf{R} + \tau \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)^{-1}(\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{A}_k)'. \tag{F.12}$$

Introduce the constant $C := \|\mathtt{P}^\star_{1:K+1}\|_{\max,\mathrm{op}}$. Using $\mathbf{R} \succeq \mathbf{I}$, we have

$$\begin{aligned}
\|\mathbf{K}_k'\| &= \|(\mathbf{R} + \tau \mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)'\|\|\mathbf{K}\| + \|(\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{A}_k)'\| \\
&= \tau\|(\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{B}_k)'\|\|\mathbf{K}_k\| + \|\mathtt{B}_k^\top \mathtt{P}_{k+1}\mathtt{A}_k)'\| \\
&\le \tau(2\|\mathtt{B}_k\|\|\mathtt{P}_{k+1}\|\|\mathbf{B}_k'\| + \|\mathtt{B}_k\|^2\|\mathtt{P}_{k+1}'\|)\|\mathbf{K}_k\| + (\|\mathtt{B}_k'\|\|\mathtt{A}_k\| + \|\mathtt{A}_k'\|\|\mathtt{B}_k\|)\|\mathtt{P}_{k+1}\| + \|\mathtt{P}_{k+1}'\|\|\mathtt{A}_k\|) \\
&\le \tau(2K_B\Delta_B\|\mathtt{P}_{k+1}\| + K_B^2\|\mathtt{P}_{k+1}'\|)\|\mathbf{K}_k\| + (\Delta_B K_A + \tau\Delta_A K_B)\|\mathtt{P}_{k+1}\| + \|\mathtt{P}_{k+1}'\|K_A K_B) \\
&\le \tau(2K_B^2 K_A\Delta_B\|\mathtt{P}_{k+1}\|^2 + K_B^3 K_A\|\mathtt{P}_{k+1}'\|\|\mathtt{P}_{k+1}\| + \Delta_A K_B\|\|\mathtt{P}_{k+1}\|) \\
&\quad + \big(\|\mathtt{P}_{k+1}\|\Delta_B K_A + \|\mathtt{P}_{k+1}'\|K_A K_B\big) \\
&\le \tau(2\cdot 1.8^2 K_B^2 K_A\Delta_B C^2 + 12(\Delta_A + K_A K_B\Delta_B)K_B^3 K_A C^4 + 1.8\Delta_A K_B C) \\
&\quad + (1.8C\Delta_B K_A + 12(\Delta_A + K_A K_B\Delta_B)K_A K_B C) \qquad \text{(Lemma F.4)} \\
&\le \tau C^4(2\cdot 1.8^2 K_B^2 K_A\Delta_B + 12(\Delta_A + K_A K_B\Delta_B)K_B^3 K_A + 1.8\Delta_A K_B) \\
&\quad + C^3\,(1.8\Delta_B K_A + 12(\Delta_A + K_A K_B\Delta_B)K_A K_B) \\
&\le \tau C^4 K_B^3 K_A^2(1.8^2\cdot 2\Delta_B + 12(\Delta_A + \Delta_B) + 1.8\Delta_A) \\
&\quad + C^3 K_A^2 K_B^2\,(1.8\Delta_B + 12(\Delta_A + \Delta_B)) \\
&\le C^3 K_A^2 K_B^2(1 + \tau C K_B)(1.8^2\cdot 2\Delta_B + 12(\Delta_A + \Delta_B) + 1.8\Delta_A) \\
&\le 20C^3 K_A^3 K_B^2(1 + \tau C K_B)(\Delta_A + \Delta_B).
\end{aligned}$$

It follows from Taylors theorem that $\|\mathtt{K}_k^{\mathrm{opt}}(\mathbf{\Theta}) - \mathtt{K}_k^{\mathrm{opt}}(\hat{\mathbf{\Theta}})\| \le 20C^3 K_A^3 K_B^2(1+\tau C K_B)(\Delta_A+\Delta_B)$. The result follows. □

F.5. Proof of Lemma F.2

Lemma F.2 is a special case of Simchowitz & Foster (2020, Corollary 3). To check this, we first establish the following special case of Theorem 13 in (Simchowitz & Foster, 2020).

Lemma F.9 (Comparison Lemma). *Fix dimensions $d_1, d_2 \ge 1$, let $\mathcal{V} \subset \mathbb{R}^{d_1}$, let $f : \mathcal{V} \to \mathbb{R}^{d_2}$ be a $\mathcal{C}^1$ map and let $\mathbf{v}(s) : [0,1] \to \mathcal{V}$ be a $\mathcal{C}^1$ curve defined on $[0,1]$. Suppose that $g(\cdot) : \mathbb{R} \to \mathbb{R}$ is non-negative and non-decreasing scalar function, and $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^{d_2}$ such that*

$$\|\frac{\mathrm{d}}{\mathrm{d}s} f(\mathbf{v}(s))\| \le g(\|f(\mathbf{v}(s))\|) \tag{F.13}$$

Finally, let $\eta > 0$ and $\tilde{g} : \mathbb{R} \to \mathbb{R}$ be a scalar function such that (a) for all $z \ge \|v(0)\|$, $\tilde{g}(z) \ge \eta + g(z)$ and (b) the following scalar ODE has a solution on $[0,1]$:

$$z(0) = \|v(0)\| + \eta, \quad z'(s) = \tilde{g}(z(s))$$

Then, it holds that

$$\forall s \in [0,1], \quad \|f(\mathbf{v}(s))\| \le z(s).$$

Proof. Theorem 13 in (Simchowitz & Foster, 2020) proves a more general result for implicit ODEs, such as those that arise in infinite-horizon Riccati equations. We do not need these complication here, so we specialize their result. Define the function $\mathcal{F} : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}^{d_2}$ via $\mathcal{F}(\mathbf{v}, \mathbf{w}) = f(\mathbf{v}(s)) - \mathbf{w}$. It is then clear that $\tilde{\mathbf{w}}(s) = f(\tilde{\mathbf{v}}(s))$ is the unique solution to $\mathcal{F}(\tilde{\mathbf{v}}(s), \tilde{\mathbf{w}}(s)) = 0$ for any $\mathcal{C}^1$ curve $\tilde{\mathbf{v}}(z)$; since f is $\mathcal{C}^1$, any such solution $\tilde{\mathbf{w}}$ is also $\mathcal{C}^1$. Thus, $\mathcal{F}$ is a "valid implicit function" on $\mathcal{V}$ in the sense of Simchowitz & Foster (2020, Definition 3.2) with . Moreover, by Eq. (F.13), $(\mathcal{F}, \mathcal{V}, g, \|\cdot\|, \mathbf{v}(\cdot))$ form a self-bounding tuple in the sense of Simchowitz & Foster (2020, Definition 3.3). The result now follows from Simchowitz & Foster (2020, Theorem 13). □

Proof of Lemma F.2. Take $g(z) = cz^p$. Define $h_\eta = c(z+\eta)^p$. For any $\eta > 0$, there exist an η' such that $h_{\eta'}(z) \le g(z) + \eta$ for $z \ge z_0$. Moreover, as η approaches 0, we can take $\eta' \to 0$ as wel. Solving the ODE $z(0) = \|f(\mathbf{v}(0))\| + \eta$ and $z'(s) = c(z+\eta)^p$, we see the solution is given by

$$\frac{\mathrm{d}z}{(z+\eta')^p} = c\mathrm{d}s.$$

As $z'(s) \ge 0$, it suffices to bound $z(1)$. For $p > 1$, the solution to this ODE when it exists satisfies

$$\frac{1}{(p-1)(\|f(\mathbf{v}(0))\| + \eta + \eta')^{p-1}} - \frac{1}{(p-1)(z(1)+\eta')^{p-1}} = c$$

Rearranging and setting $\eta, \eta' \to 0$ lets check that, as long as $\frac{1}{(p-1)\|f(\mathbf{v}(0))\|^{p-1}} > c$, Lemma F.9 yields

$$\max_{s\in[0,1]} \|f(\mathbf{v}(s))\| \le \left(\frac{1}{\|f(\mathbf{v}(0))\|^{p-1}} - (p-1)c\right)^{\frac{1}{-(p-1)}} = (1-\alpha)^{\frac{1}{-(p-1)}} \|f(\mathbf{v}(0))\|.$$

For $p = 1$, we instead get

$$\ln(z(1)+\eta') - \ln(\|f(\mathbf{v}(0))\| + \eta + \eta') = c$$

Again, taking $\eta', \eta \to 0$, Lemma F.9 yields

$$\max_{s\in[0,1]} \|f(\mathbf{v}(s))\| \le \exp(c + \ln(\|f(\mathbf{v}(0))\|) = \|f(\mathbf{v}(0))\| e^c.$$

□

F.6. Perturbation bounds for Lyapunov Solutions

Lemma F.10 (Basic Lypaunov Theory)**.** *Let $\mathtt{X}_{1:K}$ and $\mathtt{Y}_{1:K}$ be a sequence of matrices of appropriate dimension. Suppose that $\mathtt{Y}_k \succeq 0$, and let $\mathbf{Q} \succeq \mathbf{I}$. Define Λ_k as via the solution to the Lyapunov recursion*

$$\Lambda_{K+1} = \mathbf{Q}, \quad \Lambda_k = \mathtt{X}_k^\top \Lambda_{k+1} \mathtt{X}_k + \tau\mathbf{Q} + \mathtt{Y}_k.$$

and define the matrix $\Phi_{j+1,k} := (\mathbf{X}_j \cdot \mathbf{X}_{j-1} \cdots \cdot \mathbf{X}_{k+1} \cdot \mathbf{X}_k)$, with the convention $\Phi_{k,k} = \mathbf{I}$, let and define the operator $\mathcal{T}_{j,k}(\cdot) = \Phi_{j,k}^\top(\cdot)\Phi_{j,k}$. Then

(a) For any symmetric matrices $\mathtt{Z}_j$, we have

$$\tau \sum_{j=k}^{K} \mathcal{T}_{k,j}(\mathtt{Z}_j) \preceq \max_{j=k}^{K} \|\mathtt{Z}_j\| \Lambda_k.$$

(b) If, in addition, $\max_k \|\mathbf{I} - \mathtt{X}_k\|_{\mathrm{op}} \le \kappa\tau$ for some $\kappa \le 1/2\tau$, $\lambda_{\min}(\Lambda_k) \ge \min\{\frac{1}{2\kappa}, 1\}$

(c) Under the condition in part (b), we have

$$\|\Phi_{j,k}\|^2 \le \max\{1, 2\kappa\}\|\Lambda_{1:K+1}\|_{\max,\mathrm{op}}(1-\tau\gamma)^{j-k}, \quad \gamma := \frac{1}{\|\Lambda_{1:K+1}\|_{\max,\mathrm{op}}}.$$

In particular, $\|\mathcal{T}_{k,j}(\mathtt{Z}_j)\| \le \max\{1, 2\kappa\}\|\Lambda_{1:K+1}\|_{\max,\mathrm{op}}\|\mathtt{Z}_k(j)\|$.

Proof. We begin with part (a). By unfolding the recusion, we get

$$\begin{aligned}
\Lambda_k &= \mathtt{X}_k^\top \Lambda_{k+1}\mathtt{X}_k + \tau(\mathbf{Q} + \tau^{-1}\mathtt{Y}_k) \\
&= \big(\tau\,\mathcal{T}_{k,k}(\mathbf{Q} + \tau^{-1}\mathtt{Y}_k) + \mathcal{T}_{k+1,k}(\Lambda_{k+1})\big) \\
&= \tau\sum_{j=k}^{K}\mathcal{T}_{k,j}(\mathbf{Q} + \tau^{-1}\mathtt{Y}_k) + \mathcal{T}_{K+1,k}(\Lambda_{K+1}) \\
&\succeq \tau\sum_{j=k}^{K}\mathcal{T}_{k,j}(\mathcal{I})
\end{aligned}$$

where in the last line, we use $\mathbf{Q} + \tau^{-1}\mathtt{Y}_k \succeq \mathbf{Q} \succeq \mathbf{I}$. AS $\mathcal{T}_{k,j}(\cdot)$ is a matrix monotone operator, we have that symmetric matrix $\mathbf{Z}$, we have

$$\tau\sum_{j=k}^{K}\mathcal{T}_{j,k}(\mathtt{Z}_j) \preceq \tau\max_{j=k}^{K}\|\mathtt{Z}_j\|\sum_{j=k}^{K}\mathcal{T}_{j,k}(\mathbf{I}) \preceq \Lambda_k,$$

and similarly for $-\tau\sum_{j=k}^{K}\mathcal{T}_{j,k}(\mathtt{Z}_j)$.

Part (b). We argue part (b) by induction backwards on k, noting that $k = K+1$ is immediate. We have

$$\Lambda_k \succeq \mathtt{X}_k^\top\Lambda_{k+1}\mathtt{X}_k + \tau\mathbf{Q} \succeq \mathtt{X}_k^\top\Lambda_{k+1}\mathtt{X}_k + \tau$$

Hence,

$$\begin{aligned}
\lambda_{\min}(\Lambda_k) &\ge \lambda_{\min}(\Lambda_{k+1})\sigma_{\min}(\mathtt{X}_k)^2 + \tau \\
&\ge \lambda_{\min}(\Lambda_{k+1})(1 - \|\mathtt{X}_k - \mathbf{I}\|)^2 + \tau \\
&\ge \lambda_{\min}(\Lambda_{k+1})(1 - \kappa\tau)^2 + \tau \\
&\ge \lambda_{\min}(\Lambda_{k+1})(1 - 2\kappa\tau) + \tau
\end{aligned}$$

Applying the inductive hypothesis, we see that the above is at least

$$\lambda_{\min}(\Lambda_k) \ge \frac{1}{2\kappa}(1 - 2\kappa\tau) + \tau \ge \frac{1}{2\kappa}, \text{ as needed.}$$

Part (c). We have

$$\mathtt{X}_j^\top\Lambda_{j+1}\mathtt{X}_j = \Lambda_j - \tau(\mathbf{Q} + \mathtt{Y}_k) \preceq \Lambda_j(1 - \tau\Lambda_j^{-1/2}\mathbf{Q}\Lambda_j^{-1/2}) \preceq \Lambda_j(1-\tau\gamma),$$

where we recall $\gamma = 1/\|\Lambda_{1:K+1}\|_{\max,\mathrm{op}}$ and use $\mathbf{Q} \succeq \mathbf{I}$. By unfolding the bound, we find

$$\|\Phi_{j+1,k}^\top\Lambda_{j+1}\Phi_{j+1,k}\| \le (1-\tau\gamma)^{j+1-k}\|\Lambda_k\|.$$

Hence,

$$\|\Phi_{j+1,k}\|^2 \le \lambda_{\min}(\Lambda_{j+1})^{-1}\|\Phi_{j+1,k}^\top\Lambda_{j+1}\Phi_{j+1,k}\| \le 2\|\Lambda_{j+1}\|\kappa(1-\tau\gamma)^{j+1-k}.$$

Moreover, as $\kappa \ge 1$, the bound also applies to $\Phi_{j+1,j+1} = \mathbf{I}$. □

Lemma F.11 (Formula for Lyapunov Curve Derivatives)**.** *Consider curves* $\mathtt{X}_{1:K}(s), \mathtt{Y}_{1:K}(s)$*, let* $\mathbf{Q} \succeq I$*, and define*

$$\mathtt{\Lambda}_{K+1}(s) = \mathbf{Q}, \quad \mathtt{\Lambda}_k(s) = \mathtt{X}_k(s)^\top \mathtt{\Lambda}_{k+1}\mathtt{X}_k(s) + \tau\mathbf{Q} + \mathtt{Y}_k(s)$$

Again, let $\mathtt{\Phi}_{k,j} := (\mathbf{X}_{j-1} \cdot \mathbf{X}_{j-2} \cdots \cdot \mathbf{X}_{k+1} \cdot \mathbf{X}_k)$*, with the convention* $\mathtt{\Phi}_{k,k} = \mathbf{I}$*, define the operator* $\mathcal{T}_{k,j}(\cdot) = \mathtt{\Phi}_{k,j}^\top(\cdot)\mathtt{\Phi}_{j,j}$*. Then,*

$$\mathtt{\Lambda}'_k = \sum_{j=k}^{K} \mathcal{T}_{k,j}(\mathtt{\Omega}_k), \quad \mathtt{\Omega}_k := \mathsf{Sym}(\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}\mathtt{X}'_k) + \mathtt{Y}'_k.$$

Proof. We compute

$$\mathtt{\Lambda}'_k = \underbrace{\mathsf{Sym}(\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}\mathtt{X}'_k) + \mathtt{Y}'_k}_{=\mathtt{\Omega}_k} + \mathtt{X}_k(s)^\top \mathtt{\Lambda}_{k+1}\mathtt{X}_k(s).$$

The result follows by unfolding the recursion, with the base case $\mathtt{\Lambda}'_{K+1} = \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{Q} = 0$. □

We now state our Lyapunov perturbation bound:

Proposition F.12 (Lyapunov Function Perturbation)**.** *Consider curves* $\mathtt{X}_{1:K}(s), \mathtt{Y}_{1:K}(s)$*, and define for* $\mathbf{Q} \succeq \mathbf{I}$

$$\mathtt{\Lambda}_{K+1}(s) = \mathbf{Q}, \quad \mathtt{\Lambda}_k(s) = \mathtt{X}_k(s)^\top \mathtt{\Lambda}_{k+1}\mathtt{X}_k(s) + \tau\mathbf{Q} + \mathtt{Y}_k(s)$$

Then,

$$\begin{aligned}
&\mathtt{\Lambda}_k(s)' \le \|\mathtt{\Lambda}_{1:K+1}(s)\|^2_{\max,\mathrm{op}}\Delta(s), \quad \textit{where} \\
&\Delta(s) = \max_{j\in[k]} \tau^{-1}\left(2\|\mathtt{X}_j(s)'\| + \|\mathtt{\Lambda}_{1:K+1}(s)\|^{-1}_{\max,\mathrm{op}}\|\mathtt{Y}_j(s)'\|\right)
\end{aligned}$$

Moreover, as long as $\|\mathtt{\Lambda}_{1:K+1}(0)\|_{\max,\mathrm{op}} \sup_{s\in[0,1]} \Delta(s) < 1$*,*

$$\max_{s\in[0,1]} \|\mathtt{\Lambda}_{1:K+1}(s)\|_{\max,\mathrm{op}} \le \left(1 - \|\mathtt{\Lambda}_{1:K+1}(0)\|_{\max,\mathrm{op}} \sup_{s\in[0,1]} \Delta(s)\right)^{-1} \|\mathtt{\Lambda}_{1:K+1}(0)\|_{\max,\mathrm{op}},$$

The above bound also holds when $\Delta(s)$ *is replaced by the simpler term*

$$\tilde{\Delta}(s) := \max_{j\in[k]} \tau^{-1}\left(2\|\mathtt{X}_j(s)'\| + \|\mathtt{Y}_j(s)'\|\right) \tag{F.14}$$

Proof. We write

$$\mathtt{\Lambda}'_k = \sum_{j=k}^{K} \mathcal{T}_{k,j}(\mathtt{\Omega}_k), \quad \mathtt{\Omega}_k := \mathsf{Sym}(\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}\mathtt{X}'_k) + \mathtt{Y}'_k \tag{F.15}$$

We have

$$\begin{aligned}
\|\mathtt{\Omega}_k\| &\le 2\|\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}\|\|\mathtt{X}'_k\| + \|\mathtt{Y}'_k\| \\
&\le 2\|\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}^{\frac{1}{2}}\|\|\mathtt{\Lambda}_{k+1}^{\frac{1}{2}}\|\|\mathtt{X}'_k\| + \|\mathtt{Y}'_k\|.
\end{aligned}$$

Observe that $\|\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}^{\frac{1}{2}}\| = \|\mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}\mathtt{X}_k\|^{\frac{1}{2}}$. As $0 \preceq \mathtt{X}_k^\top \mathtt{\Lambda}_{k+1}\mathtt{X}_k^\top = \mathtt{\Lambda}_k - \mathtt{Y}_k \preceq \mathtt{\Lambda}_k$, we conclude,

$$\begin{aligned}
\|\mathtt{\Omega}_k\| &\le 2\|\mathtt{\Lambda}_k\|^{\frac{1}{2}}\|\mathtt{\Lambda}_{k+1}^{\frac{1}{2}}\|\|\mathtt{X}'_k\| + \|\mathtt{Y}'_k\| \\
&\le \|\mathtt{\Lambda}_{1:K+1}\|_{\max,\mathrm{op}} 2\|\mathtt{X}'_k\| + \|\mathtt{\Lambda}_{1:K+1}(s)\|^{-1}_{\max,\mathrm{op}}\|\mathtt{Y}'_k\| \\
&\le \|\mathtt{\Lambda}_{1:K+1}\|_{\max,\mathrm{op}}\left(2\|\mathtt{X}'_k\| + \|\mathtt{\Lambda}_{1:K+1}(s)\|^{-1}_{\max,\mathrm{op}}\|\mathtt{Y}'_k\|\right) \\
&\le \|\mathtt{\Lambda}_{1:K+1}\|_{\max,\mathrm{op}} \max_{j\in[K]}\left(2\|\mathtt{X}'_j\| + \|\mathtt{\Lambda}_{1:K+1}(s)\|^{-1}_{\max,\mathrm{op}}\|\mathtt{Y}'_j\|\right) \le \tau\|\mathtt{\Lambda}_{1:K+1}\|_{\max,\mathrm{op}}\Delta(s).
\end{aligned}$$

Thus, from Eq. (F.15) and Lemma F.10, we conclude

$$\begin{aligned}\Lambda_k(s)' &\le \tau^{-1}\cdot\tau\|\Lambda_k\|\|\Lambda_{1:K+1}(s)\|_{\max,\mathrm{op}}\Delta(s)\\ &\le \|\Lambda_{1:K+1}(s)\|^2_{\max,\mathrm{op}}\Delta(s)\end{aligned}$$

The final result follows by applying Lemma F.2 with $p=2$, $c=\max_{s\in[0,1]}\Delta$, and $\alpha=\Delta(s)\|\Lambda_{1:K+1}(0)\|_{\max,\mathrm{op}}$. That Eq. (F.14) follows from the fact that if $\mathbf{Q}\succeq\mathbf{I}$, $\|\Lambda_{1:K+1}(s)\|_{\max,\mathrm{op}}\ge 1$. □

Lemma F.13 (Average Perturbation)**.** *Let $\kappa\le 1/2\tau$. Consider a curve $\mathtt{X}_{1:K}(s)$ such $\max_k\sup_{s\in[0,1]}\|\mathbf{I}-\mathtt{X}_k(s)\|_{\mathrm{op}}\le\kappa\tau$. Then,*

$$\Lambda_{K+1}(s)=\mathbf{I},\quad \Lambda_k(s)=\mathtt{X}_k(s)^\top\Lambda_{k+1}\mathtt{X}_k(s)+\tau\mathbf{I}.$$

Fix

$$\Delta_{\mathrm{sum}}=3\sup_{s\in[0,1]}\max\{1,2\kappa\}\sum_{k=1}^K\|\mathtt{X}_k'(s)\|.$$

Then, as long as $\|\Lambda_{1:K+1}(0)\|_{\max,\mathrm{op}}\Delta_{\mathrm{sum}}<1$, we have

$$\|\Lambda_{1:K+1}(1)\|_{\max,\mathrm{op}}\le(1-\|\Lambda_{1:K+1}(0)\|_{\max,\mathrm{op}}\Delta_{\mathrm{sum}})^{-1}\|\Lambda_{1:K+1}(0)\|_{\max,\mathrm{op}}.$$

Proof. We have that

$$\Lambda_k'=\sum_{j=k}^K\mathcal{T}_{k,j}(\Omega_k),\quad \Omega_k:=\mathsf{Sym}(\mathtt{X}_k^\top\Lambda_{k+1}\mathtt{X}_k')., \tag{F.16}$$

so by Lemma F.10(c),

$$\begin{aligned}\|\Lambda_k'\|_{\mathrm{op}}&\le\|\Lambda_{1:K+1}\|_{\max,\mathrm{op}}\max\{1,2\kappa\}\sum_{j=k}^K\|\Omega_j\|\\ &\le 2\|\Lambda_{1:K+1}\|_{\max,\mathrm{op}}\max\{1,2\kappa\}\sum_{j=k}^K\|\mathtt{X}_k\|\|\mathtt{X}_k'\|\|\Lambda_{k+1}\|\\ &\le 2\|\Lambda_{1:K+1}\|^2_{\max,\mathrm{op}}\max\{1,2\kappa\}\sum_{j=k}^K\|\mathtt{X}_k'\|\|\mathtt{X}_k\|\\ &\le 2\|\Lambda_{1:K+1}\|^2_{\max,\mathrm{op}}\max\{1,2\kappa\}\underbrace{(1+\kappa\tau)}_{\le\frac{3}{2}}\sum_{j=k}^K\|\mathtt{X}_k'\|\\ &\le\Delta_{\mathrm{sum}}\|\Lambda_{1:K+1}\|^2_{\max,\mathrm{op}}.\end{aligned}$$

The result now follows from Lemma F.2. □

G. Instantiantions of Certainty Equivalence Bound

Definition G.1. Given a sequence of gains $\tilde{\mathtt{K}}_{1:K}\in(\mathbb{R}^{d_{\mathsf{u}}\times d_{\mathsf{x}}})^K$, we define the discrete cost-to-go matrix as

$$\mathtt{P}^\pi_{K+1}=\mathbf{I},\quad \mathtt{P}^\pi_k[\tilde{\mathtt{K}}_{k:K}]=(\mathtt{A}^\pi_{\mathrm{ol},k}+\mathtt{B}^\pi_{\mathrm{ol},k}\tilde{\mathtt{K}}_k)^\top\mathtt{P}^\pi_{k+1}[\tilde{\mathtt{K}}_{k+1:K}](\mathtt{A}^\pi_{\mathrm{ol},k}+\mathtt{B}^\pi_{\mathrm{ol},k}\tilde{\mathtt{K}}_k)+\tau(\mathbf{I}+\tilde{\mathtt{K}}_k^\top\tilde{\mathtt{K}}_k).$$

The follow is standard (see, e.g. Anderson & Moore (2007, Section 2.4)).

Lemma G.1. *There exists a unique minimizer sequence $\mathtt{K}^{\pi,\star}_{1:K}$ such that, for all other $\tilde{\mathtt{K}}_{1:K}$, $\mathtt{P}^\pi_k[\mathtt{K}^{\pi,\star}_{k:K}]\preceq\mathtt{P}^\pi_k[\tilde{\mathtt{K}}_{k:K}]$. We denote this minimize P-matrix $\mathtt{P}^\pi_{\mathrm{opt},k}:=\mathtt{P}^\pi_k[\mathtt{K}^{\pi,\star}_{k:K}]$.*

Proposition G.2. *Recall the definition of* τ_{ric} *from Definition A.2,*

$$\tau_{\mathrm{ric}} := \frac{1}{4\mu_{\mathrm{ric}}^2\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_f^2(1+L_f\mu_{\mathrm{ric}})^2\right)} = \frac{1}{\mathcal{O}_\star(1)}$$

Then, as long as $\tau \le \min\{\tau_{\mathrm{ric},\pi}, 1/4L_f\}$*, it holds that for any feasible policy* π*,* $\max_{k\in[K+1]}\|\mathtt{P}^\pi_{\mathrm{opt},k}\| \le 2\mu_{\mathrm{ric}}$.

The following lemma bounding the constant K_π for the initial policy π can be estabilished along the same lines of Proposition G.2. Its proof is given in Appendix G.2.

Lemma G.3. *Suppose that* $\tau \le \tau_{\mathrm{ric}}$*. For* $\pi = \pi^{(1)}$*,* $\mu_{\pi,\star} \le 2\mu_{\mathrm{ric}}$ *and* $L_\pi = 1$.

Proposition G.4 (Certainty Equivalence Bound)**.** *Let* $\hat{\mathtt{A}}^\pi_k$ *and* $\hat{\mathtt{B}}^\pi_k$ *be estimates of* $\mathtt{A}^\pi_{\mathrm{ol},k}$ *and* $\mathtt{B}^\pi_{\mathrm{ol},k}$*, and let* $\hat{\mathtt{K}}_k$ *denote the corresponding certainty equivalence controller sythesized by solving the following recursion given by* $\hat{\mathtt{P}}_{K+1} = \mathbf{I}$*, and for* $k \in [k_0 : K]$*, setting*

$$\hat{\mathtt{P}}_k = (\hat{\mathtt{A}}^\pi_k)^\top\hat{\mathtt{P}}_{k+1}\hat{\mathtt{A}}^\pi_k - \left(\hat{\mathtt{B}}^\pi_k\hat{\mathtt{P}}_{k+1}\hat{\mathtt{A}}^\pi_k\right)^\top(\tau^{-1}\mathbf{I} + (\hat{\mathtt{B}}^\pi_k)^\top\hat{\mathtt{P}}_{k+1}\hat{\mathtt{B}}^\pi_k)^{-1}\left(\hat{\mathtt{B}}^\pi_k\hat{\mathtt{P}}_{k+1}\hat{\mathtt{A}}^\pi_k\right) + \tau\mathbf{I}$$
$$\hat{\mathtt{K}}_k = -(\tau^{-1}\mathbf{I} + (\hat{\mathtt{B}}^\pi_k)^\top\hat{\mathtt{P}}_k\hat{\mathtt{B}}^\pi_k)^{-1}(\hat{\mathtt{B}}^\pi_k)^\top\hat{\mathtt{P}}_k\hat{\mathtt{A}}^\pi_k,$$

Then, as long as $\max_{k\in[k_0:K]}\|\hat{\mathtt{A}}^\pi_k - \mathtt{A}^\pi_{\mathrm{ol},k}\|_{\mathrm{op}} \vee \|\hat{\mathtt{B}}^\pi_k - \mathtt{B}^\pi_{\mathrm{ol},k}\|_{\mathrm{op}} \le (2^{17}\tau\mu^4_{\mathrm{ric}}\max\{1, L_f^3\})^{-1}$*, and* $\tau \le \min\{\tau_{\mathrm{ric}}, 1/4L_fL_\pi\}$*, we have*

$$\max_{k\ge k_0}\|\mathtt{P}^\pi_k[\hat{\mathtt{K}}_{k:K}]\| \le 4\mu_{\mathrm{ric}}, \quad \textit{and} \quad \max_{k\ge k_0}\|\hat{\mathtt{K}}_k\| \le 6\max\{1, L_f\}\mu_{\mathrm{ric}}.$$

We can now prove Proposition A.14.

Proof. Let $\hat{\mathtt{K}}_{1:K}$ be the gains synthesized according to Algorithm 1(Line 7-10), and $\pi' = (\mathtt{u}^\pi_{1:K}, \hat{\mathtt{K}}_{1:K})$ be the policy with the same inputs as π but with these new gains. Definining the shorthand $\tilde{\mathtt{P}}_k := \|\mathtt{P}^\pi_k[\hat{\mathtt{K}}_{k:K}]\|$, Proposition G.4 then implies that

$$\max_{k\ge k_0}\|\tilde{\mathtt{P}}_k\| \le 4\mu_{\mathrm{ric}}, \quad \text{and} \quad \max_{k\ge k_0}\|\hat{\mathtt{K}}_k\| \le 6\max\{1, L_f\}\mu_{\mathrm{ric}}.$$

Since $\hat{\mathtt{K}}_k = 0$ for $k < k_0$, we conclude $\max_{k\in[K+1]}\|\hat{\mathtt{K}}_k\| \le 6\max\{1, L_f\}\mu_{\mathrm{ric}}$, which we note is ≥ 1 as $\mu_{\mathrm{ric}} \ge 1$. Thus, we can take $L_{\pi'} = 6\max\{1, L_f\}\mu_{\mathrm{ric}}$. Moreover, for this policy π, we have $\mathtt{A}^{\pi'}_{\mathrm{cl},k} = (\mathtt{A}^\pi_{\mathrm{ol},k} + \mathtt{B}^\pi_{\mathrm{ol},k}\hat{\mathtt{K}}_k)$, so that the matrices $\tilde{\mathtt{P}}_k$ are given by the recursion

$$\tilde{\mathtt{P}}_{K+1} = \mathbf{I}, \quad \tilde{\mathtt{P}}_k = (\mathtt{A}^{\pi'}_{\mathrm{cl},k})^\top\tilde{\mathtt{P}}_{k+1}(\mathtt{A}^{\pi'}_{\mathrm{cl},k}) + \tau(\mathbf{I} + \hat{\mathtt{K}}_k^\top\hat{\mathtt{K}}_k).$$

Hence, $\tilde{\mathtt{P}}_k \succeq \Lambda^{\pi'}_k$, where we recal that $\Lambda^{\pi'}_k$ satisfy the recursion

$$\Lambda^{\pi'}_{K+1} = \mathbf{I}, \quad \Lambda^{\pi'}_k = (\mathtt{A}^{\pi'}_{\mathrm{cl},k})^\top\Lambda^{\pi'}_{k+1}(\mathtt{A}^{\pi'}_{\mathrm{cl},k}) + \tau\mathbf{I}.$$

Thus, $\mu_{\pi',\star} := \max_{k\in[k_0:K]}\|\Lambda^{\pi'}_k\| \le \max_{k\in[k_0:K]}\|\tilde{\mathtt{P}}_k\| \le 4\mu_{\mathrm{ric}}$. □

G.1. Proof of Proposition G.4

Essentially, we instantiate Theorem 4 with appropriate bounds on parameters, and using the last part of the recursion for $k \ge k_0$. Fix an index $k_0 \in [K]$, let $K_0 = K - k_0 - 1$, and recall $[k_0 : j] := \{k_0, \dots, j\}$. Throughout, we suppose

$$\tau \le 1/4L_f\max\{1, L_\pi, \mu_{\mathrm{ric}}\} \tag{G.1}$$

Suppose we have givens estimates $\hat{\mathtt{A}}^\pi_k$ and $\hat{\mathtt{B}}^\pi_k$ satisfying

$$\max_{k\in[k_0:K]}\tau^{-1}\|\hat{\mathtt{A}}^\pi_k - \mathtt{A}^\pi_{\mathrm{ol},k}\|_{\mathrm{op}} \le \epsilon_A, \quad \max_{k\in[k_0:K]}\tau^{-1}\|\hat{\mathtt{B}}^\pi_k - \mathtt{B}^\pi_{\mathrm{ol},k}\|_{\mathrm{op}} \le \epsilon_B$$

We apply Theorem 4 with the substitutions

$$K \leftarrow K_0, \quad \hat{\mathtt{B}}_k \leftarrow \tau\hat{\mathtt{B}}^{\pi}_{k+k_0-1}, \quad \mathtt{B}_k \leftarrow \tau\mathtt{B}^{\pi}_{\mathrm{ol},k+k_0-1}, \quad \hat{\mathtt{A}}_k \leftarrow \hat{\mathtt{A}}^{\pi}_{k+k_0-1}, \quad \mathtt{A}_k \leftarrow \tau\mathtt{A}^{\pi}_{\mathrm{ol},k+k_0-1},$$

So that, with $\boldsymbol{\Theta} = (\mathtt{A}_j, \mathtt{B}_j)_{j\in[K_0]}$ and $\hat{\boldsymbol{\Theta}} = (\hat{\mathtt{A}}_j, \hat{\mathtt{B}}_j)_{j\in[K_0]}$, we have

$$\hat{\mathtt{K}}_k = \begin{cases} 0 & k < k_0 \\ \mathtt{K}^{\mathrm{opt}}_{k-k_0}(\hat{\boldsymbol{\Theta}}) & k \geq k_0 \end{cases}$$

and thus,

$$\mathtt{P}^{\mathrm{opt}}_j(\boldsymbol{\Theta}) = \mathtt{P}^{\pi}_{\mathrm{opt},k}, \quad \mathtt{P}^{\mathrm{ce}}_k(\boldsymbol{\Theta}; \hat{\boldsymbol{\Theta}}) = \mathtt{P}^{\pi}_{j+k_0-1}[\hat{\mathtt{K}}_{k_0:K}].$$

With the above substitutions, we can apply Proposition G.2 as long as τ satisfies the condition stipulated in that proposition, we have

$$\max_{j\in[K_0+1]} \|\mathtt{P}^{\mathrm{opt}}_j(\boldsymbol{\Theta})\| \leq 2\mu_{\mathrm{ric}}. \tag{G.2}$$

Moreover, we have that by Lemmas I.3, I.4 and I.7, the following holds for $\tau \leq 1/4L_f \max\{1, L_\pi\}$,

$$\begin{aligned} &\max_k \tau^{-1}\|\mathtt{B}^{\pi}_{\mathrm{ol},k}\| \leq \exp(1/4)L_f \\ &\max_k \|\mathtt{A}^{\pi}_{\mathrm{ol},k}\| = \max_k \|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k+1,k}\| \leq \frac{5}{3} \\ &\max_{k\in[K]} \|\mathtt{A}_k - \mathbf{I}\| = \max_{k\in[K]} \|\boldsymbol{\Phi}^{\pi}_{\mathrm{ol}}(t_{k+1}, t_k) - \mathbf{I}\| \leq \exp(1/4)\tau L_f \end{aligned} \tag{G.3}$$

Hence Conditions F.1 and F.4 hold for

$$\begin{aligned} K_A &= 1 \vee \max_{k\in[k_0:K]} \|\mathtt{A}^{\pi}_{\mathrm{ol},k}\| \vee \|\hat{\mathtt{A}}^{\pi}_k\| \leq \frac{5}{3} \\ K_B &= 1 \vee \max_{k\in[k_0:K]} \tau^{-1}(\|\mathtt{B}^{\pi}_{\mathrm{ol},k}\| \vee \|\hat{\mathtt{B}}^{\pi}_k\|) \leq \exp(1/4)\max\{1, L_f\} \\ \kappa_A &:= \tau^{-1} \max_{k\in[k_0:K]} \|\mathtt{A}_k - \mathbf{I}\| = \exp(1/4)L_f. \end{aligned} \tag{G.4}$$

Moreover, Condition F.2 holds with $\Delta_A = \epsilon_A$, $\Delta_B = \epsilon_B$. We can now apply Theorem 4. We take and for $\epsilon_A \leq 1/3$ and $\epsilon_B \leq L_f/2$, we may take

$$\begin{aligned} \Delta_{\mathrm{ce}} &:= 80C^4K_A^3K_B^3(1 + \tau C K_B)(\Delta_A + \Delta_B) \\ C &:= \max_{j\in[K_0+1]} \|\mathtt{P}^{\mathrm{opt}}_j(\boldsymbol{\Theta})\| \leq 2\mu_{\mathrm{ric}} \qquad \text{(by Eq. (G.2))} \end{aligned}$$

And we can bound (recalling $\tau \leq 1/4L_f\mu_{\mathrm{ric}}$ and $\mu_{\mathrm{ric}} \geq 1$)

$$\begin{aligned} \Delta_{\mathrm{ce}} &\leq 80\mu_{\mathrm{ric}}^4 \cdot (16 \cdot (5/3)^3 \cdot \exp(3/4))\max\{1, L_f^3\}(1 + 4\tau L_f\mu_{\mathrm{ric}})(\epsilon_A + \epsilon_B) \\ &\leq 2^{14}\mu_{\mathrm{ric}}^4 \cdot \max\{1, L_f^3\}(\epsilon_A + \epsilon_B)(1 + 4\tau L_f\mu_{\mathrm{ric}}) \\ &\leq 2^{15}\mu_{\mathrm{ric}}^4 \cdot \max\{1, L_f^3\}(\epsilon_A + \epsilon_B). \end{aligned}$$

Hence, as long as

$$2^{16}\mu_{\mathrm{ric}}^4 \max\{1, L_f^3\}(\epsilon_A + \epsilon_B) \leq 1,$$

we have

$$\Delta_{\mathrm{ce}} \leq 1/2 \tag{G.5}$$

and therefore, by Theorem 4(a),

$$\max_{k\in[k_0:K+1]} \|\mathtt{P}_k^{\pi}[\hat{\mathtt{K}}_{k_0:K}]\| = \max_{j\in[K_0+1]} \|\mathtt{P}_j^{\mathrm{ce}}(\boldsymbol{\Theta};\hat{\boldsymbol{\Theta}})\| \le 2 \max_{j\in[K_0+1]} \|\mathtt{P}_j^{\mathrm{opt}}(\boldsymbol{\Theta})\| \le 4\mu_{\mathrm{ric}}.$$

Next, Theorem 4(b), we can take $L_{\pi'} = 6\max\{1, L_f\}\mu_{\mathrm{ric}}$:

$$\begin{aligned}\max_{k\in[K]} \|\hat{\mathbf{K}}^{\pi'}\| &= \max_{j\in[K_0]} \|\mathtt{K}_j^{\mathrm{opt}}(\hat{\boldsymbol{\Theta}})\| \\ &\le \frac{5}{4}K_B K_A C \le \frac{5}{4}(5/3)\exp(1/1)\max\{1, L_f\}\cdot 2\mu_{\mathrm{ric}} \le L_{\pi'} := 6\max\{1, L_f\}\mu_{\mathrm{ric}}.\end{aligned}$$

□

G.2. Proof of Proposition G.2

G.2.1. Preliminaries.

We recall the following, standard definition of continuous-time cost to-go matrices (see, e.g. (**?**)Section 2.3]anderson2007optimal):

Definition G.2 (Cost-to-Go Matrices)**.** Given a *policy* π, and a sequence of controls $\tilde{\mathbf{u}}(\cdot) \in \mathcal{U}$, let $\mathbf{P}^{\pi}(\cdot \mid \tilde{\mathbf{u}})$ as the cost-to-go matrix satisfying $\xi^{\top}\mathbf{P}^{\pi}(t \mid \tilde{\mathbf{u}})\xi = \int_{s=t}^{T}(\|\tilde{\mathbf{x}}(s)\|^2 + \|\tilde{\mathbf{u}}(s)\|^2)\mathrm{d}s + \|\tilde{\mathbf{x}}(T)\|^2$, under the dynamics $\frac{\mathrm{d}}{\mathrm{d}s}\tilde{\mathbf{x}}(s) = \mathbf{A}_{\mathrm{ol}}^{\pi}(s)\tilde{\mathbf{x}}(s) + \mathbf{B}_{\mathrm{ol}}^{\pi}(s)\tilde{\mathbf{u}}(s), \quad \tilde{\mathbf{x}}(t) = \xi$. We let $\mathbf{P}_{\mathrm{opt}}^{\pi}(t)$ denote the optimal cost-to-go matrix, i.e., the matrix satisfying $\xi^{\top}\mathbf{P}_{\mathrm{opt}}^{\pi}(t)\xi = \min_{\tilde{\mathbf{u}}\in\mathcal{U}}\xi^{\top}\mathbf{P}^{\pi}(t \mid \tilde{\mathbf{u}})\xi := V^{\pi}(t \mid \tilde{\mathbf{u}}, \xi)$.

Recall that Assumption 4.3 implies $V^{\pi}(t \mid \tilde{\mathbf{u}}, \xi) \le \mu_{\mathrm{ric}}\|\xi\|^2$, so that $\|\mathbf{P}_{\mathrm{opt}}^{\pi}(t)\| \le \mu_{\mathrm{ric}}$. In what follows, we supress superscript dependence on π, assume π is feasible, and adopt the shorthand $\mathbf{P}(t) = \mathbf{P}_{\mathrm{opt}}^{\pi}(t)$, $\mathbf{A}(t) = \mathbf{A}_{\mathrm{ol}}^{\pi}(t)$, $\mathbf{B}(t) = \mathbf{B}_{\mathrm{ol}}^{\pi}(t)$, $\mathbf{x}(t) = \mathbf{x}^{\pi}(t)$, and $\mathbf{u}(t) = \mathbf{u}^{\pi}(t)$. We also use the shorthand

$$L_{\mathrm{cl}} := L_f(1 + L_f\mu_{\mathrm{ric}}). \tag{G.6}$$

The optimal input defining $\mathbf{P}(t)$ in Assumption 4.3 selects $\tilde{\mathbf{u}}(t) = \mathbf{K}(t)\tilde{\mathbf{x}}(t)$, where $\mathbf{K}(t) = \mathbf{B}(t)^{\top}\mathbf{P}(t)$ (again, Anderson & Moore (2007, Section 2.3)). Introduce the evaluations of the *continuous* value function $\mathbf{P}(t)$ and $\mathbf{K}(t)$ at the time steps t_k:

$$\mathtt{P}_k^{\mathrm{ct}} := \mathbf{P}(t_k), \quad \mathtt{K}_k^{\mathrm{ct}} := \mathbf{K}(t_k) \tag{G.7}$$

We also define an *suboptimal* discrete-time value function by taking $\mathtt{P}_k^{\mathrm{sub}} = \mathtt{P}_k^{\pi}[\mathtt{K}_{1:K}^{\mathrm{ct}}]$, defined in Definition G.1, which satisfies

$$\mathtt{P}_k^{\mathrm{sub}} \succeq \mathtt{P}_{\mathrm{opt},k}^{\pi}.$$

by optimality of $\mathtt{P}_{\mathrm{opt},k}^{\pi}$. Hence, it suffices to bound $\mathtt{P}_k^{\mathrm{sub}}$. To do this, first express both $\mathtt{P}_k^{\mathrm{sub}}$ and $\mathtt{P}_k^{\mathrm{ct}}$ as discrete Lyapunov recusions. To do so, we require the relevant transition operators.

Definition G.3 (Relevant Transitions Operators)**.** For $k \in [K]$ and $s \in \mathcal{I}_k$, let $\boldsymbol{\Phi}_1(s, t_k)$ and $\boldsymbol{\Phi}_2(s, t_k)$ denote the solution to the ODEs

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\boldsymbol{\Phi}_1(s, t_k) &= (\mathbf{A}(s) + \mathbf{B}(s)\mathbf{K}(s))\boldsymbol{\Phi}_1(s, t) \\ \frac{\mathrm{d}}{\mathrm{d}s}\boldsymbol{\Phi}_2(s, t_k) &= \mathbf{A}(s)\boldsymbol{\Phi}_2(s, t) + \mathbf{B}(s)\mathtt{K}_k(t_k).\end{aligned} \tag{G.8}$$

with initial conditions $\boldsymbol{\Phi}_1(t_k, t_k) = \boldsymbol{\Phi}_2(t_k, t_k) = \mathbf{I}$. We define

$$\mathtt{X}_k^{\mathrm{ct}} := \boldsymbol{\Phi}_1(t_{k+1}, t_k), \quad \mathtt{X}_k^{\mathrm{sub}} := \boldsymbol{\Phi}_2(t_{k+1}, t_k)$$

Definition G.4 (Relevant Cost Matrices)**.** For $k \in [K]$, define

$$\begin{aligned}\mathtt{Y}_k^{\mathrm{ct}} &:= \int_{s=t_k}^{t_k} \boldsymbol{\Phi}_1(s, t_k)^{\top}(\mathbf{I} + \mathbf{K}(s)^{\top}\mathbf{K}(s))\boldsymbol{\Phi}_1(s, t_k)\mathrm{d}s \\ \mathtt{Y}_k^{\mathrm{sub}} &:= \tau(\mathbf{I} + \mathbf{K}(t_k)^{\top}\mathbf{K}(t_k))\end{aligned}$$

Lemma G.5. *The cost-to-go matrices* $\mathtt{P}_k^{\mathrm{ct}}$ *and* $\mathtt{P}_k^{\mathrm{sub}}$ *are given by the following Lyapunov recursions, with initial conditions* $\mathtt{P}_{K+1}^{\mathrm{ct}} = \mathtt{P}_{K+1}^{\mathrm{sub}} = \mathbf{I}$*:*

$$\mathtt{P}_k^{\mathrm{ct}} = (\mathtt{X}_k^{\mathrm{ct}})^\top \mathtt{P}_{k+1}^{\mathrm{ct}} \mathtt{X}_k^{\mathrm{ct}} + \mathtt{Y}_k^{\mathrm{ct}}$$
$$\mathtt{P}_k^{\mathrm{sub}} = (\mathtt{X}_k^{\mathrm{sub}})^\top \mathtt{P}_{k+1}^{\mathrm{sub}} \mathtt{X}_k^{\mathrm{sub}} + \mathtt{Y}_k^{\mathrm{sub}}$$

Proof of Lemma G.5. The recursion for $\mathtt{P}_k^{\mathrm{sub}}$ is directly from Definition G.1, and the fact that $\mathtt{X}_k^{\mathrm{sub}} = \mathtt{A}_{\mathrm{cl},k}^{\pi}$ due to Lemma C.10. To verify the recursion for $\mathtt{P}_k^{\mathrm{ct}}$, we note that we can express $\mathbf{P}(t) = \mathbf{P}_{\mathrm{opt}}^{\pi}(t)$ in Definition G.2 as satisfying the following ODE (see Anderson & Moore (2007, Section 2.3)):

$$\mathbf{P}(T) = \mathbf{I}, \quad -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) = (\mathbf{A}(t) + \mathbf{B}(t)\mathbf{K}(t))^\top \mathbf{P}(t)(\mathbf{A}(t) + \mathbf{B}(t)\mathbf{K}(t)) + \mathbf{I} + \mathbf{K}(t)^\top \mathbf{K}(t)$$

It can be checked then by computing derivatives and using existence and uniqueness of ODEs that

$$\mathbf{P}(s,t) = \mathbf{\Phi}_1(s,t)^\top \mathbf{P}(s)\mathbf{\Phi}_1(s,t) + \int_{s'=t}^{s} \mathbf{\Phi}_1(s',t)^\top (\mathbf{I} + \mathbf{K}(s')^\top \mathbf{K}(s'))\mathbf{\Phi}(s',t)\mathrm{d}s'$$

Specializing to $s = t_{k+1}$ and $t = t_k$ verifies the desired recursion. □

As $\mathtt{P}_k^{\mathrm{ct}} = \mathbf{P}(t_k)$, the terms $\mathtt{P}_k^{\mathrm{ct}}$ are bounded whenever $\mathbf{P}(\cdot)$ is. Therefore, we use a Lyapunov perturbation bound to bound $\mathtt{P}_k^{\mathrm{sub}}$ in terms of $\mathtt{P}_k^{\mathrm{ct}}$. This requires reasoning about the differences $\mathtt{X}_k^{\mathrm{ct}} - \mathtt{X}_k^{\mathrm{sub}}$ and $\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}$, which we do in just below.

G.2.2. Controlling the rate of change of $\mathbf{K}(t)$.

Our first step in controlling the perturbation term is to argue that the optimal controller $\mathbf{K}(t)$ does not change too rapidly. As $\mathbf{K}(t) = \mathbf{B}(t)^\top \mathbf{P}(t)$, we begin by bounding the change in $\mathbf{B}(t)$.

Claim G.1 (Change in $\mathbf{B}(t)$)**.** $\mathbf{B}(t)$ *is differentiable in* t *on for* $t \in \mathrm{int}(\mathcal{I}_k)$*, and satisfies* $\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t)\| \le M_f \kappa_f$

Proof. Recall that $\mathbf{B}(t) = \partial_u f(\mathbf{x}(t), \mathbf{u}(t))$. For $t \in \mathrm{int}(\mathcal{I}_k)$, $\mathbf{u}(t)$ is constant, and $\mathbf{x}(t)$, being the solution to an ODE, is also t-differentiable. We now bound $\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t)\|$. We have

$$\begin{aligned}
\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t)\| &= \|\frac{\mathrm{d}}{\mathrm{d}t}\partial_u f(\mathbf{x}(t), \mathbf{u}(t))\| \\
&= \|\partial_{uu} f(\mathbf{x}(t), \mathbf{u}(t))\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}(t) + \partial_{xu} f(\mathbf{x}(t), \mathbf{u}(t))\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t)\| \\
&\le M_f \|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t)\| \le M_f \kappa_f
\end{aligned}$$

where the second-to-last inequality is the limiting consequence holds from Assumption 4.1, and where the term $\partial_{uu} f(\mathbf{x}(t), \mathbf{u}(t))\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{u}(t)$ vanishes vanishes because $\mathbf{u}(t) = \mathbf{u}^\pi(t)$ is constant on $\mathcal{I}_k$. □

Next, we bound the change in $\mathbf{P}(t)$:

Claim G.2 (Change in $\mathbf{P}(t)$)**.** $\mathbf{P}(t)$ *is differentiable in* t*, and* $\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t)\| \le (L_{\mathrm{cl}}/L_f)^2$.

Proof. Note that $\mathbf{P}(t)$ is given by the ODE

$$\mathbf{P}(T) = \mathbf{I}, \quad -\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t) = \mathbf{A}(t)^\top \mathbf{P}(t) + \mathbf{P}(t)\mathbf{A}(t) - \mathbf{P}(t)\mathbf{B}(t)\mathbf{B}(t)^\top \mathbf{P}(t) + \mathbf{I},$$

which ensures differentiability. Thus, as $\|\mathbf{A}(t)\| \vee \|\mathbf{A}(t)\| \vee 1 \le L_f$ by Assumption 4.1 and Assumption 4.3,

$$\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t)\| \le 1 + 2L_f\|\mathbf{P}(t)\| + L_f^2\|\mathbf{P}(t)\|^2 \le (1 + 2L_f\mu_{\mathrm{ric}} + L_f^2\mu_{\mathrm{ric}}^2) \le (1 + L_f\mu_{\mathrm{ric}})^2,$$

which is precisely $(L_{\mathrm{cl}}/L_f)^2$. □

We now establish a bound on the change in $\mathbf{K}(t)$.

Claim G.3 (Continuity of Optimal Controller)**.** *For all $t \in \mathcal{I}_k$,*

$$\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{K}(t)\| \le M_f \kappa_f \mu_{\mathrm{ric}} + L_f^{-1} L_{\mathrm{cl}}^2.$$

Proof of Claim G.3. By Claims G.1 and G.2, we have

$$\begin{aligned}\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{K}(t)\| &\le \|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t)\|\|\mathbf{P}(t)\| + \|\mathbf{B}(t)\|\|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{P}(t)\| \\ &\le \|\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t)\|\mu_{\mathrm{ric}} + L_f(L_{\mathrm{cl}}/L_f)^2 \\ &\le M_f\kappa_f\mu_{\mathrm{ric}} + L_f^{-1}L_{\mathrm{cl}}^2.\end{aligned}$$

□

By integrating, we arrive at the next claim.

Claim G.4. *The following bound holds*

$$\sup_{s\in\mathcal{I}_k}\|\mathbf{K}(s) - \mathbf{K}(t_k)\| \le \tau\left(M_f\kappa_f\mu_{\mathrm{ric}} + L_f^{-1}L_{\mathrm{cl}}^2\right).$$

Proof of Claim G.4. Directly from Claim G.3. □

G.2.3. Controlling differences in $\|\mathtt{x}_k^{\mathrm{ct}} - \mathtt{x}_k^{\mathrm{sub}}\|$ and $\|\mathtt{y}_k^{\mathrm{ct}} - \mathtt{y}_k^{\mathrm{sub}}\|$

We first state a bound on the magnitudes of various quantities of interest.

Claim G.5. $\|\mathbf{K}(t)\| \le \mu_{\mathrm{ric}} L_f$ *and* $\|\mathbf{A}(t) + \mathbf{B}(t)\mathbf{K}(t)\| \le L_{\mathrm{cl}}$*, where we recall* $L_{\mathrm{cl}} := L_f(1 + L_f\mu_{\mathrm{ric}})$.

Proof. Recall that $\mathbf{K}(t) = \mathbf{B}(t)^\top \mathbf{P}(t)$. From Assumption 4.3, $\|\mathbf{P}(t)\| \le \mu_{\mathrm{ric}}$, and $\|\mathbf{B}(t)\| \le L_f$ by Assumption 4.1, which gives $\|\mathbf{K}(t)\| \le \mu_{\mathrm{ric}} L_f$. Bounding $\|\mathbf{A}(t)\| \vee \|\mathbf{B}(t)\|$ by L_f (again, invoking Assumption 4.1), concludes the demonstration. □

Next, we show that $\mathbf{\Phi}_1(s, t_k)$ is close to the identity for sufficiently small τ.

Claim G.6. *Suppose that $\tau L_{\mathrm{cl}} \le 1/2$. Then,*

$$\|\mathbf{I} - \mathbf{\Phi}_1(s, t_k)\| \le \tau L_{\mathrm{cl}} \exp(1/2) \le \min\{1, 2\tau L_{\mathrm{cl}}\}$$

Proof of Claim G.6. It suffices to bound, for all $\xi \in \mathbb{R}^{d_{\mathtt{x}}} : \|\xi\| = 1$ the differences $\|\mathbf{y}_1(s) - \xi\|$ where $\mathbf{y}_1 = \Phi_1(s, t_k)\xi$. We do this via Picard's lemma.

Specifically, write $\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{y}_1(s) = \tilde{f}(\mathbf{y}_1(s), s)$, where $\tilde{f}(y, s) = (\mathbf{A}(s) + \mathbf{B}(s)\mathbf{K}(s))y$, and $\mathbf{z}(s) = \xi$. As $\tilde{f}(y, s)$ is $\sup_{s\in\mathcal{I}_k}\|\mathbf{A}(s) + \mathbf{B}(s)\mathbf{K}(s)\| \le L_{\mathrm{cl}}$ Lipchitz is y (here, we use Claim G.5) and as $\frac{\mathrm{d}}{\mathrm{d}s}\xi = 0$, and the Picard Lemma (Lemma C.9) gives

$$\begin{aligned}\|\xi - \mathbf{y}_1(s)\| &\le \exp((s - t_k)(2L_f^2\mu_{\mathrm{ric}}))\int_{s'=t_k}^{s}\|(\mathbf{A}(s') + \mathbf{B}(s')\mathbf{K}(s'))\xi\|\mathrm{d}s' \\ &\le \exp((s - t_k)L_{\mathrm{cl}})\int_{s'=t_k}^{s}\|(\mathbf{A}(s') + \mathbf{B}(s')\mathbf{K}(s'))\|\mathrm{d}s' && (\|\xi\| \le 1) \\ &\le \exp((s - t_k)L_{\mathrm{cl}}) \cdot (s - t_k)L_{\mathrm{cl}}, \\ &\le \exp(\tau L_{\mathrm{cl}}) \cdot \tau L_{\mathrm{cl}}, \\ &\le \exp(1/2)\tau L_{\mathrm{cl}}\end{aligned}$$

where we assume $\tau L_{\mathrm{cl}} \le 1/2$. □

We can now bound the differences between $\|\mathtt{X}_k^{\mathrm{ct}} - \mathtt{X}_k^{\mathrm{sub}}\| = \|\mathbf{\Phi}_2(t_{k+1}, t_k) - \mathbf{\Phi}_1(t_{k+1}, t_k)\|$.

Lemma G.6. *For $k \in [K]$ and $s \in \mathcal{I}_k$, let $\mathbf{\Phi}_1(s, t_k)$ and $\mathbf{\Phi}_2(s, t_k)$ denote the solution to the ODEs*

$$\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{\Phi}_1(s, t_k) = (\mathbf{A}(s) + \mathbf{B}(s)\mathbf{K}(s))\mathbf{\Phi}_1(s, t), \quad \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{\Phi}_2(s, t_k) = \mathbf{A}(s)\mathbf{\Phi}_2(s, t) + \mathbf{B}(s)\mathtt{K}_k(t_k).$$

with initial conditions $\mathbf{\Phi}_1(t_k, t_k) = \mathbf{\Phi}_2(t_k, t_k) = \mathbf{I}$. Then, if $\tau L_{\mathrm{cl}} \le 1/2$,

$$\|\mathtt{X}_k^{\mathrm{ct}} - \mathtt{X}_k^{\mathrm{sub}}\| = \|\mathbf{\Phi}_2(t_{k+1}, t_k) - \mathbf{\Phi}_1(t_{k+1}, t_k)\| \le 2\tau^2 \left(L_f \mu_{\mathrm{ric}} M_f \kappa_f + 3L_{\mathrm{cl}}^2\right).$$

Proof. It suffices to bound, for all initial conditions, $\xi \in \mathbb{R}^{d_{\mathsf{x}}}$ with $\|\xi\| = 1$, the solutions $\mathbf{y}_i(s) = \mathbf{\Phi}_i(s)\xi$. We apply the Picard Lemma, with $\mathbf{z}(s) \leftarrow \mathbf{y}_1(s)$, and express $\mathbf{y}_2(s) = \tilde{f}(\mathbf{y}_2(s), s)$, where $\tilde{f}(y, s) = \mathbf{A}(s)y + \mathbf{B}(s)\mathbf{K}(t)$. As $\|\mathbf{A}(s)\| \le L_f$, the Picard Lemma (Lemma C.9) yields

$$\begin{aligned}
\|\mathbf{y}_1(t_{k+1}) - \mathbf{y}_2(t_{k+1})\| &\le \exp(L_f(t-s)) \int_{s=t}^{t_{k+1}} \|\mathbf{A}(s)\mathbf{y}_1(s) + \mathbf{B}(s)\mathbf{K}(t_k)\xi - \frac{\mathrm{d}}{\mathrm{d}s}\mathbf{y}_1(s)\| \mathrm{d}s \\
&\le \exp(L_f\tau) \int_{s=t}^{t_{k+1}} \|\mathbf{A}(s)\mathbf{y}_1(s) + \mathbf{B}(s)\mathbf{K}(t_k)\xi - (\mathbf{A}(s) + \mathbf{B}(s)\mathbf{K}(s))\mathbf{y}_1(s)\| \mathrm{d}s \\
&\le \exp(L_f\tau) \int_{s=t}^{t_{k+1}} \|\mathbf{B}(s)(\mathbf{K}(s)\mathbf{y}_1(s) - \mathbf{K}(t_k)\xi)\| \mathrm{d}s \\
&\le L_f \exp(L_f\tau) \int_{s=t}^{t_{k+1}} \|\mathbf{K}(s)\mathbf{y}_1(s) - \mathbf{K}(t_k)\xi\| \mathrm{d}s \\
&\le L_f \exp(L_f\tau) \int_{s=t}^{t_{k+1}} (\|(\mathbf{K}(s) - \mathbf{K}(t_k))\xi\| + \|\mathbf{K}(s)(\xi - \mathbf{y}_1(s))\|) \mathrm{d}s \\
&\le L_f \exp(L_f\tau) \int_{s=t}^{t_{k+1}} (\|\mathbf{K}(s) - \mathbf{K}(t_k)\| + L_f\mu_{\mathrm{ric}}\|\xi - \mathbf{y}_1(s)\|) \mathrm{d}s \\
&\le \exp(1/2) L_f\tau \max_{s \in \mathcal{I}_k} (\|\mathbf{K}(s) - \mathbf{K}(t_k)\| + L_f\mu_{\mathrm{ric}}\|\xi - \mathbf{y}_1(s)\|)
\end{aligned}$$

where the second-to-last line uses $\|\xi\| = 1$ and $\|\mathbf{K}(s)\| \le L_u\mu_{\mathrm{ric}}$, and the last uses $\tau \le 1/2L_x$ and well as a bound of an integral by a maximum. By claims Claims G.4 and G.6,

$$\begin{aligned}
&\max_{s \in \mathcal{I}_k} (\|\mathbf{K}(s) - \mathbf{K}(t_k)\| + L_f\mu_{\mathrm{ric}}\|\xi - \mathbf{y}_1(s)\|) \\
&\le \tau \left(M_f\kappa_f\mu_{\mathrm{ric}} + L_f^{-1}L_{\mathrm{cl}}^2\right) + L_f\mu_{\mathrm{ric}} \exp(1/2)\tau L_{\mathrm{cl}} \\
&= \tau \left(\mu_{\mathrm{ric}}M_f\kappa_f + L_{\mathrm{cl}}(L_f\mu_{\mathrm{ric}} \exp(1/2) + L_f^{-1}L_{\mathrm{cl}})\right) \\
&\le \tau \left(\mu_{\mathrm{ric}}M_f\kappa_f + 3L_f^{-1}L_{\mathrm{cl}}^2\right),
\end{aligned}$$

where in the last inequality we use $\exp(1/2) \le 2$ and $L_f\mu_{\mathrm{ric}} \le (1 + L_f\mu_{\mathrm{ric}}) = L_{\mathrm{cl}}L_f^{-1}$. Therefore, again using $\exp(1/2) \le 2$,

$$\|\mathbf{y}_1(t_{k+1}) - \mathbf{y}_2(t_{k+1})\| \le 2\tau^2 \left(L_f\mu_{\mathrm{ric}}M_f\kappa_f + 3L_{\mathrm{cl}}^2\right).$$

Quantifying over all unit-norm initial conditions ξ concludes the proof. □

We now establish a qualitatively similar bound on $\tau\|\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}\|$.

Lemma G.7. $\|\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}\| \le 2\tau^2 \left(M_f\kappa_f\mu_{\mathrm{ric}}^2L_f + 7\mu_{\mathrm{ric}}L_{\mathrm{cl}}^2\right)$.

Proof. Recall the definitions

$$\begin{aligned}
\mathtt{Y}_k^{\mathrm{ct}} &:= \int_{s=t_k}^{t_k} \mathbf{\Phi}_1(s, t_k)^\top (\mathbf{I} + \mathbf{K}(s)^\top \mathbf{K}(s)) \mathbf{\Phi}_1(s, t_k) \mathrm{d}s \\
\mathtt{Y}_k^{\mathrm{sub}} &:= \tau(\mathbf{I} + \mathbf{K}(t_k)^\top \mathbf{K}(t_k))
\end{aligned}$$

We can then express

$$\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}} = \mathtt{Y}_k^{\mathrm{ct}} - \tau(\mathbf{I} + \mathbf{K}(t_k)\mathbf{K}(t_k)^\top) = \int_{s=t_k}^{t_k} \mathbf{Z}_k(s),$$
$$\mathbf{Z}_k(s) := \left\{\mathbf{\Phi}_1(s,t_k)^\top(\mathbf{I} + \mathbf{K}(s)^\top\mathbf{K}(s))\mathbf{\Phi}_1(s,t_k) - (\mathbf{I} + \mathbf{K}(t_k)^\top\mathbf{K}(t_k))\right\} \mathrm{d}s$$

Thus,

$$\|\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}\| \le \tau \max_{s\in\mathcal{I}_k} \|\mathbf{Z}_k(s)\|. \tag{G.9}$$

With numerous applications of the triangle inequality,

$$\begin{aligned}\|\mathbf{Z}_k(s)\| \le{}& \|\mathbf{I} - \mathbf{\Phi}_1(s,t_k)\|\|\mathbf{I} + \mathbf{K}(s)^\top\mathbf{K}(s)\|\|\mathbf{\Phi}_1(s,t_k)\| \\ &+ \|\mathbf{I} + \mathbf{K}(s)^\top\mathbf{K}(s)\|\|\mathbf{I} - \mathbf{\Phi}_1(s,t_k)\| + \|\mathbf{K}(s) - \mathbf{K}(t_k)\|(\|\mathbf{K}(s)\| + \|\mathbf{K}(t_k)\|).\end{aligned}$$

Using $\|\mathbf{K}(s)\| \vee \|\mathbf{K}(t_k)\| \le L_f\mu_{\mathrm{ric}}$ due to Claim G.3, we have

$$\begin{aligned}\|\mathbf{Z}_k(s)\| &\le (1 + L_f^2\mu_{\mathrm{ric}}^2)(1 + \|\mathbf{\Phi}_1(s,t_k)\|)\|\mathbf{I} - \mathbf{\Phi}_1(s,t_k)\| + 2\mu_{\mathrm{ric}}L_f\|\mathbf{K}(s) - \mathbf{K}(t_k)\| \\ &\le 3(1 + L_f^2\mu_{\mathrm{ric}}^2)\|\mathbf{I} - \mathbf{\Phi}_1(s,t_k)\| + 2\mu_{\mathrm{ric}}L_f\|\mathbf{K}(s) - \mathbf{K}(t_k)\| && \text{(Claim G.6)} \\ &\le 6\tau L_{\mathrm{cl}}(1 + L_f^2\mu_{\mathrm{ric}}^2) + 2\mu_{\mathrm{ric}}L_f\|\mathbf{K}(s) - \mathbf{K}(t_k)\| && \text{(Claim G.6)} \\ &\le 12\tau L_f^2\mu_{\mathrm{ric}}(1 + L_f^2\mu_{\mathrm{ric}}^2) + 2\tau\mu_{\mathrm{ric}}L_f\left(M_f\kappa_f\mu_{\mathrm{ric}} + L_f^{-1}L_{\mathrm{cl}}^2\right). && \text{(Claim G.4)}\end{aligned}$$

We can upper bound $L_f^2(1 + L_f^2\mu_{\mathrm{ric}}^2) \le L_f^2(1 + L_f\mu_{\mathrm{ric}})^2 = L_{\mathrm{cl}}^2$, and simplify $2\tau\mu_{\mathrm{ric}}L_f\left(M_f\kappa_f\mu_{\mathrm{ric}} + L_f^{-1}L_{\mathrm{cl}}^2\right) = 2\tau\left(M_f\kappa_f\mu_{\mathrm{ric}}^2L_f + \mu_{\mathrm{ric}}L_{\mathrm{cl}}^2\right)$. This gives

$$\|\mathbf{Z}_k(s)\| \le 12\tau\mu_{\mathrm{ric}}L_{\mathrm{cl}}^2 + 2\tau\left(M_f\kappa_f\mu_{\mathrm{ric}}^2L_f + \mu_{\mathrm{ric}}L_{\mathrm{cl}}^2\right) = 2\tau\left(M_f\kappa_f\mu_{\mathrm{ric}}^2L_f + 7\mu_{\mathrm{ric}}L_{\mathrm{cl}}^2\right).$$

Plugging the above bound into Eq. (G.9) concludes. □

G.2.4. Concluding the proof of Proposition G.2

From Lemmas G.6 and G.7, we have

$$\|\mathtt{X}_k^{\mathrm{ct}} - \mathtt{X}_k^{\mathrm{sub}}\| \le 2\tau^2\left(L_f\mu_{\mathrm{ric}}M_f\kappa_f + 3L_{\mathrm{cl}}^2\right), \quad \|\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}\| \le 2\tau^2\mu_{\mathrm{ric}}\left(M_f\kappa_f\mu_{\mathrm{ric}}L_f + 7L_{\mathrm{cl}}^2\right)$$

Therefore, using $\mu_{\mathrm{ric}} \ge 1$

$$2\|\mathtt{X}_k^{\mathrm{ct}} - \mathtt{X}_k^{\mathrm{sub}}\| + \|\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}\| \le 2\tau^2\mu_{\mathrm{ric}}\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_{\mathrm{cl}}^2\right). \tag{G.10}$$

Now, we invoke Proposition F.12. We construct linear interpolation (here, $s \in [0,1]$ parametrizes the interpolation and not time)

$$\mathtt{X}_k(s) = (1-s)\mathtt{X}_k^{\mathrm{ct}} + s\mathtt{X}_k^{\mathrm{sub}}, \quad \mathtt{Y}_k(s) = (1-s)\mathtt{Y}_k^{\mathrm{ct}} + s\mathtt{Y}_k^{\mathrm{sub}}.$$

Then, by Lemma G.5, the interpolator $\mathtt{\Lambda}_k(s)$ defined in Proposition F.12 satisfies $\mathtt{\Lambda}_k(0) = \mathtt{P}_k^{\mathrm{ct}}$ and $\mathtt{\Lambda}_k(1) = \mathtt{P}_k^{\mathrm{sub}}$. In particular,

$$\begin{aligned}\|\mathtt{\Lambda}_{1:K+1}(0)\|_{\max,\mathrm{op}} &= \max_{k\in[K+1]} \|\mathtt{P}_k^{\mathrm{ct}}\| && \text{(since } \mathtt{\Lambda}_k = \mathtt{P}_k^{\mathrm{ct}} \text{ and definition of } \|\cdot\|_{\max,\mathrm{op}}) \\ &= \max_{k\in[K+1]} \|\mathbf{P}(t_k)\| && \text{(by Eq. (G.7))} \\ &\le \sup_{t\in[T]} \|\mathbf{P}(t)\| \\ &\le \mu_{\mathrm{ric}}, && \text{(G.11)}\end{aligned}$$

where the last inequality is by Assumption 4.3. Moreover, the term $\tilde{\Delta}(s)$ defined in Eq. (F.14) satisfies

$$\begin{aligned}\forall s \in [0,1], \tilde{\Delta}(s) &= \tau^{-1}(2\|\mathtt{X}_k^{\mathrm{ct}} - \mathtt{X}_k^{\mathrm{sub}}\| + \tau\|\mathtt{Y}_k^{\mathrm{ct}} - \mathtt{Y}_k^{\mathrm{sub}}\| \\ &\le \tau \cdot 2\mu_{\mathrm{ric}}\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_{\mathrm{cl}}^2\right). \qquad \text{(by Eq. (G.10))}\end{aligned}$$

Hence, recalling $\|\mathtt{\Lambda}_{1:K+1}(0)\|_{\mathrm{max,op}} \le \mu_{\mathrm{ric}}$ due to Eq. (G.11), it holds that long as $2\tau\mu_{\mathrm{ric}}^2\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_{\mathrm{cl}}^2\right) \le \frac{1}{2}$, it holds that

$$\max_{k\in[K+1]} \|\mathtt{P}_k^{\mathrm{sub}}\| = \|\mathtt{\Lambda}_{1:K+1}(1)\|_{\mathrm{max,op}} \le 2\|\mathtt{\Lambda}_{1:K+1}(0)\|_{\mathrm{max,op}} \le 2\mu_{\mathrm{ric}}.$$

Lastly, we note the condition $2\tau\mu_{\mathrm{ric}}^2\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_{\mathrm{cl}}^2\right) \le \frac{1}{2}$ is equivalent to

$$\begin{aligned}\tau &\le \frac{1}{4\mu_{\mathrm{ric}}^2\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_{\mathrm{cl}}^2\right)} \\ &\le \frac{1}{4\mu_{\mathrm{ric}}^2\left(3M_f\kappa_f\mu_{\mathrm{ric}}L_f + 13L_f^2(1+L_f\mu_{\mathrm{ric}})^2\right)} := \tau_{\mathrm{ric}} \qquad \text{(Definition of } L_{\mathrm{cl}} \text{ in Eq. (G.6))}\end{aligned}$$

This concludes the proof of Proposition G.2. □

G.3. Proof of Lemma G.3

The proof is similar to Proposition G.2. Let $\pi = \pi^{(1)}$. As $\mathtt{K}_k^\pi = 0$ for all k, we that $L_\pi = 1$, and that $\mathtt{\Lambda}_{K+1}^\pi = \mathbf{I}$, and

$$\mathtt{\Lambda}_k^\pi = (\mathtt{A}_{\mathrm{ol},k}^\pi)^\top \mathtt{\Lambda}_{k+1}^\pi \mathtt{A}_{\mathrm{ol},k}^\pi + \tau\mathbf{I} \tag{G.12}$$

On the other hand, following the arguments of Proposition G.2, we it can be shown that $V^\pi(t_k; \mathbf{u} = 0, \xi) = \xi^\top \mathtt{P}_k^{\mathrm{ct}} \xi$, where $\mathtt{P}_k^{\mathrm{ct}}$ satisfies the recursion $\mathtt{P}_{K+1}^{\mathrm{ct}} = \mathbf{I}$ and

$$\mathtt{P}_k^{\mathrm{ct}} = \mathbf{\Phi}_1(t_{k+1}, t_k)^\top \mathtt{P}_{k+1}^{\mathrm{ct}} \mathbf{\Phi}_1(t_{k+1}, t_k) + \mathtt{Y}_k^{\mathrm{ct}}, \quad \mathtt{Y}_k^{\mathrm{ct}} := \int_{s=t_k}^{t_k} \mathbf{\Phi}_1(s, t_k)^\top \mathbf{\Phi}_1(s, t_k)\mathrm{d}s,$$

where $\Phi_1(s,s) = \mathbf{I}$ and where (using that we consider $V^\pi(t_k; \mathbf{u} = 0, \xi)$ with $\mathbf{u} = 0$, so the corresponding $\mathbf{K}(t)$ in Eq. (G.8) vanishes)

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{\Phi}_1(t,s) = \mathbf{A}_{\mathrm{ol}}^\pi(s)\mathbf{\Phi}_1(t,s).$$

Hence, $\mathbf{\Phi}_1(t_{k+1}, t_k) = \mathtt{A}_{\mathrm{ol},k}^\pi$, so that

$$\mathtt{P}_k^{\mathrm{ct}} = (\mathtt{A}_{\mathrm{ol},k}^\pi)^\top \mathtt{P}_{k+1}^{\mathrm{ct}} \mathtt{A}_{\mathrm{ol},k}^\pi + \mathtt{Y}_k^{\mathrm{ct}}, \quad \mathtt{Y}_k^{\mathrm{ct}} := \int_{s=t_k}^{t_k} \mathbf{\Phi}_1(s, t_k)^\top \mathbf{\Phi}_1(s, t_k)\mathrm{d}s$$

Along the lines of Lemma G.7, it can be shown that for $\tau \le \tau_{\mathrm{ric}}$, $\mathbf{\Phi}_1(s,t_k)^\top \mathbf{\Phi}_1(s,t_k) \succeq \frac{1}{2}\mathbf{I}$ for all $t \in \mathcal{I}_k$. Thus,

$$\mathtt{P}_k^{\mathrm{ct}} \succeq (\mathtt{A}_{\mathrm{ol},k}^\pi)^\top \mathtt{P}_{k+1}^{\mathrm{ct}} \mathtt{A}_{\mathrm{ol},k}^\pi + \frac{1}{2}\mathbf{I}.$$

Comparing to Eq. (G.12), we find that $\mathtt{P}_k^{\mathrm{ct}} \succeq \frac{1}{2}\mathtt{\Lambda}_k^\pi$. As $\|\mathtt{P}_k^{\mathrm{ct}}\| = \sup_{\xi:\|\xi\|=1} V^\pi(t_k; \mathbf{u} = 0, \xi) \le \mu_{\mathrm{ric}}$, we conclude $\|\mathtt{\Lambda}_k^\pi\| \le 2\mu_{\mathrm{ric}}$, as neeeded.

G.4. Proof of Lemma A.1

We recall the definitions $\kappa_{\pi,\infty} := \max_{1\le j\le k\le K+1} \|\mathtt{\Phi}_{\mathrm{cl},k,j}^\pi\|$, and

$$\begin{aligned}\kappa_{\pi,1} &:= \max_{k\in[K+1]} \tau\left(\sum_{j=1}^{k} \|\mathtt{\Phi}_{\mathrm{cl},k,j}^\pi\| \vee \sum_{j=k}^{K+1} \|\mathtt{\Phi}_{\mathrm{cl},j,k}^\pi\|\right) \\ \kappa_{\pi,2}^2 &:= \max_{k\in[K+1]} \tau\left(\sum_{j=1}^{k} \|\mathtt{\Phi}_{\mathrm{cl},k,j}^\pi\|^2 \vee \sum_{j=k}^{K+1} \|\mathtt{\Phi}_{\mathrm{cl},j,k}^\pi\|^2\right),\end{aligned}$$

and recall the definitions $\Lambda^\pi_{K+1} = \mathbf{I}$, and $\Lambda^\pi_k = (\mathtt{A}^\pi_{\mathrm{cl},k})^\top \Lambda^\pi_{k+1} \mathtt{A}^\pi_{\mathrm{cl},k} + \tau \mathbf{I}$, and $\mu_{\pi,\star} := \max_{k \in \{k_0, k_0+1, \dots, K+1\}} \|\Lambda^\pi_k\|$.

Let us first bound $\|\Phi^\pi_{\mathrm{cl},k,j}\|$ for $j \geq k_0$.

Claim G.7. *For $j \geq k_0$, and $\tau \leq 1/6L_f L_\pi$, $\|\Phi^\pi_{\mathrm{cl},k,j}\| \leq \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}(1-\tau/\mu_{\pi,\star})^{k-j}}$*

Proof. We apply Lemma F.10 with $\mathbf{X}_k \leftarrow \mathtt{A}^\pi_{\mathrm{ol},k}$, $\mathbf{Q} = \mathbf{I}$, and $\mathbf{Y}_k = 0$, and only take the recursion back to $k = k_0$. That lemma shows that, as long as $\|\mathtt{A}^\pi_{\mathrm{ol},k} - \mathbf{I}\| \leq \kappa\tau$ for some $\kappa \leq 2/\tau$, it holds that (for $j \geq k_0$)

$$\|\Phi^\pi_{\mathrm{cl},k,j}\|^2 \leq \max\{1, 2\kappa\}\mu_{\pi,\star}(1-\tau/\mu_{\pi,\star})^{k-j}.$$

Lemmas I.3 and I.4, and using $L_\pi \geq 1$, we have that for $\tau \leq 1/4L_f$, $\|\mathtt{A}^\pi_{\mathrm{cl},k} - \mathbf{I}\| \leq \|\mathtt{A}^\pi_{\mathrm{ol},k} - \mathbf{I}\| + \|\mathtt{B}^\pi_{\mathrm{ol},k}\mathtt{K}^\pi_k\| \leq \exp(1/4)\tau L_f(1+L_\pi) \leq 2\exp(1/4)\tau L_f L_\pi \leq 3\tau L_f L_\pi$. Hence, for $\tau \leq 1/6L_f L_\pi$, we can take $\kappa := 3L_f L_\pi$ and have $\kappa \leq 1/2\tau$. For this choice of κ, we get

$$\|\Phi^\pi_{\mathrm{cl},k,j}\|^2 \leq \max\{1, 6L_f L_\pi\}\mu_{\pi,\star}(1-\tau/\mu_{\pi,\star})^{k-j}.$$

□

Next, we bound $\|\Phi^\pi_{\mathrm{cl},k,j}\|$ for $k \leq k_0$.

Claim G.8. *For $k \geq k_0$, $\|\Phi^\pi_{\mathrm{cl},k,j}\| \leq \exp(k_0\tau L_f)$.*

Proof. For $j \leq k \leq k_0$, we have $\mathtt{K}^\pi_j = 0$. Hence $\mathtt{A}^\pi_{\mathrm{cl},j} = \mathtt{A}^\pi_{\mathrm{ol},j}$, and from Lemma I.4, we get $\|\mathtt{A}^\pi_{\mathrm{cl},j}\| \leq \exp(\tau L_f)$. Thus $\|\Phi^\pi_{\mathrm{cl},k,j}\| \leq \prod_{j=1}^k \|\mathtt{A}^\pi_{\mathrm{ol},j}\| \leq \exp(k\tau L_f) \leq \exp(k_0\tau L_f)$. □

Finally, we bound we bound $\|\Phi^\pi_{\mathrm{cl},k,j}\|$ for $j \leq k_0$.

Claim G.9. *For $j \leq k_0$, $\|\Phi^\pi_{\mathrm{cl},k,j}\| \leq \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}}\exp(\tau k_0 L_f)$.*

Proof. For $j < k_0$, we have $\|\Phi^\pi_{\mathrm{cl},k,j}\| = \|\Phi^\pi_{\mathrm{cl},k,k_0}\Phi^\pi_{\mathrm{cl},k,j}\| \leq \|\Phi^\pi_{\mathrm{cl},k,k_0}\|\|\Phi^\pi_{\mathrm{cl},k_0,j}\|$. The first term is at most $\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}$ by Claim G.7, and the second term at most $\exp(\tau k_0 L_f)$ by Claim G.8. □

We can now bound all terms of interest. Directly from the dichotmoty in Claim G.7, we have $\kappa_{\pi,\infty} \leq \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}}\exp(\tau k_0 L_f)$. Next, for any k, we can bound via Claims G.7 and G.9

$$\begin{aligned}
\tau\sum_{j=1}^k \|\Phi^\pi_{\mathrm{cl},k,j}\|^2 &\leq \tau\sum_{j=1}^{k_0} \|\Phi^\pi_{\mathrm{cl},k,j}\|^2 + \tau\sum_{j=k_0}^k \|\Phi^\pi_{\mathrm{cl},k,j}\|^2 \\
&\leq \max\{1, 6L_f L_\pi\}\mu_{\pi,\star}\left((\tau k_0)\exp(2\tau k_0 L_f) + \tau\sum_{j=k_0}^k (1-\tau/\mu_{\pi,\star})^{k-k_0}\right) \qquad \text{(Claims G.7 and G.9)} \\
&\leq \max\{1, 6L_f L_\pi\}\mu_{\pi,\star}\left((\tau k_0)\exp(2\tau k_0 L_f) + \tau\underbrace{\sum_{n=0}^\infty (1-\tau/\mu_{\pi,\star})^n}_{=\frac{1}{1-(1-\tau/\mu_{\pi,\star})} = \mu_{\pi,\star}/\tau}\right) \\
&\leq \max\{1, 6L_f L_\pi\}\mu_{\pi,\star}\left((\tau k_0)\exp(2\tau k_0 L_f) + \mu_{\pi,\star}\right).
\end{aligned}$$

and show the same bound for $\tau \sum_{j=k}^{K+1} \|\Phi^{\pi}_{\mathrm{cl},j,k}\|^2$, which yields the desired upper bound on $\kappa_{\pi,2}$. Finally, to bound $\kappa_{\pi,1}$,

$$\begin{aligned}
\tau \sum_{j=1}^{k} \|\Phi^{\pi}_{\mathrm{cl},k,j}\| &\leq \tau \sum_{j=1}^{k_0} \|\Phi^{\pi}_{\mathrm{cl},k,j}\| + \tau \sum_{j=k_0}^{k} \|\Phi^{\pi}_{\mathrm{cl},k,j}\| \\
&\leq \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}} \left((\tau k_0) \exp(\tau k_0 L_f) + \tau \sum_{j=k_0}^{k} \sqrt{(1-\tau/\mu_{\pi,\star})^{k-k_0}} \right) \quad \text{(Claims G.7 and G.9)} \\
&\leq \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}} \left((\tau k_0) \exp(\tau k_0 L_f) + \tau \sum_{n=0}^{\infty} \sqrt{(1-\tau/\mu_{\pi,\star})^{n}} \right) \\
&\overset{(i)}{\leq} \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}} \left((\tau k_0) \exp(\tau k_0 L_f) + 2\tau \sum_{n=0}^{\infty} (1-\tau/\mu_{\pi,\star})^{n} \right) \\
&= \sqrt{\max\{1, 6L_f L_\pi\}\mu_{\pi,\star}} \left((\tau k_0) \exp(\tau k_0 L_f) + 2\mu_{\pi,\star} \right),
\end{aligned}$$

where in (i), we use that $\sum_{n\geq 0} \sqrt{(1-\gamma)^n} = \sum_{n\geq 0} \sqrt{(1-\gamma)^{2n}} + \sqrt{(1-\gamma)^{2n+1}} \leq 2\sum_{n\geq 0} \sqrt{(1-\gamma)^{2n}} = 2\sum_{n\geq 0}(1-\gamma)^n$. One can establish the same bound for $\tau \sum_{j=k}^{K+1} \|\Phi^{\pi}_{\mathrm{cl},j,k}\|$, which gives the desired bound on $\kappa_{\pi,1}$. □

H. Optimization Proofs

H.1. Proof of Descent Lemma (Lemma A.13)

Proof of Lemma A.13. For simplicity, write $\mathrm{Err}_\nabla = \mathrm{Err}_{\nabla,\pi^{(n)}}(\delta)$, $L_{\nabla,\pi,\infty} = L_{\nabla,\pi^{(n)},\infty}$, and $\mathrm{Err}_{\hat{x}} = \mathrm{Err}_{\hat{x}}(\delta)$ and $\mathrm{Err}_\nabla(\delta)$ and $M \geq M_{\mathcal{J},\mathrm{tay},\pi^{(n)}}$. Note that if π and $\tilde{\pi}$ have the same input sequence but possibly different gains, $\mathcal{J}_T^{\mathrm{disc}}(\pi) = \mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi})$. Therefore,

$$\mathcal{J}_T^{\mathrm{disc}}(\pi^{(n+1)}) = \mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi}^{(n+1)}).$$

Define the input

$$\check{\mathtt{u}}_k^{(n)} = \mathtt{u}_k^{(n)} - \eta\hat{\nabla}_k^{(n)} + \mathtt{K}_k^{\pi^{(n)}}(\mathtt{x}_k^{\pi} - \hat{\mathtt{x}}_k)$$

Then, as in Eq. (E.3),

$$\mathtt{u}_k^{\tilde{\pi}^{(n)}} = \mathtt{u}_{\mathrm{orac},k}^{\pi^{(n)}}(\tilde{\mathtt{u}}_{1:K}) = \tilde{\mathtt{u}}_k^{\pi^{(n)}}(\check{\mathtt{u}}_{1:K}), \quad \mathtt{x}_k^{\tilde{\pi}^{(n)}} = \mathtt{x}_{\mathrm{orac},k}^{\pi^{(n)}}(\tilde{\mathtt{u}}_{1:K}) = \tilde{\mathtt{x}}_k^{\pi^{(n)}}(\check{\mathtt{u}}_{1:K}), \quad \forall k \in [K].$$

Consequently, we have the quality

$$\mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi}^{(n)}) := \mathcal{J}_T^{\tilde{\pi}^{(n+1)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\tilde{\pi}^{(n)}}) = \mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\check{\mathtt{u}}_{1:K}) = \mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}} + \delta\check{\mathtt{u}}_k^{(n)}),$$

where we introduced

$$\delta\check{\mathtt{u}}_k^{(n)} := \check{\mathtt{u}}_k^{(n)} - \mathtt{u}_k^{(n)} = \underbrace{-\frac{\eta}{\tau}\hat{\nabla}_k^{(n)}}_{=\delta\check{\mathtt{u}}_k^{(n;1)}} + \underbrace{\mathtt{K}_k^{\pi^{(n)}}(\mathtt{x}_k^{\pi} - \hat{\mathtt{x}}_k)}_{=\delta\check{\mathtt{u}}_k^{(n;2)}}$$

Claim H.1. *We have*

$$\sqrt{\tau}\|\delta\check{\mathtt{u}}_{1:K}^{(n)}\|_{\ell_2} \leq \sqrt{T}(\eta(L_{\nabla,\pi,\infty} + \frac{1}{\tau}\mathrm{Err}_\nabla) + \mathrm{Err}_{\hat{x}})$$
$$\max_k \|\delta\check{\mathtt{u}}_k^{(n)}\| \leq (\eta(L_{\nabla,\pi,\infty} + \frac{1}{\tau}\mathrm{Err}_\nabla) + \mathrm{Err}_{\hat{x}})$$

Proof. The first bound follows from the second. We have that

$$\begin{aligned}
\max_k \|\delta\check{\mathtt{u}}_k^{(n)}\| &\leq \max_k \|\delta\check{\mathtt{u}}_k^{(n;1)}\| + \max_k \|\delta\check{\mathtt{u}}_k^{(n;2)}\| \\
&\leq \frac{\eta}{\tau}\max_k \|\hat{\nabla}_k^{(n)}\| + \mathrm{Err}_{\hat{x}} \\
&\leq \frac{\eta}{\tau}\left(\max_k \|\mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}})\| + \mathrm{Err}_\nabla\right) + \mathrm{Err}_{\hat{x}} \\
&\leq \frac{\eta}{\tau}(\tau L_{\nabla,\pi,\infty} + \mathrm{Err}_\nabla) + \mathrm{Err}_{\hat{x}} && \text{(Lemma A.8)} \\
&= \eta(L_{\nabla,\pi,\infty} + \frac{1}{\tau}\mathrm{Err}_\nabla) + \mathrm{Err}_{\hat{x}}
\end{aligned}$$

□

As a consequence of the above claim, it holds that if

$$(\eta(L_{\nabla,\pi,\infty} + \frac{1}{\tau}\mathrm{Err}_\nabla) + \mathrm{Err}_{\hat{x}}) \leq \min\left\{\frac{R_{\mathrm{feas}}}{8}, B_{\mathrm{stab},\pi}, B_{\mathrm{tay,inf},\pi}, \frac{B_{\mathrm{tay},2,\pi}}{\sqrt{T}}\right\},$$

then (a) Lemma A.7 implies stability of $\tilde{\pi}^{(n)}$:

$$\mu_{\tilde{\pi}^{(n)},\star} \leq 2\mu_{\pi^{(n)},\star}, \quad L_{\tilde{\pi}^{(n)}} = L_{\pi^{(n)}},$$

and (b) the Taylor expansion in Lemma A.6 implies that

$$\begin{aligned}
&\mathcal{J}_T^{\mathrm{disc}}(\tilde{\pi}^{(n)}) = \mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}} + \delta\check{\mathtt{u}}_k^{(n)}) \\
&= \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)}) + \left\langle \delta\check{\mathtt{u}}_{1:K}^{(n)}, \nabla\mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}}) \right\rangle + \frac{R\tau}{2}\|\delta\check{\mathtt{u}}_{1:K}^{(n)}\|_{\ell_2}^2 \\
&\le \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)}) + \sum_{i=1}^{2} \underbrace{\left\langle \delta\check{\mathtt{u}}_{1:K}^{(n;i)}, \nabla\mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}}) \right\rangle + M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;i)}\|_{\ell_2}^2}_{\mathrm{Term}_i}. \qquad \text{(AM-GM)}
\end{aligned}$$

It remains to massage the above display to obtain the descent descent guarantee:

$$\begin{aligned}
\mathrm{Term}_1 &= \left\langle \delta\check{\mathtt{u}}_{1:K}^{(n;1)}, \nabla\mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}}) \right\rangle + M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;1)}\|_{\ell_2}^2 \\
&\le \langle \delta\check{\mathtt{u}}_{1:K}^{(n;1)}, \hat{\nabla}_{1:K}^{(n)} \rangle + \|\delta\check{\mathtt{u}}_k^{(n;1)}\|_{\ell_2}\sqrt{K}\mathrm{Err}_\nabla + M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;1)}\|_{\ell_2}^2 \\
&\le \langle \delta\check{\mathtt{u}}_{1:K}^{(n;1)}, \hat{\nabla}_{1:K}^{(n)} \rangle + \frac{K}{4R\tau}\mathrm{Err}_\nabla^2 + 2M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;1)}\|_{\ell_2}^2 \\
&= (-\frac{\eta}{\tau} + 2M\frac{\eta^2}{\tau})\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + \frac{T}{4R\tau^2}\mathrm{Err}_\nabla^2 \\
&\ge -\frac{\eta}{2\tau}\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + \frac{T}{4M\tau^2}\mathrm{Err}_\nabla^2,
\end{aligned}$$

where the last step uses $\eta \le \frac{1}{4M}$. Then,

$$\begin{aligned}
\mathrm{Term}_2 &= \left\langle \delta\check{\mathtt{u}}_{1:K}^{(n;2)}, \nabla\mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}}) \right\rangle + M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;2)}\|_{\ell_2}^2 \\
&\le \|\delta\check{\mathtt{u}}_{1:K}^{(n;2)}\|_{\ell_2}\|\nabla\mathcal{J}_T^{\pi^{(n)},\mathrm{disc}}(\mathtt{u}_{1:K}^{\pi^{(n)}})\|_{\ell_2} + M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;2)}\|_{\ell_2}^2 \\
&\le \sqrt{\tau}\|\delta\check{\mathtt{u}}_{1:K}^{(n;2)}\|_{\ell_2} \cdot \sqrt{K}\tau L_{\nabla,\pi,\infty} + M\tau\|\delta\check{\mathtt{u}}_{1:K}^{(n;2)}\|_{\ell_2}^2 \qquad \text{(Lemma A.8)} \\
&\le \sqrt{K}\max_k\|\delta\check{\mathtt{u}}_k^{(n;2)}\| \cdot \sqrt{K}\tau L_{\nabla,\pi,\infty} + M\tau K\max_k\|\delta\check{\mathtt{u}}_k\|^2 \\
&= T(\mathrm{Err}_{\hat{x}}L_{\nabla,\pi,\infty} + M\mathrm{Err}_{\hat{x}}^2).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathcal{J}_T^{\mathrm{disc}}(\pi^{(n+1)}) - \mathcal{J}_T^{\mathrm{disc}}(\pi^{(n)}) \\
&\le \mathrm{Term}_1 + \mathrm{Term}_2 \\
&\le -\frac{\eta}{2\tau}\|\hat{\nabla}_{1:K}^{(n)}\|_{\ell_2}^2 + T(\frac{1}{4M\tau^2}\mathrm{Err}_\nabla^2 + \mathrm{Err}_{\hat{x}}L_{\nabla,\pi,\infty} + M\mathrm{Err}_{\hat{x}}^2).
\end{aligned}$$

□

H.2. Proof of Proposition 4.1

The proof of Proposition 4.1 makes liberal use of the definitions of the linearizations given in Appendix C.1, which we recall without further comment. Going forward, introduce the Jacobian linearization of the stabilized cost:

$$\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) := V(\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})) + \int_0^T Q(\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}), \tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}), t)\mathrm{d}t.$$

We now characterize some properties of $\mathcal{J}_T^{\pi,\mathrm{jac}}$.

Lemma H.1 (Valid First-Order Approximation)**.** *We have that* $\nabla_{\bar{\mathbf{u}}}\mathcal{J}_T^{\pi}(\bar{\mathbf{u}}) = \nabla_{\bar{\mathbf{u}}}\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}})$.

Proof. Immediate from the chain rule, and the fact that the Jacobian linearizations are defined as the first-order Taylor expansion of the true dynamics. □

Lemma H.2 (Congruence with the Open-Loop)**.**

$$\inf_{\tilde{\mathbf{u}}} \mathcal{J}_T^{\pi,\mathrm{jac}}(\tilde{\mathbf{u}}) = \inf_{\tilde{\mathbf{u}}} \mathcal{J}_T^{\mathrm{jac}}(\tilde{\mathbf{u}}; \mathbf{u}^\pi).$$

Proof. We prove $\inf_{\tilde{\mathbf{u}}} \mathcal{J}_T^{\pi,\mathrm{jac}}(\tilde{\mathbf{u}}) \le \inf_{\tilde{\mathbf{u}}} \mathcal{J}_T^{\mathrm{jac}}(\tilde{\mathbf{u}}; \mathbf{u}^\pi)$; the converse can be proved similarly. Fix any $\bar{\mathbf{u}}_1 \in \mathcal{U}$. It suffices to exhibit some $\bar{\mathbf{u}}_2 \in \mathcal{U}$ such that, for all $t \in [0, T]$,

$$\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}_2) = \mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}_1; \mathbf{u}^\pi), \quad \tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}_2) = \bar{\mathbf{u}}_1(t).$$

By substracting off $\mathbf{x}^\pi(t)$ and $\mathbf{u}^\pi(t)$, it suffices to show that

$$\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}_2) = \delta\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}_1; \mathbf{u}^\pi), \quad \delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}_2) = \delta\bar{\mathbf{u}}_1(t).$$

It can be directly checked from Lemmas C.1 and C.2 that the input $\bar{\mathbf{u}}_2(t) = \bar{\mathbf{u}}_1(t) - \mathtt{K}^\pi_{k(t)}\delta\mathbf{x}^{\mathrm{jac}}(t \mid \bar{\mathbf{u}}_1; \mathbf{u}^\pi)$ ensures the above display holds. □

The last lemma contains our main technical endeavor, and its proof is defered to Lemma H.3 just below.

Lemma H.3 (Strong Convexity)**.** *Suppose* $\tau \le \min\{\frac{1}{4L_f}, \frac{1}{16L_\pi L_f}\}$. *Then,* $\bar{\mathbf{u}} \mapsto \mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) - \alpha_\pi\|\bar{\mathbf{u}}\|^2_{\mathcal{L}_2(\mathcal{U})}$ *is convex, where* $\alpha_\pi := \frac{\alpha}{64\max\{1, L_\pi^2\}}$.

We may now conclude the proof of our proposition.

Proof of Proposition 4.1. Suppose that π satisfies Definition 4.7, and suppose $\tau \le \min\{\frac{1}{4L_f}, \frac{1}{16L_\pi L_f}\}$ and $\|\nabla_{\bar{\mathbf{u}}}\mathcal{J}_T^\pi(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^\pi}\|_{\mathcal{L}_2(\mathcal{U}))} \le \epsilon_0$. By Lemma H.1, $\|\nabla_{\bar{\mathbf{u}}}\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}})\big|_{\bar{\mathbf{u}}=\mathbf{u}^\pi}\|_{\mathcal{L}_2(\mathcal{U}))} \le \epsilon_0$. By Lemma H.3 and the fact that strong convex functions satisfy the PL-inequality (e.g. Karimi et al. (2016, Theorem 2)), we have

$$\mathcal{J}_T^{\pi,\mathrm{jac}}(\mathbf{u}^\pi) \le \inf_{\bar{\mathbf{u}}} \mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) + \frac{\epsilon_0^2}{\alpha_\pi} = \inf_{\bar{\mathbf{u}}} \mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) + \frac{64\epsilon_0^2\max\{1, L_\pi^2\}}{\alpha}.$$

Finally, by Lemma H.2, $\inf_{\bar{\mathbf{u}}} \mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) = \inf_{\bar{\mathbf{u}}} \mathcal{J}_T^{\mathrm{jac}}(\bar{\mathbf{u}}; \mathbf{u})$, which implies the proposition. □

H.2.1. Proof of Lemma H.3

Proof. We claim that suffices to show the following PSD lower bound:

$$\forall \bar{\mathbf{u}} \in \mathcal{U}, \mathcal{Q}^\pi(\bar{\mathbf{u}}) \ge \frac{\alpha_\pi}{\alpha}\|\bar{\mathbf{u}}\|^2_{\mathcal{L}_2}, \tag{H.1}$$

where we define

$$\mathcal{Q}^\pi(\bar{\mathbf{u}}) := \int_0^T (\|\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})\|^2 + \|\delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})\|^2)\mathrm{d}t,$$

and where we define the deviations $\delta(\cdot)$ as in Lemma C.2.

Claim H.2. *If Eq. (H.1) holds, then Lemma H.3 holds.*

Proof of Claim H.2. Note that $\bar{\mathbf{u}} \mapsto \tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})$ and $\bar{\mathbf{u}} \mapsto \tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})$ are affine, that $\delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) = \tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \mathbf{x}^\pi(t)$ and $\delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) = \tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \mathbf{u}^\pi(t)$ are linear (no affine term), and that the diferences $\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) = \mathbf{x}^\pi(t)$ and $\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) - \delta\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}) = \mathbf{u}^\pi(t)$ are independent of $\bar{\mathbf{u}}$. Hence, we cocnlude $\mathcal{Q}^\pi(\bar{\mathbf{u}})$ is a quadratic function with no linear term, and $\tilde{\mathcal{Q}}^\pi(\bar{\mathbf{u}}) - \mathcal{Q}^\pi(\bar{\mathbf{u}})$ is linear, where we define

$$\tilde{\mathcal{Q}}^\pi(\bar{\mathbf{u}}) := \int_0^T (\|\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})\|^2 + \|\tilde{\mathbf{u}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}})\|^2)\mathrm{d}t,$$

Assumption 2.1 implies that $\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) - \alpha\tilde{\mathcal{Q}}^\pi(\bar{\mathbf{u}})$ is convex, and since the difference $\tilde{\mathcal{Q}}^\pi(\bar{\mathbf{u}}) - \mathcal{Q}^\pi(\bar{\mathbf{u}})$ is linear, that $\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) - \alpha\tilde{\mathcal{Q}}^\pi(\bar{\mathbf{u}})$ is also convex. Lastly, as $\mathcal{Q}^\pi(\bar{\mathbf{u}})$ is quadratic with no linear term, Eq. (H.1) implies $\alpha\mathcal{Q}^\pi(\bar{\mathbf{u}}) - \alpha_\pi\|\bar{\mathbf{u}}\|^2_{\mathcal{L}_2}$ is convex. Thus, $\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) - \alpha_\pi\|\bar{\mathbf{u}}\|^2_{\mathcal{L}_2} = (\mathcal{J}_T^{\pi,\mathrm{jac}}(\bar{\mathbf{u}}) - \alpha\tilde{\mathcal{Q}}^\pi(\bar{\mathbf{u}})) - (\alpha\mathcal{Q}^\pi(\bar{\mathbf{u}}) - \alpha_\pi\|\bar{\mathbf{u}}\|^2_{\mathcal{L}_2})$ is convex. □

To verify Eq. (H.1), let us define a few salient operators. Let $\mathcal{X}$ denote the space of $\mathcal{L}_2$ bounded curves $\mathbf{x}(t) \in \mathbb{R}^{d_\mathbf{x}}$. We define linear operators $\mathbb{T}_1 : \mathcal{U} \to \mathcal{X}$ and $\mathbb{T}_2 : \mathcal{U} \to \mathcal{X}$ and $\mathbb{K} : \mathcal{X} \to \mathcal{U}$ via

$$\mathbb{T}_1[\bar{\mathbf{u}}](t) = \delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t_{k(t)} \mid \bar{\mathbf{u}}), \quad \mathbb{T}_2[\bar{\mathbf{u}}](t) = \delta\tilde{\mathbf{x}}^{\pi,\mathrm{jac}}(t \mid \bar{\mathbf{u}}), \quad \mathbb{K}[\bar{\mathbf{x}}](t) = \mathtt{K}^\pi_{k(t)}\bar{\mathbf{x}}(t).$$

Then, letting $\mathbf{I}_\mathcal{U}$ denote the identity operator of $\mathcal{U}$, we can write

$$\mathcal{Q}^\pi(\bar{\mathbf{u}}) = \|\mathbb{T}_2[\bar{\mathbf{u}}]\|^2_{\mathcal{L}_2(\mathcal{X})} + \|(\mathbf{I}_\mathcal{U} + \mathbb{K}\mathbb{T}_1)[\bar{\mathbf{u}}]\|^2_{\mathcal{L}_2(\mathcal{U})}.$$

Next, we relate $\mathbb{T}_2$ and $\mathbb{T}_1$. Define the operators $\mathbb{L} : \mathcal{X} \to \mathcal{X}$ and $\mathbb{W} : \mathcal{U} \to \mathcal{X}$ by

$$\mathbb{L}[\bar{\mathbf{x}}](t) = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t_{k(t)})\bar{\mathbf{x}}(t), \quad \mathbb{W}[\bar{\mathbf{u}}](t) = \int_{s=t_{k(t)}}^{t} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t,s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\bar{\mathbf{u}}(s)\mathrm{d}s.$$

Then, it can be checked from Lemmas C.2 and C.10 that

$$\mathbb{T}_2[\bar{\mathbf{u}}] = \mathbb{L}\mathbb{T}_1[\bar{\mathbf{u}}] + \mathbb{W}[\bar{\mathbf{u}}].$$

Hence,

$$\mathcal{Q}^\pi(\bar{\mathbf{u}}) = \|(\mathbb{L}\mathbb{T}_1 + \mathbb{W})[\bar{\mathbf{u}}]\|^2_{\mathcal{L}_2(\mathcal{X})} + \|(\mathbf{I}_\mathcal{U} + \mathbb{K}\mathbb{T}_1)[\bar{\mathbf{u}}]\|^2_{\mathcal{L}_2(\mathcal{U})}.$$

With this representation of $\mathcal{Q}^\pi$, we establish a lower bound by applying the following lemma, whose proof we below:

Lemma H.4. *Let $\mathcal{A}, \mathcal{B}$ be Hilbert spaces with norms $\|\cdot\|_\mathcal{A}$ and $\|\cdot\|_\mathcal{B}$, let $\mathbf{I}_\mathcal{A}$ denote the identity operator on $\mathbf{U}$, and let $\mathbb{T}, \mathbb{W} : \mathcal{A} \to \mathcal{B}$, $\mathbb{L} : \mathcal{B} \to \mathcal{B}$, and $\mathbf{K} : \mathcal{B} \to \mathcal{A}$ be linear operators, and let $\|\cdot\|_{\mathrm{op}}$ denote operator norms. Then, if $\|\mathbb{W}\|_{\mathrm{op}} \le \frac{\min\{1, \sigma_{\min}(\mathbb{L})\}}{4\|\mathbb{K}\|}$, it holds for any $\mathbf{a} \in \mathcal{A}$,*

$$\|(\mathbb{L}\mathbb{T} + \mathbb{W})[\mathbf{a}]\|^2_\mathcal{B} + \|(\mathbf{I}_\mathcal{A} + \mathbb{K}\mathbb{T})[\mathbf{a}]\|^2_\mathcal{A} \ge \|\mathbf{a}\|^2 \cdot \frac{\min\{1, \sigma_{\min}(\mathbb{L})^2\}}{16\max\{1, \|\mathbb{K}\|^2_{\mathrm{op}}\}},$$

where $\sigma_{\min}(\mathbb{L}) := \inf_{\mathbf{a}:\|\mathbf{a}\|_\mathcal{A}=1} \|\mathbb{L}\mathbf{a}\|_\mathcal{A}$.

To apply the lemma, we first perform a few computations. Throughout, we use $\|\cdot\|_{\mathrm{op}}$ to denote operator norm, and $\sigma_{\min}$ to denote minimal singular value as an operator, e.g.

$$\|\mathbb{K}\|_{\mathrm{op}} = \sup_{\bar{\mathbf{x}} \neq 0} \frac{\|\mathbb{K}[\bar{\mathbf{x}}]\|_{\mathcal{L}_2(\mathcal{U})}}{\|\bar{\mathbf{x}}\|_{\mathcal{L}_2(\mathcal{X})}}, \sigma_{\min}(\mathbb{L}) = \inf_{\bar{\mathbf{x}} \neq 0} \frac{\|\mathbb{L}[\bar{\mathbf{x}}]\|_{\mathcal{L}_2(\mathcal{U})}}{\|\bar{\mathbf{x}}\|_{\mathcal{L}_2(\mathcal{X})}}.$$

Claim H.3. *Under Definition 4.7, $\|\mathbb{K}\|_{\mathrm{op}} \le L_\pi$.*

Proof. Since $\mathbb{K}$ is a (block-)diagonal operator in t, i.e. $\mathbb{K}[\bar{\mathbf{x}}](t) = \mathtt{K}^\pi_{k(t)}\bar{\mathbf{x}}(t)$, its $\mathcal{L}_2(\mathcal{X}) \to \mathcal{L}_2(\mathcal{U})$ operator is bounded by $\max_k \|\mathtt{K}^\pi_k\|$, which is at most L_π under Definition 4.7. □

Claim H.4. *For $\tau \le L_f/4$, $\sigma_{\min}(\mathbb{L}) \ge 1 - \tau L_f \exp(\tau L_f) \ge \frac{1}{2}$.*

Proof. Again, since $\mathbb{L}$ is a block-diagonal operator in t, i.e. $\mathbb{L}[\bar{\mathbf{x}}](t) = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t_{k(t)})\bar{\mathbf{x}}(t)$, $\sigma_{\min}(\mathbb{L}) = \inf_{t\in[0,T]} \sigma_{\min}(\mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t_{k(t)}))$. By …, $\|\mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t_{k(t)}) - \mathbf{I}\| \le \tau L_f \tau(\tau L_f) \le 1/2$. Thus,$\inf_{t\in[0,T]} \sigma_{\min}(\mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t_{k(t)})) \ge 1/2$. □

Claim H.5. *$\|\mathbb{W}\|_{\mathrm{op}} \le \tau L_f \exp(\tau L_f)$, which is at most $2\tau L_f$ for $\tau \le L_f/4$.*

Proof. For any $\bar{\mathbf{u}}$, we bound

$$\begin{aligned}
\|\mathbb{W}[\bar{\mathbf{u}}]\|_{\mathcal{L}_2(\mathcal{U})}^2 &= \int_{t=0}^{T} \left\| \int_{s=t_{k(t)}}^{t} \mathbf{\Phi}_{\mathrm{ol}}^{\pi}(t,s)\mathbf{B}_{\mathrm{ol}}^{\pi}(s)\bar{\mathbf{u}}(s)\mathrm{d}s \right\|^2 \\
&\le \int_{t=0}^{T} (t - t_{k(t)}) \int_{s=t_{k(t)}}^{t} \|\mathbf{\Phi}_{\mathrm{ol}}^{\pi}(t,s)\mathbf{B}_{\mathrm{ol}}^{\pi}(s)\bar{\mathbf{u}}(s)\mathrm{d}s\|^2 && \text{(Cauchy-Schwartz)} \\
&\le \tau \exp(2\tau L_f) \int_{t=0}^{T} \int_{s=t_{k(t)}}^{t} \|\mathbf{B}_{\mathrm{ol}}^{\pi}(s)\|^2 \|\bar{\mathbf{u}}(s)\mathrm{d}s\|^2 && (t - t_{k(t)} \le \tau, \text{ Lemma I.4}) \\
&\le \tau L_f^2 \exp(2\tau L_f) \int_{t=0}^{T} \int_{s=t_{k(t)}}^{t} \|\bar{\mathbf{u}}(s)\mathrm{d}s\|^2 && \text{(Assumption 4.1)} \\
&= \tau L_f^2 \exp(2\tau L_f) \int_{s=0}^{T} \int_{t=s}^{t_{k+1}(t)} \|\bar{\mathbf{u}}(s)\mathrm{d}s\|^2 \mathrm{d}s \\
&\le \tau^2 L_f^2 \exp(2\tau L_f) \int_{s=0}^{T} \|\bar{\mathbf{u}}(s)\mathrm{d}s\|^2 \mathrm{d}s \\
&= (\tau L_f \exp(\tau L_f))^2 \|\bar{\mathbf{u}}\|_{\mathcal{L}_2(\mathcal{U})}^2.
\end{aligned}$$

□

With the above three claims, Lemma H.4 implies that as long as $\tau \le \min\{\frac{1}{4L_f}, \frac{1}{16L_\pi L_f}\}$,

$$\mathcal{Q}^{\pi}(\bar{\mathbf{u}}) = \|(\mathbb{L}\mathbb{T}_1 + \mathbb{W})[\bar{\mathbf{u}}]\|_{\mathcal{L}_2(\mathcal{X})}^2 + \|(\mathbf{I}_{\mathcal{U}} + \mathbb{K}\mathbb{T}_1)[\bar{\mathbf{u}}]\|_{\mathcal{L}_2(\mathcal{U})}^2 \ge \frac{1}{64\max\{1, L_\pi^2\}} = \frac{\alpha_\pi}{\alpha}.$$

□

Proof of Lemma H.4. Without loss of generality, assume $\|\mathbf{a}\|_{\mathcal{A}} = 1$, and define $\eta = \frac{\min\{1, \sigma_{\min}(\mathbb{L})\}}{2\|\mathbb{K}\|_{\mathrm{op}}}$. We consider two cases: First, if $\|\mathbb{T}\mathbf{a}\| \le \eta$, then

$$\|(\mathbb{L}\mathbb{T} + \mathbb{W})[\mathbf{a}]\|_{\mathcal{B}}^2 + \|(\mathbf{I}_{\mathcal{A}} + \mathbb{K}\mathbb{T})[\mathbf{a}]\|_{\mathcal{A}}^2 \ge \|(\mathbf{I}_{\mathcal{A}} + \mathbb{K}\mathbb{T})[\mathbf{a}]\|_{\mathcal{A}}^2 \ge (\|\mathbf{a}\| - \|\mathbb{K}\|_{\mathrm{op}}\eta)^2 = (1 - \frac{1}{2})^2 \ge \frac{1}{4}.$$

Otherwise, suppose $\|\mathbb{T}\mathbf{a}\| > \eta$. Then,

$$\|(\mathbb{L}\mathbb{T} + \mathbb{W})[\mathbf{a}]\|_{\mathcal{B}}^2 + \|(\mathbf{I}_{\mathcal{A}} + \mathbb{K}\mathbb{T})[\mathbf{a}]\|_{\mathcal{A}}^2 \ge \|(\mathbb{L}\mathbb{T} + \mathbb{W})[\mathbf{a}]\|_{\mathcal{B}}^2 \ge (\eta - \|\mathbb{W}\|_{\mathrm{op}}\|\mathbf{a}\|)^2 = (\frac{\min\{1, \sigma_{\min}(\mathbb{L})\}}{2\|\mathbb{K}\|_{\mathrm{op}}} - \|\mathbb{W}\|_{\mathrm{op}})^2.$$

Hence, if $\|\mathbb{W}\|_{\mathrm{op}} \le \frac{\min\{1, \sigma_{\min}(\mathbb{L})\}}{4\|\mathbb{K}\|_{\mathrm{op}}}$, the above is at least

$$\frac{\min\{1, \sigma_{\min}(\mathbb{L})^2\}}{16\max\{1, \|\mathbb{K}\|_{\mathrm{op}}^2\}}.$$

□

I. Discretization Arguments

In this section, we use discretizations of the Markov and transition operators. Again, we assume π is feasible. We define the shorthand.

Definition I.1 (Useful Shorthand)**.** Define the short hand

$$\psi_\pi(k_2, k_1) := \|\mathbf{\Phi}_{\mathrm{cl},k_2,k_1+1}^{\pi}\|, \quad L_{\mathrm{ol}} := \exp(\tau L_f), \tag{I.1}$$

and note that $\psi_\pi(k_2, k_1) \le \kappa_{\pi,\infty}$ due to Lemma A.1.

I.1. Discretization of Open-Loop Linearizations

All lemmas in this section assume π is feasible.

Lemma I.1 (Continuity of $\mathbf{x}^\pi(\cdot)$)**.** *Then,*

$$\|\mathbf{x}^\pi(t) - \mathbf{x}^\pi(t')\| \le \kappa_f |t - t'|$$

Proof. Assume without loss of generality that $t' \ge t$. By Assumption 4.1 feasibility of π,

$$\|\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{x}^\pi(s)\| = \|f_{\mathrm{dyn}}(\mathbf{x}^\pi(s), \mathbf{u}^\pi(s)\| \le \kappa_f.$$

Hence, as $\mathbf{x}^\pi(t') = \mathbf{x}^\pi(t) + \int_{s=t}^{t'} f_{\mathrm{dyn}}(\mathbf{x}^\pi(s), \mathbf{u}^\pi(s))\mathrm{d}s$, the bound follows. □

Lemma I.2. *For all $k, s \in \mathcal{I}_k$, we have $\|\mathbf{B}^\pi_{\mathrm{ol}}(t) - \mathbf{B}^\pi_{\mathrm{ol}}(s)\| \vee \|\mathbf{A}^\pi_{\mathrm{ol}}(t) - \mathbf{A}^\pi_{\mathrm{ol}}(s)\| \le \tau\kappa_f M_f$.*

Proof. We bound $\|\mathbf{B}^\pi_{\mathrm{ol}}(t) - \mathbf{B}^\pi_{\mathrm{ol}}(s)\|$ as $\|\mathbf{A}^\pi_{\mathrm{ol}}(t) - \mathbf{A}^\pi_{\mathrm{ol}}(s)\|$ is similar. Assume $s \le t$ without loss of generality. Then,

$$\begin{aligned}
\|\mathbf{B}^\pi_{\mathrm{ol}}(t) - \mathbf{B}^\pi_{\mathrm{ol}}(s)\| &= \|\partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(t)) - \partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(s), \mathbf{u}^\pi(s)) \\
&= \|\partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(s)) - \partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(s), \mathbf{u}^\pi(s))\| && (\mathbf{u}^\pi \in \mathcal{U}_\tau) \\
&= \|\partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(t), \mathbf{u}^\pi(s)) - \partial_u f_{\mathrm{dyn}}(\mathbf{x}^\pi(s), \mathbf{u}^\pi(s))\| && (\mathbf{u}^\pi \in \mathcal{U}_\tau) \\
&\le \|\mathbf{x}^\pi(t) - \mathbf{x}^\pi(s)\| \max_{\alpha\in[0,1]} \|\partial_u f_{\mathrm{dyn}}(\alpha\mathbf{x}^\pi(t) + (1-\alpha)\mathbf{x}^\pi(s), \mathbf{u}^\pi(s))\| && \text{(Mean Value Theorem)} \\
&\le \|\mathbf{x}^\pi(t) - \mathbf{x}^\pi(s)\| M_f && \text{(Assumption 4.1 and convexity of feasibility)} \\
&\le (t-s)\kappa_f M_f \le \tau\kappa_f M_f. && \text{(Lemma I.1)}
\end{aligned}$$

□

Lemma I.3 (Bound on $\mathtt{B}^\pi_{\mathrm{ol},k}$)**.** *For any $k \in [K]$, $\|\mathtt{B}^\pi_{\mathrm{ol},k}\| \le \tau L_{\mathrm{ol}} L_f = \tau L_f \exp(\tau L_f)$.*

Proof. $\|\mathtt{B}^\pi_{\mathrm{ol},k}\| = \int_{s=t_k}^{t_{k+1}} \mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, s)\mathbf{B}^\pi_{\mathrm{ol}}(s)\mathrm{d}s \le \tau \max_{s\in\mathcal{I}_k} \|\mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, s)\|\|\mathbf{B}^\pi_{\mathrm{ol}}(s)\|$. We bound $\|\mathbf{B}^\pi_{\mathrm{ol}}(s)\| \le L_f$ by Assumption 4.1 and $\|\mathbf{\Phi}^\pi_{\mathrm{ol}}(t_{k+1}, s)\| \le L_{\mathrm{ol}}$ by Lemma I.4 below. □

I.2. Discretization of Transition and Markov Operators

We begin by discretizing the open-loop transition operator.

Lemma I.4 (Discretization of Open-Loop Transition Operator)**.** *Recall $L_{\mathrm{ol}} = \exp(\tau L_{\mathrm{ol}})$. $\|\mathbf{\Phi}^\pi_{\mathrm{ol}}(t', t) - \mathbf{I}\| \le (t' - t)L_f \exp((t'-t)L_f)$. Moreover, $\mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t') \le \exp((t'-t)L_f)$. In particular, if $t, t' \in \mathcal{I}_k$, then*

$$\|\mathbf{\Phi}^\pi_{\mathrm{ol}}(t', t) - \mathbf{I}\| \le \tau L_f L_{\mathrm{ol}}, \quad \mathbf{\Phi}^\pi_{\mathrm{ol}}(t', t) \le L_{\mathrm{ol}} = \exp(\tau L_f)$$

Proof of Lemma I.4. For the first part, it suffices to bound the ODE $\mathbf{y}(t') = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t', t)\xi$, where $\xi \in \mathbb{R}^{d_{\mathsf{x}}}$ is an arbitrary initial condition with $\|\xi\| = 1$. Then $\mathbf{y}(t) = \mathbf{\Phi}^\pi_{\mathrm{ol}}(t, t)\xi = \xi$, and $\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{y}(s) = \mathbf{A}^\pi_{\mathrm{ol}}(s)\mathbf{y}(s)$. Hence, $\|\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{y}(s)\| \le L_f\|\mathbf{y}(s)\|$. The result now follows by comparison to the constant ODE $\mathbf{z}(t) = \mathbf{y}(t)$, $\frac{\mathrm{d}}{\mathrm{d}s}\mathbf{z}(s) = 0$, and Picard's Lemma (Lemma C.9). The second part follows from Picard's lemma with comparison the to the stationary curve $\mathbf{z}(0) = 0$. □

Next, we bound the difference between the operator $\tilde{\mathbf{\Phi}}^\pi_{\mathrm{cl}}$ in the definition of $\mathbf{\Phi}^\pi_{\mathrm{cl}}$, and the identity matrix.

Lemma I.5. *Recall the definition $L_{\mathrm{ol}} = \exp(\tau L_f)$ and*

$$\tilde{\mathbf{\Phi}}^\pi_{\mathrm{cl}}(s, t_k) = \mathbf{\Phi}^\pi_{\mathrm{ol}}(s, t_k) + (\int_{s'=t_k}^{s} \mathbf{\Phi}^\pi_{\mathrm{ol}}(s, s')\mathbf{B}^\pi_{\mathrm{ol}}(s')\mathrm{d}s)\mathtt{K}_k.$$

Then,

$$\|\tilde{\mathbf{\Phi}}^\pi_{\mathrm{cl}}(t, t_{k(t)}) - \mathbf{I}\| \le \tau L_f L_{\mathrm{ol}}(1 + L_\pi).$$

Similary,

$$\|\tilde{\mathbf{\Phi}}^\pi_{\mathrm{cl}}(t, t_{k(t)}) - \mathbf{I}\| \le \tau L_f L_{\mathrm{ol}}(1 + L_\pi).$$

Proof of Lemma I.5. Let $t_k = t_{k(t)}$ for shorthand. We have

$$\begin{aligned}
\|\tilde{\mathbf{\Phi}}^{\pi}_{\mathrm{cl}}(t, t_k) - \mathbf{I}\| &= \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t, t_k) + \int_{s=t_k}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t, s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\mathrm{d}s)\mathtt{K}^{\pi}_k - \mathbf{I}\| \\
&\le \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t, t_k) - \mathbf{I}\| + |t - t_k|\|\mathtt{K}^{\pi}_k\| \max_{s\in[t_k,t]} \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t, s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\| \\
&\le \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t, t_k) - \mathbf{I}\| + |t - t_k|L_\pi L_f \max_{s\in[t_k,t]} \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t, s)\| && \text{(Assumption 4.1)} \\
&\le |t - t_k|L_f \exp(L_f(t - t_k)) + |t - t_k|L_\pi L_f \max_{s\in[t_k,t]} \exp(L_f(t - t_k)) && \text{(Lemma I.4)} \\
&\le \tau L_f(1 + L_\pi)\exp(L_f\tau) = \tau L_f(1 + L_\pi)L_{\mathrm{ol}}.
\end{aligned}$$

The bound on $\|\tilde{\mathbf{\Phi}}^{\pi}_{\mathrm{cl}}(t, t_{k(t)}) - \mathbf{I}\|$ can be derived similarly.

□

Lemma I.6 (Discretization of Closed-Loop Transition Operator)**.** *Let $s > t$ such that $t_{k(s)} > t_{k(t)}$. Then, under Definition 4.7,*

$$\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(s, t_{k(t)+1}) - \mathtt{\Phi}^{\pi}_{\mathrm{cl},k(s),k(t)+1}\| \le 2L_f L_{\mathrm{ol}} L_\pi \psi_\pi(k(s), k(t)).$$

Proof of Lemma I.6. As $t_{k(s)} > t_{k(t)}$, we can write $s \in \mathcal{I}_{k_2}$ and $t \in \mathcal{I}_{k_1}$ for $k_2 > k_1$; then

$$k_2 = k(s), \quad k_1 = k(t).$$

We now have

$$\begin{aligned}
\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(s, t_{k(t)+1}) - \mathtt{\Phi}^{\pi}_{\mathrm{cl},k(s),k(t)+1}\| &= \|\tilde{\mathbf{\Phi}}^{\pi}_{\mathrm{cl}}(s, t_{k_2}) - \mathbf{I}\| \cdot \|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k_2,k_1+1}\| \\
&= \|\tilde{\mathbf{\Phi}}^{\pi}_{\mathrm{cl}}(s, t_{k_2}) - \mathbf{I}\| \cdot \psi_\pi(k_2, k_1)
\end{aligned}$$

Directly from Lemma I.5 and $k_2 = k(s)$, $\|\tilde{\mathbf{\Phi}}^{\pi}_{\mathrm{cl}}(s, t_{k_2}) - \mathbf{I}\| \le \tau L_f L_{\mathrm{ol}}(1 + L_\pi)$. This yields, with $L_\pi \ge 1$,

$$\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(s, t_{k(t)+1}) - \mathtt{\Phi}^{\pi}_{\mathrm{cl},k(s),k(t)+1}\| \le \tau L_f L_{\mathrm{ol}}(1 + L_\pi)\psi_\pi(k_2, k_1) \le 2L_f L_{\mathrm{ol}} L_\pi \psi_\pi(k_2, k_1).$$

□

Lemma I.7. *Suppose Definition 4.7 holds and that $\tau \le 1/4L_f \max\{1, L_\pi\}$. Then, $\|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k+1,k}\| \le 5/3$.*

Proof. We have

$$\|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k+1,1}\| = \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k+1}, t_k) + \mathtt{B}^{\pi}_{\mathrm{ol},k}\mathtt{K}_k\| \le \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k+1}, t_k)\| + L_\pi\|\mathtt{B}^{\pi}_{\mathrm{ol},k}\|,$$

where we use $\|\mathtt{K}_k\| \le L_\pi$ under Definition 4.7. By Lemma I.3, $\|\mathtt{B}^{\pi}_{\mathrm{ol},k}\| \le \tau L_f \exp(\tau L_f)$ and by Lemma I.4,$\|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k+1}, t_k)\| \le L_{\mathrm{ol}} = \exp(\tau L_f)$. Then, $\|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k+1,1}\| \le \exp(\tau L_f)(1+\tau L_f L_\pi)$. For $\tau \le 1/L_f \max\{1, L_\pi\}$, we have $\|\mathtt{\Phi}^{\pi}_{\mathrm{cl},k+1,1}\| \le \exp(1/4)(1 + 1/4) \le 5/3$. □

Finally, we turn to a discretization of the Markov operator:

Lemma I.8 (Discretization of Closed-Loop Markov Operator)**.** *The following bounds hold:*

(a) For any $k_2 > k_1$, we have

$$\max_{t\in\mathcal{I}_{k_1}, s\in\mathcal{I}_{k_2}} \|\tau^{-1}\mathtt{\Psi}^{\pi}_{\mathrm{cl},k_2,k_1} - \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(s, t)\| \le \tau\psi_\pi(k_2, k_1) \underbrace{L_{\mathrm{ol}}\left(\kappa_f M_f + 4L_f^2 L_\pi\right)}_{L_{\mathrm{cl},1}}.$$

(b) For any $k_2 > k_1$, we have

$$\sup_{t\in\mathcal{I}_{k_1}} \|\tau^{-1}\Psi^{\pi}_{\mathrm{cl},k_2,k_1} - \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(k_2,t)\| \le \tau\psi_\pi(k_2,k_1)\underbrace{L_{\mathrm{ol}}\left(\kappa_f M_f + 2L_f^2\right)}_{L_{\mathrm{cl},2}}$$
$$\le \tau\kappa_{\pi,\infty}L_{\mathrm{ol}}\left(\kappa_f M_f + 2L_f^2\right).$$

(c) For any (s,t) with $t_{k(s)} = t_{k(t)}$, we have

$$\|\Psi^{\pi}_{\mathrm{cl},t_{k(s)},t_{k(t)}}\| \le L_{\mathrm{ol}}L_f$$

(d) For any $1 \le k_1 < k_2 \le K$ and t for which $t_{k(t)} = k_1$,

$$\frac{1}{\tau}\|\Psi^{\pi}_{\mathrm{cl},k_2,k_1}\| \vee \|\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_{k_2},t)\| \le L_{\mathrm{ol}}L_f\psi_\pi(k_2,k_1) \le L_{\mathrm{ol}}L_f\kappa_{\pi,\infty}.$$

Proof. Let us start with the first bound. Set $k_1 = k(t)$, and $k_2 = k(s)$. Note that $\Psi^{\pi}_{\mathrm{cl},k_2,k_1} := \Phi^{\pi}_{\mathrm{cl},k_2,k_1+1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1}$

$$\begin{aligned}
&\|\tau^{-1}\Psi^{\pi}_{\mathrm{cl},k_2,k_1} - \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(s,t)\| \\
&= \|\tau^{-1}\Phi^{\pi}_{\mathrm{cl},k_2,k_1+1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1} - \mathbf{\Phi}^{\pi}_{\mathrm{cl}}(s,t_{k_1+1})\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t)\| \\
&\le \|\Phi^{\pi}_{\mathrm{cl},k_2,k_1+1}(\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \tau^{-1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1})\| + \|(\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(s,t_{k_1+1}) - \Phi^{\pi}_{\mathrm{cl},k_2,k_1+1})\mathbf{B}^{\pi}_{\mathrm{ol}}(t)\| \\
&\le \psi_\pi(k_2,k_1)\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \tau^{-1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1}\| + L_f\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(s,t_{k_1+1}) - \Phi^{\pi}_{\mathrm{cl},k_2,k_1+1}\| \quad \text{(Assumption 4.1 and Eq. (I.1))} \\
&\le \psi_\pi(k_2,k_1)\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \tau^{-1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1}\| + 2\tau L_f^2 L_{\mathrm{ol}}L_\pi\psi_\pi(k_2,k_1). \quad \text{(Lemma I.6)}
\end{aligned}$$

Finally, we bound

$$\begin{aligned}
&\|\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \tau^{-1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1}\| \\
&= \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \tau^{-1}\int_{s=t_{k_1}}^{t_{k_1+1}} \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},s)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\mathrm{d}s\| \\
&\le \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \tau^{-1}\int_{s=t_{k_1}}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\mathrm{d}s\| + \|\tau^{-1}\int_{s=t_{k_1}}^{t} \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},s) - \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\mathrm{d}s\| \\
&\le \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\| \max_{s\in\mathcal{I}_{k_1}} \|\mathbf{B}^{\pi}_{\mathrm{ol}}(t) - \mathbf{B}^{\pi}_{\mathrm{ol}}(s)\| + L_f \max_{s\in\mathcal{I}_{k_1}} \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},s) - \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\|\|\mathbf{B}^{\pi}_{\mathrm{ol}}(s)\| \\
&\le \tau L_{\mathrm{ol}}\kappa_f M_f + L_f \max_{s\in\mathcal{I}_{k_1}} \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},s) - \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\| \\
&\le \tau L_{\mathrm{ol}}\kappa_f M_f + 2\tau L_f^2 L_{\mathrm{ol}}
\end{aligned}$$

where the second-tolast step uses Lemmas I.2 and I.4 and Assumption 4.1, and the last step uses $\|\|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},s) - \mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t)\| \le \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},s) - \mathbf{I}\| + \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(t_{k_1+1},t) - \mathbf{I}\| \le 2\tau L_f L_{\mathrm{ol}}$ by Lemma I.4. Combining with the previous display,

$$\begin{aligned}
&\|\tau^{-1}\Psi^{\pi}_{\mathrm{cl},k_2,k_1} - \mathbf{\Psi}^{\pi}_{\mathrm{cl}}(s,t)\| \\
&\le \tau\psi_\pi(k_2,k_1)\left(L_{\mathrm{ol}}\kappa_f M_f + 2L_f^2 L_{\mathrm{ol}} + 2L_f^2 L_{\mathrm{ol}}L_\pi\right) \\
&\le \tau L_{\mathrm{ol}}\psi_\pi(k_2,k_1)\left(\kappa_f M_f + 4L_f^2 L_\pi\right) \qquad (L_\pi \ge 1)
\end{aligned}$$

This concludes the proof of (a).

For (b), the argument is the same, but the contribution of $2L_f^2 L_{\mathrm{ol}}L_\pi$ vanishes, as $\mathbf{\Phi}^{\pi}_{\mathrm{cl}}(t_{k(s)},t_{k_1+1}) = \Phi^{\pi}_{\mathrm{cl},k_2,k_1+1}$.

For (c), we note that if $t_{k(s)} = t_{k(t)}$, $\|\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(s,t)\| = \|\mathbf{\Phi}^{\pi}_{\mathrm{ol}}(s,t)\mathbf{B}^{\pi}_{\mathrm{ol}}(t)\| \le L_{\mathrm{ol}}L_f$ by Assumption 4.1 and Lemma I.4. For the final inequality, we have by Eq. (I.1) and Lemma I.3 that

$$\|\Psi^{\pi}_{\mathrm{cl},t_{k(s)},t_{k(t)}}\| = \|\Phi^{\pi}_{\mathrm{cl},k_2,k_1+1}\mathsf{B}^{\pi}_{\mathrm{ol},k_1}\| \le \tau L_{\mathrm{ol}}L_f\psi_\pi(k_2,k_1).$$

The bound on $\|\mathbf{\Psi}^{\pi}_{\mathrm{cl}}(t_{k_2},t)\|$ is similar. □

I.3. Discretization of the Gradient (Proof of Proposition A.4)

We use the shorthand from Definition I.1. Recall that $\tilde{\nabla}\mathcal{J}_T(\pi) = \frac{1}{\tau}\tau(\nabla\mathcal{J}_T^{\text{disc}}(\pi))$ is the continuous-time inclusion of the discrete-time gradient, renormalized by τ^{-1}. Thus, from Lemma C.7,

$$\begin{aligned}\tilde{\nabla}\mathcal{J}_T(\pi)(t) = \tau^{-1}Q_u^\pi(t_{k(t)}) + \tau^{-1}(\Psi_{\text{cl},T,t_{k(t)}}^\pi)^\top(\partial_x V(\mathbf{x}^\pi(T))) \\ + \sum_{j=k(t)+1}^{K}(\Psi_{\text{cl},j,k(t)}^\pi)^\top(Q_x^\pi(t_j) + \mathtt{K}_j Q_u^\pi(t_j)).\end{aligned}$$

From Lemma C.6, we can write

$$\begin{aligned}\nabla\mathcal{J}_T(\pi)(t) = Q_u^\pi(t) + \mathbf{\Psi}_{\text{cl}}^\pi(T,t)^\top(\partial_x V(\mathbf{x}^\pi(T))) \\ + \int_{s=t_{k(t)}}^{t}\mathbf{\Psi}_{\text{cl}}^\pi(s,t)^\top Q_x^\pi(s)\mathrm{d}s \\ + \sum_{j=k(t)+1}^{K}\int_{s\in\mathcal{I}_j}\left(\mathbf{\Psi}_{\text{cl}}^\pi(s,t)Q_x^\pi(s) + \mathbf{\Psi}_{\text{cl}}^\pi(t_j,t)^\top\mathtt{K}_j{}^\top Q_u^\pi(s)\right)\mathrm{d}s,\end{aligned}$$

and therefore decompose the error into five terms via the triangle inequality.

$$\begin{aligned}\|\tilde{\nabla}\mathcal{J}_T(\pi)(t) - \nabla\mathcal{J}_T(\pi)(t)\| \le \underbrace{\|Q_u^\pi(t) - Q_u^\pi(t_{k(t)})\|}_{:=\text{Term}_1} + \underbrace{\|(\mathbf{\Psi}_{\text{cl}}^\pi(T,t) - \tau^{-1}(\Psi_{\text{cl},T,t_{k(t)}}^\pi))^\top(\partial_x V(\mathbf{x}^\pi(T)))\|}_{:=\text{Term}_2} \\ + \underbrace{\int_{s=t}^{t_{k(t)+1}}\|\mathbf{\Psi}_{\text{cl}}^\pi(t_{k(s)},t_{k(t)})^\top Q_x^\pi(s)\|\mathrm{d}s}_{:=\text{Term}_3} \\ + \tau\sum_{j=k(t)+1}^{K}\text{Term}_{4,j} + \text{Term}_{5,j},\end{aligned}$$

where we further define

$$\begin{aligned}\text{Term}_{4,j} &:= \sup_{s\in\mathcal{I}_j}\|\mathbf{\Psi}_{\text{cl}}^\pi(s,t)Q_x^\pi(s) - \tau^{-1}(\Psi_{\text{cl},j,k(t)}^\pi)^\top(Q_x^\pi(t_j)\| \\ \text{Term}_{5,j} &:= \sup_{s\in\mathcal{I}_j}\|\mathbf{\Psi}_{\text{cl}}^\pi(t_j,t)^\top(\mathtt{K}_j)^\top Q_u^\pi(s) - \tau^{-1}(\Psi_{\text{cl},j,k(t)}^\pi)^\top(\mathtt{K}_j)^\top Q_u^\pi(t_j)\|\end{aligned}$$

Before contitnuous, we apply the following lemma.

Lemma I.9 (Discretization of Cost-Gradient)**.** *For $z \in \{x, u\}$,*

$$\|Q_x^\pi(t) - Q_x^\pi(t_k(t))\| \vee \|Q_u^\pi(t) - Q_u^\pi(t_k(t))\| \le \tau M_{\text{cost}}(1+\kappa_f).$$

Proof. We bound $\|Q_x^\pi(t) - Q_x^\pi(t_k(t))\|$ as the bound on $\|Q_u^\pi(t) - Q_u^\pi(t_k(t))\|$ is similar.

$$\begin{aligned}\|Q_x^\pi(t) - Q_x^\pi(t_k(t))\| &= \|\partial_x Q(\mathbf{x}^\pi(t),\mathbf{u}^\pi(t),t)) - \partial_x Q(\mathbf{x}^\pi(t_{k(t)})\mathbf{u}^\pi(t_{k(t)}),t_{k(t)})\| \\ &= \|\partial_x Q(\mathbf{x}^\pi(t),\mathbf{u}^\pi(t_{k(t)}),t)) - \partial_x Q(\mathbf{x}^\pi(t_{k(t)}),\mathbf{u}^\pi(t_{k(t)}),t_{k(t)})\| && (\mathbf{u}^\pi(\cdot)\in\mathcal{U}_\tau) \\ &\le M_{\text{cost}}|t - t_{k(t)}| + M_{\text{cost}}\|\mathbf{x}^\pi(t_{k(t)}) - \mathbf{x}^\pi(t)\| && (\text{Integrating Assumption 4.2}) \\ &\le M_{\text{cost}}\tau + M_{\text{cost}}\kappa_f\tau = \tau M_{\text{cost}}(1+\kappa_f). && (\text{Lemma I.1})\end{aligned}$$

□

From Lemma I.9, we bound

$$\text{Term}_1 \le \tau M_{\text{cost}}(1+\kappa_f).$$

Next, using that T/τ is integral by assumption, i.e. $t = t_{k(T)}$, we have

$$\begin{aligned}
\text{Term}_2 &= \|(\mathbf{\Psi}^\pi_{\text{cl}}(T,t) - \tau^{-1}(\Psi^\pi_{\text{cl},T,t_{k(t)}}))^\top (\partial_x V(\mathbf{x}^\pi(T)))\| \\
&\le L_{\text{cost}} \|\mathbf{\Psi}^\pi_{\text{cl}}(T,t) - \tau^{-1}(\Psi^\pi_{\text{cl},T,t_{k(t)}})\| && \text{(Assumption 4.2)} \\
&\le \tau L_{\text{cost}} L_{\text{cl},2} \psi_\pi(k(T), k(t)) = \tau L_{\text{cost}} L_{\text{cl},2} \psi_\pi(K+1, k(t)), && \text{(Lemma I.8(b))}
\end{aligned}$$

For the third term, we have

$$\begin{aligned}
\text{Term}_3 &= \int_{s=t}^{t_{k(t)+1}} \|\mathbf{\Psi}^\pi_{\text{cl}}(t_{k(s)}, t_{k(t)})^\top Q^\pi_x(s)\| \mathrm{d}s \\
&\le L_{\text{cost}} \int_{s=t}^{t_{k(t)+1}} \|\mathbf{\Psi}^\pi_{\text{cl}}(t_{k(s)}, t_{k(t)})^\top\| \mathrm{d}s && \text{(Assumption 4.2)} \\
&\le \tau L_{\text{cost}} \max_{s \in [t, t_{k(t)+1})} \|\mathbf{\Psi}^\pi_{\text{cl}}(t_{k(s)}, t_{k(t)})^\top\| && \text{(ignoring interval endpoint due to integration)} \\
&= L_{\text{cost}} L_{\text{ol}} L_f,
\end{aligned}$$

where in the last step we use Lemma I.8(c). Summarizing these the bounds on the first and third term,

$$\text{Term}_1 + \text{Term}_2 + \text{Term}_3 \le \tau(L_{\text{cost}} L_{\text{ol}} L_f + M_{\text{cost}}(1+\kappa_f) + L_{\text{cost}} L_{\text{cl},2} \psi_\pi(K+1, k(t)) \tag{I.2}$$

Next, we turn to the fourth and and fifth terms. We bound

$$\begin{aligned}
\text{Term}_{4,j} &:= \sup_{s \in \mathcal{I}_j} \|\mathbf{\Psi}^\pi_{\text{cl}}(s,t) Q^\pi_x(s) - \tau^{-1}(\Psi^\pi_{\text{cl},j,k(t)})^\top (Q^\pi_x(t_j)\| \\
&\le \|\tau^{-1}\Psi^\pi_{\text{cl},j,k(t)}\| \cdot \sup_{s \in \mathcal{I}_j} \|Q^\pi_x(s) - Q^\pi_x(t_j)\| + \|\tau^{-1}\Psi^\pi_{\text{cl},j,k(t)} - \mathbf{\Psi}^\pi_{\text{cl}}(s,t)\| \|Q^\pi_x(t_j)\| \\
&\le L_{\text{ol}} L_f \psi_\pi(j, k(t)) \cdot \sup_{s \in \mathcal{I}_j} \|Q^\pi_x(s) - Q^\pi_x(t_j)\| + \tau L_{\text{cl},1} \psi_\pi(j, k(t)) \|Q^\pi_x(t_j)\| && \text{(Lemma I.8(a\&c))} \\
&\le L_{\text{ol}} L_f \psi_\pi(j, k(t)) \cdot \tau M_{\text{cost}}(1+\kappa_f)) + \tau L_{\text{cl},1} \psi_\pi(j, k(t)) L_{\text{cost}} && \text{(Assumption 4.2 and Lemma I.9)} \\
&= \tau \psi_\pi(j, k(t)) (L_{\text{ol}} L_f M_{\text{cost}}(1+\kappa_f) + L_{\text{cost}} L_{\text{cl},1}).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\text{Term}_{5,j} &= \sup_{s \in \mathcal{I}_j} \|\mathbf{\Psi}^\pi_{\text{cl}}(t_j, t)^\top (\mathtt{K}_j)^\top Q^\pi_u(s) - \tau^{-1}(\Psi^\pi_{\text{cl},j,k(t)})^\top (\mathtt{K}_j)^\top Q^\pi_u(t_j)\| \\
&\le \|\tau^{-1}(\Psi^\pi_{\text{cl},j,k(t)})\| \|\mathtt{K}_j\| \sup_{s \in \mathcal{I}_j} \|Q^\pi_u(s) - Q^\pi_u(t_j)\| \\
&\quad + \sup_{s \in \mathcal{I}_j} \|\tau^{-1}(\Psi^\pi_{\text{cl},j,k(t)}) - \mathbf{\Psi}^\pi_{\text{cl}}(t_j, t)\| \|\mathtt{K}_j\| \|Q^\pi_u(t_j)\| \\
&\le L_\pi \|\tau^{-1}(\Psi^\pi_{\text{cl},j,k(t)})\| \sup_{s \in \mathcal{I}_j} \|Q^\pi_u(s) - Q^\pi_u(t_j)\| \\
&\quad + L_\pi \sup_{s \in \mathcal{I}_j} \|\tau^{-1}(\Psi^\pi_{\text{cl},j,k(t)}) - \mathbf{\Psi}^\pi_{\text{cl}}(t_j, t)\| \|Q^\pi_u(t_j)\| && \text{(Definition 4.7)} \\
&\le L_\pi L_{\text{ol}} L_f \psi_\pi(j, k(t)) \sup_{s \in \mathcal{I}_j} \|Q^\pi_u(s) - Q^\pi_u(t_j)\| && \text{(Lemma I.8(d))} \\
&\quad + L_\pi L_{\text{cl},2} \psi_\pi(j, k(t)) \|Q^\pi_u(t_j)\| && \text{(Lemma I.8(b))} \\
&\le \tau L_\pi L_{\text{ol}} L_f \psi_\pi(j, k(t)) M_{\text{cost}}(1+\kappa_f) + L_\pi L_{\text{cl},2} \psi_\pi(j, k(t)) L_{\text{cost}} && \text{(Assumption 4.2 and Lemma I.9)} \\
&= \tau \psi_\pi(j, k(t)) \left(L_\pi L_{\text{ol}} L_f M_{\text{cost}}(1+\kappa_f) + L_\pi L_{\text{cl},2} L_{\text{cost}}\right).
\end{aligned}$$

Hence,

$$\text{Term}_{4,j} + \text{Term}_{5,j} \le \tau \psi_\pi(j, k(t)) \left((1+L_\pi) L_{\text{ol}} L_f M_{\text{cost}}(1+\kappa_f) + L_{\text{cost}}(L_{\text{cl},1} + L_\pi L_{\text{cl},2})\right)$$

Using the definitions of $L_{\mathrm{cl},1}, L_{\mathrm{cl},2}$ in Lemma I.8 and $L_\pi \geq 1$, we have

$$\begin{aligned}(L_{\mathrm{cl},1} + L_\pi L_{\mathrm{cl},2}) &= L_{\mathrm{ol}}(\kappa_f M_f + 4L_f^2 L_\pi + L_\pi(\kappa_f M_f + 2L_f^2)) \\ &= L_{\mathrm{ol}}((1+L_\pi)\kappa_f M_f + 6L_f^2 L_\pi \leq L_\pi(2\kappa_f M_f + 6L_f^2))\end{aligned}$$

Substituing into the the previous display and again using $L_\pi, \geq 1$,

$$\begin{aligned}\mathrm{Term}_{4,j} + \mathrm{Term}_{5,j} &\leq \tau\psi_\pi(j,k(t))L_{\mathrm{ol}}\left(2L_\pi L_f M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}L_\pi(2\kappa_f M_f + 6L_f^2)\right) \\ &\leq 2L_\pi\tau\psi_\pi(j,k(t))L_{\mathrm{ol}}\left(L_f M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}(\kappa_f M_f + 3L_f^2)\right) \\ &\leq \tau\psi_\pi(j,k(t))\underbrace{2L_\pi L_{\mathrm{ol}}\left(L_f M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}(\kappa_f M_f + 3L_f^2)\right)}_{:=L_{\mathrm{cl},3}}.\end{aligned}$$

Hence,

$$\begin{aligned}\tau\sum_{j=k(t)+1}^{K}\mathrm{Term}_{4,j} + \mathrm{Term}_{5,j} &\leq \tau L_{\mathrm{cl},3}\cdot(\tau\sum_{j=k(t)+1}^{K}\psi_\pi(j,k(t))) \\ &\leq \tau L_{\mathrm{cl},3}\kappa_{\pi,1}. \qquad (\tau\textstyle\sum_{j=k(t)+1}^{K}\psi_\pi(j,k(t)) = \tau\sum_{j=k(t)+1}^{K}\|\Phi_{\mathrm{cl},j,k}^\pi\| \leq \kappa_{\pi,1})\end{aligned}$$

In sum, we conclude that

$$\begin{aligned}&\|\tilde{\nabla}\mathcal{J}_T(\pi)(t) - \nabla\mathcal{J}_T(\pi)(t)\| \\ &\leq \tau\left(L_{\mathrm{cost}}L_{\mathrm{ol}}L_f + M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}L_{\mathrm{cl},2}\psi_\pi(K+1,k(t)) + \kappa_{\pi,1}L_{\mathrm{cl},3}\right) \\ &\leq \tau\left(L_{\mathrm{cost}}L_{\mathrm{ol}}L_f + M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}L_{\mathrm{cl},2}\kappa_{\pi,\infty} + \kappa_{\pi,1}L_{\mathrm{cl},3}\right) \\ &\leq \tau\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}\left(L_{\mathrm{cost}}L_{\mathrm{ol}}L_f + M_{\mathrm{cost}}(1+\kappa_f) + (L_{\mathrm{cost}}L_{\mathrm{cl},2} + L_{\mathrm{cl},3})\right).\end{aligned}$$

Finally, using the definition of $L_{\mathrm{cl},2} := L_{\mathrm{ol}}\left(\kappa_f M_f + 2L_f^2\right)$ in Lemma I.8(b) and the definition of $L_{\mathrm{cl},3} := 2L_\pi L_{\mathrm{ol}}\left(L_f M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}(\kappa_f M_f + 3L_f^2)\right)$ defined above, and $L_\pi \geq 1$,

$$M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}L_{\mathrm{ol}}L_f + L_{\mathrm{cost}}L_{\mathrm{cl},2} + L_{\mathrm{cl},3} \leq L_\pi L_{\mathrm{ol}}\left((1+L_f)M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}(3\kappa_f M_f + 8L_f^2 + L_f)\right).$$

Thus, recalling $L_{\mathrm{ol}} = \exp(\tau L_f)$,

$$\|\tilde{\nabla}\mathcal{J}_T(\pi)(t) - \nabla\mathcal{J}_T(\pi)(t)\| \leq \tau e^{\tau L_f}\max\{\kappa_{\pi,\infty},\kappa_{\pi,1},1\}L_\pi\left((1+L_f)M_{\mathrm{cost}}(1+\kappa_f) + L_{\mathrm{cost}}(3\kappa_f M_f + 8L_f^2 + L_f)\right),$$

as needed.

□

Part II

Experiments

J. Experiments Details

J.1. Implementation Details

`trajax` (Frostig et al., 2021) is used for nonlinear iLQR trajectory optimization and `haiku`+`optax` (Hennigan et al., 2020; Babuschkin et al., 2020) for training neural network dynamics models.

J.2. Environments

J.2.1. PENDULUM

We consider simple pendulum dynamics with state $(\theta, \dot{\theta})$ and input u:

$$\ddot{\theta} = \sin(\theta) + u.$$

To integrate these dynamics, we use a standard forward Euler approximation with step size $\tau = 0.15$, applying a zero-order hold to the input. The goal is to swing up the pendulum to the origin state $(0, 0)$. We consider the cost function:

$$c((\theta, v), u) = \theta^2 + v^2 + u^2.$$

Evaluation details. All methods were evaluated over a horizon of length $T = 25$ on initial states sampled from $\mathrm{Unif}([-1 + \pi, 1 + \pi] \times \{0\})$.

Random state sampling distribution. For learning from random states and actions, we sample the initial condition from $\mathrm{Unif}([-5, 5]^2)$ and random actions from $\mathrm{Unif}([-1, 1])$.

Optimization Details. We use $N = 100$ samples

J.2.2. QUADROTOR

The 2D quadrotor 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 $m = 0.8$, $g = 0.1$, and $I_{xx} = 0.5$. Again, we integrate these dynamics using forward Euler with step size $\tau = 0.1$. The task is to move the quadrotor to the origin state. The cost function we use is:

$$c((x, z, \phi, \dot{x}, \dot{z}, \dot{\phi}), (u_1, u_2)) = x^2 + z^2 + 10\phi^2 + 0.1(\dot{x}^2 + \dot{z}^2 + \dot{\phi}^2) + 0.1(u_1^2 + u_2^2).$$

Evaluation details. All methods were evaluated over a horizon of length $T = 25$ on initial states sampled from $\mathrm{Uniform}([-0.5, 0.5]^2 \times \{0\}^4)$

Random state sampling distribution. For learning from random states and actions, we sample the initial condition from $\mathrm{Unif}([-1, 1]^6)$ and random actions from $\mathrm{Unif}([-0.5, 0.5])$.

J.3. Neural network training

For modeling environment dynamics, we consider three layer fully connected neural networks For pendulum, we set the width to 96, the learning rate to 10^{-3}, and the activation to swish. For quadrotor, we set the width to 128, the learning rate to 5×10^{-3}, and the activation to gelu. We use the Adam optimizer with 10^{-4} additive weight decay and a cosine decay learning schedule.

J.4. Least Squares

While Algorithm 1 features a method of moments estimator to estimate the Markov transition operators, our implementation relies on using regularized least squares. Specifically, we solve:

$$\underbrace{\begin{bmatrix} \hat{\Psi}_{j,1} \\ \hat{\Psi}_{j,2} \\ \cdots \\ \hat{\Psi}_{j,j-1} \end{bmatrix}^\top}_{\in \mathbb{R}^{d_{\mathtt{x}} \times (j-1) d_{\mathtt{u}}}} = \left(\sum_{i=1}^{N} \underbrace{\begin{bmatrix} \mathtt{w}_1^{(i)} \\ \mathtt{w}_2^{(i)} \\ \cdots \\ \mathtt{w}_{j-1}^{(i)} \end{bmatrix}}_{\in \mathbb{R}^{(j-1) d_{\mathtt{u}}}} \begin{bmatrix} \mathtt{w}_1^{(i)} \\ \mathtt{w}_2^{(i)} \\ \cdots \\ \mathtt{w}_{j-1}^{(i)} \end{bmatrix}^\top + \lambda \mathbf{I} \right)^{-1} \cdot \sum_{i=1}^{N} \begin{bmatrix} \mathtt{w}_1^{(i)} \\ \mathtt{w}_2^{(i)} \\ \cdots \\ \mathtt{w}_{j-1}^{(i)} \end{bmatrix} (\mathtt{y}_j^{(i)} - \hat{\mathtt{x}}_j)^\top \tag{J.1}$$

J.5. Scaling the gain matrix

In order to stabilize the gain computation during gain estimation (Algorithm 3), we scale the update to the $\hat{\mathtt{P}}_k$ matrix as:

$$\hat{\mathtt{P}}_k \leftarrow \hat{\mathtt{P}}_k \cdot \frac{1}{1 + 0.01 \|\hat{\mathtt{P}}_k\|_F}.$$