Foundations of Safe Online Reinforcement Learning in the Linear Quadratic Regulator: $\sqrt{T}$-Regret

Benjamin Schiffer BSCHIFFER1@G.HARVARD.EDU and **Lucas Janson** LJANSON@FAS.HARVARD.EDU
Department of Statistics, Harvard University

Editors: G. Sukhatme, L. Lindemann, S. Tu, A. Wierman, N. Atanasov

Abstract

Understanding how to efficiently learn while adhering to safety constraints is essential for using online reinforcement learning in practical applications. However, proving rigorous regret bounds for safety-constrained reinforcement learning is difficult due to the complex interaction between safety, exploration, and exploitation. In this work, we seek to establish foundations for safety-constrained reinforcement learning by studying the canonical problem of controlling a one-dimensional linear dynamical system with unknown dynamics. We study the safety-constrained version of this problem, where the state must with high probability stay within a safe region, and we provide the first safe algorithm that achieves regret of $\tilde{O}_T(\sqrt{T})$. Furthermore, the regret is with respect to the baseline of truncated linear controllers, a natural baseline of non-linear controllers that are well-suited for safety-constrained linear systems. In addition to introducing this new baseline, we also prove several desirable continuity properties of the optimal controller in this baseline. In showing our main result, we prove that whenever the constraints impact the optimal controller, the non-linearity of our controller class leads to a faster rate of learning than in the unconstrained setting.

1. Introduction

1.1. Background and Motivation

Online reinforcement learning (RL) algorithms are powerful tools for interacting with and learning about unknown environments (Levine et al., 2016; Lillicrap et al., 2015; Tewari and Murphy, 2017). The core idea behind many successful RL algorithms is carefully balancing exploration and exploitation. However, in many real world applications, online RL algorithms must satisfy a set of safety constraints. Importantly, these safety constraints must be satisfied even while the algorithm learns, leading to a complex interaction between safety and learning. Safety constraints reduce an algorithm's ability to explore because the algorithm must take actions that are known to be safe. Similarly, safety constraints reduce an algorithm's ability to exploit because actions that exploit known information may lead to unsafe states. As an example, consider a self-driving car that uses online RL to learn how to navigate a new environment in real time. To do this, an RL algorithm must make adjustments to speed and acceleration that account for unknown environmental factors such as wind speed and friction. However, the algorithm controlling a car in the real world must keep the car in safe states and avoid crashing into other objects. Therefore, it is critical that the algorithm learns while being safe. A better understanding of the relationship between learning and safety constraints is crucial for deploying online reinforcement learning algorithms in the real world. In this paper, we focus on understanding how safety and learning interact for a canonical learning problem in control theory known as *online linear quadratic regulator (LQR) learning*. While online LQR learning is one of the simplest learning problems with a continuous action space, this problem highlights the inherent differences between learning without safety constraints and learning with safety constraints.

1.2. Setting and Motivation

We study the problem of learning and controlling a discrete-time linear dynamical system when the dynamics are unknown and safety must be maintained during online learning. At each time step, the algorithm observes the current state and chooses a control (action). The state at the next time step then depends on the current state, the chosen control, and random noise. The way in which the next state depends on the current state and chosen control is referred to as the *dynamics of the system*. The goal of the problem is to choose actions that minimize a quadratic cost by keeping the state close to the origin while using minimal control. This model is used, e.g., in robotics when a robot (drone, submarine, rocket, etc.) attempts to stay close to a single point while being subject to random environmental forces (Rubio et al., 2016). In practice, the dynamics of the system (such as air resistance) are not known a priori. Therefore, we study this problem when the dynamics are unknown, and the algorithm must minimize cost while learning the unknown dynamics. To model safety in this setting, we assume that the state must stay within a predefined 'safe region.' For example, the robot described above cannot move to states that make the robot crash into other objects.

When the dynamics are known and there are no safety constraints, the optimal algorithm is the Linear Quadratic Regulator, which is well-studied in the field of control theory (Rawlings and Mayne, 2009). However, the addition of state constraints significantly complicates even this simple problem, and there no longer exists a closed-form solution for the constrained version of this problem with known dynamics (Rawlings and Mayne, 2009). In order to make the problem more tractable, we study this problem when both states and controls are one-dimensional; Schiffer and Janson (2024) take the same approach in analyzing the one-dimensional constrained linear systems. One-dimensional linear systems have been frequently studied as a first step toward understanding other complex aspects of control theory, see e.g. Fefferman et al. (2021); Abeille and Lazaric (2017). Furthermore, some real-world problems can be represented as one-dimensional LQR problems. As an example, consider the simple setting of controlling the temperature of a room, a common problem in control (Oldewurtel et al., 2008). The possible actions include adding different amounts of hot air or cold air to the room, and a natural goal is to minimize costs (the amount of energy used) while also keeping the room close to a specific temperature. In this setting, state constraints would consist of constraining the temperature to stay within a 'safe' region that is not too hot and not too cold.

1.3. Our Contribution

The goal of this paper is to provide foundations for analyzing safety-constrained LQR learning using non-linear baselines of controllers that are better suited for the constrained problem. Our main result is the first algorithm for safety-constrained one-dimensional LQR with unknown dynamics that with high probability guarantees $\tilde{O}_T(\sqrt{T})$ regret. In this setting, our work improves upon the previous best results, in particular Li et al. (2021b); Dean et al. (2019) prove $\tilde{O}_T(T^{2/3})$ regret bounds and only for bounded noise distributions.

The rate of $\tilde{O}_T(\sqrt{T})$ matches the optimal rate of regret in the unconstrained LQR learning problem. Note that unconstrained LQR learning is a special case of constrained LQR learning with sufficiently loose constraints. Therefore, because the lower bound for unconstrained LQR learning is $\tilde{\Omega}_T(\sqrt{T})$ regret (Ziemann and Sandberg, 2024), it is impossible to in general do better than $\tilde{\Omega}_T(\sqrt{T})$ regret for the constrained problem. In addition to improving the rate of regret, the $\tilde{O}_T(\sqrt{T})$ regret is also with respect to a stronger baseline than previous works. More specifically, the regret is defined with respect to the best controller from the baseline class of truncated linear controllers, which consists of linear controllers corrected to obey the safety constraints. This baseline is naturally well-suited for safe control and is a significantly stronger baseline than studied in previous works (see Section 2.3 for more details). Because the controllers in this class are frequently non-linear, we also introduce new theoretical tools for analyzing this type of non-linear controller. Therefore, our $\tilde{O}_T(\sqrt{T})$ regret result is strictly stronger than the previous $\tilde{O}_T(T^{2/3})$ regret results of Li et al. (2021b); Dean et al. (2019) applied to our setting, in both the regret baseline and the rate

of regret. Note that these previous works also assume bounded noise distribution, while our results hold for any sub-gaussian distribution. Informally, our main theorem can be stated as follows:

Theorem 1 (Informal) *For safety-constrained one-dimensional LQR with unknown dynamics and any sub-gaussian noise distribution, there exists an algorithm that with high probability is safe and has regret that scales with $\sqrt{T}$ compared to the best truncated linear controller with known dynamics.*

To prove Theorem 1, we show that either the constraints are tight enough to give faster learning rates or loose enough that the problem is approximately unconstrained. This dichotomy is the main idea behind our algorithm achieving $\tilde{O}_T(\sqrt{T})$ regret for all possible noise distributions. We also show that the class of truncated linear controllers satisfies desirable continuity properties, which may be of independent interest.

1.4. Related Work

Safe reinforcement learning has been studied in many different contexts with various definitions of safety, including reachability of safe sets and long term stability (Ganai et al., 2024; Garg et al., 2024; Gu et al., 2022; Moldovan and Abbeel, 2012; Wachi et al., 2018, 2024; Yao et al., 2024). Specifically in control theory, there exist many methods that satisfy different notions of safety for specific control tasks (Fulton and Platzer, 2018; Cheng et al., 2019; Marvi and Kiumarsi, 2021; Fisac et al., 2018).

The LQR learning problem has recently gained significant attention after Abbasi-Yadkori and Szepesvári (2011) showed that $\tilde{O}_T(\sqrt{T})$ regret is possible in the unconstrained LQR learning problem. Subsequent works have built on these results with many variations and more efficient algorithms (Dean et al., 2018; Mania et al., 2019, 2020; Simchowitz et al., 2018; Cohen et al., 2019; Wang and Janson, 2021, 2022; Abeille and Lazaric, 2017; Zheng and Li, 2020; Sun et al., 2020; Khosravi and Smith, 2020; Sattar and Oymak, 2022; Ye et al., 2024; Athrey et al., 2024; Ziemann and Sandberg, 2024; Lee et al., 2024; Simchowitz and Foster, 2020; Faradonbeh et al., 2018; Wang and Janson, 2022).

Two previous works have focused on regret bounds for variants of the constrained LQR learning problem. Dean et al. (2019) and Li et al. (2021b) both consider the problem of constrained LQR learning specifically with bounded noise distributions. These works both give algorithms that achieve $\tilde{O}_T(T^{2/3})$ regret for this problem, and their regret results are with respect to the baseline of the best safe linear controller. While the results in these works hold in higher dimensions, our work improves on these results in two ways. The first is that the regret rate we achieve is with respect to the baseline of the best truncated linear controller, which is a strictly stronger (and often significantly stronger) baseline than the best safe linear controller. Furthermore, the regret of our algorithm is $\tilde{O}_T(\sqrt{T})$.

The most closely related work is Schiffer and Janson (2024), which proves a weaker regret bound that applies to any baseline class of controllers satisfying a set of assumptions. Specifically, Schiffer and Janson (2024) shows that for any baseline class of controllers satisfying certain natural but abstract assumptions, it is possible to achieve $\tilde{O}_T(T^{2/3})$ regret with respect to that baseline. That paper also shows that $\tilde{O}_T(\sqrt{T})$ is possible for such a baseline in the special case when the noise distribution has sufficiently large support. Importantly, however, that paper does not provide any concrete examples of baselines satisfying its assumptions, and therefore the regret improvements are hypothetical. This paper establishes just such a concrete example of a baseline class that satisfies these assumptions, namely, the class of truncated linear controllers. This class of controllers is well-adapted to safe LQR yet, due to its nonlinearity, presents a number of significant technical challenges (see Appendices C and D). Furthermore, our result in Theorem 2 is significantly stronger than simply applying the results in Schiffer and Janson (2024) to truncated linear controllers because our algorithm achieves regret of $\tilde{O}_T(\sqrt{T})$ all of the time rather than $\tilde{O}_T(T^{2/3})$ some of the time. Algorithm 2 requires a number of new technical ideas and tools developed in our paper that are specific to the class of truncated linear controllers (see Section 3 and Appendix E).

2. Preliminaries

2.1. Dynamics and Cost

Let T be the number of steps. For $t \in [T]$, we denote the state at time t as $x_t \in \mathbb{R}$ and the control at time t as $u_t \in \mathbb{R}$. Unless otherwise stated, we let $x_0 = 0$. Denote the (unknown) dynamics of the system as $\theta^* = (a^*, b^*) \in \mathbb{R}^2$. Then the state at time $t+1$ is $x_{t+1} = a^* x_t + b^* u_t + w_t$, where $w_t \overset{\text{i.i.d.}}{\sim} \mathcal{D}$ and $\mathcal{D}$ is a known continuous distribution with mean 0, variance $\sigma_{\mathcal{D}}^2 = 1$, cumulative distribution function $F_{\mathcal{D}}$, and bounded probability density function $f_{\mathcal{D}}$ (bounded by constant B_P). Note that the assumptions that the noise distribution is mean 0 and sub-gaussian are standard in LQR learning (Abbasi-Yadkori and Szepesvári, 2011; Li et al., 2021b; Dean et al., 2019). The assumption of unit variance is made only for expositional simplicity, and our main results still hold for noise distributions with arbitrary variances. Define $W = \{w_t\}_{t=0}^{T-1}$ as the set of noise random variables for the T steps. The goal of the algorithm is to minimize the total cost over all T steps, where the cost at time t is $qx_t^2 + ru_t^2$ for known $q, r \in \mathbb{R}_{>0}$. A controller C at time t chooses a control $u_t = C(H_t)$, where H_t is the history up to time t and is defined as $H_t := (x_0, u_0, ..., u_{t-1}, x_t)$. The average cost over T steps for controller C starting at state x_0 under dynamics θ is defined as

$$J(\theta, C, T, x_0, W) := \frac{1}{T}\left(qx_T^2 + \sum_{t=0}^{T-1} qx_t^2 + ru_t^2\right), \tag{1}$$

$$\text{where } u_t = C(H_t),\ \ x_{t+1} = ax_t + bu_t + w_t,\ \ w_t \overset{i.i.d.}{\sim} \mathcal{D}.$$

$J(\cdot)$ is an average cost, and therefore the total cost over T steps of controller C is $T \cdot J(\theta, C, T, x_0, W)$. We also define the expected cost of controller C as $J^*(\theta, C, T, x_0) = \mathbb{E}[J(\theta, C, T, x_0, W) \mid \theta, C, T, x_0]$. Finally, for ease of notation we define $J^*(\theta, C, T) = J^*(\theta, C, T, 0)$.

2.2. Safety Constraints

In this paper, we formulate safety as constraints on the expected state, which we show is *strictly more general* than just restricting the realized position as studied in previous works (Li et al., 2021b; Dean et al., 2019). More specifically, when the noise distribution is bounded, our safety definition is equivalent to the safety definition in Li et al. (2021b); Dean et al. (2019). However, our safety definition can generalize to unbounded noise distributions unlike the safety definitions in (Li et al., 2021a; Dean et al., 2019). Because w_t is a mean-0 random variable, we know that the conditional expectation of the next state given the current state and control is $\mathbb{E}[x_{t+1} \mid x_t, u_t] = a^* x_t + b^* u_t$. The safety constraints as defined in Definition 1 constrain this expected state to always stay within a known safe region between $D_{\text{L}}^{\mathbb{E}[x]}$ and $D_{\text{U}}^{\mathbb{E}[x]}$.

Definition 1 *Controls $\{u_t\}_{t=0}^{T-1}$ are safe for dynamics θ^* and boundaries $(D_{\text{L}}^{\mathbb{E}[x]}, D_{\text{U}}^{\mathbb{E}[x]})$ if for all t,*

$$D_{\text{L}}^{\mathbb{E}[x]} \le a^* x_t + b^* u_t \le D_{\text{U}}^{\mathbb{E}[x]}. \tag{2}$$

Similarly, a controller C is safe for dynamics θ^* and boundaries $(D_{\text{L}}^{\mathbb{E}[x]}, D_{\text{U}}^{\mathbb{E}[x]})$ if the resulting controls $\{C(H_t)\}_{t=0}^{T-1}$ under true dynamics θ^* are safe for dynamics θ^*.

Assumption 1 *The safety constraint boundaries satisfy that $D_{\text{L}}^{\mathbb{E}[x]} < 0 < D_{\text{U}}^{\mathbb{E}[x]}$, that $D_{\text{L}}^{\mathbb{E}[x]}, D_{\text{U}}^{E[x]} = O_T(1)$, and that $D_{\text{U}}^{\mathbb{E}[x]} - D_{\text{L}}^{\mathbb{E}[x]} \ge \frac{1}{\log(T)}$.*

The assumptions that the origin is in the safe set and that the boundaries are bounded above by constants are standard for safety-constrained LQR learning (Li et al., 2021b; Dean et al., 2019). These other works

consider a similar constrained LQR problem but require that the controller satisfies strict constraints on the state. In these works, the algorithm must choose controls such that for all t, $D_{\mathrm{L}}^x \le x_t \le D_{\mathrm{U}}^x$ for some $D_{\mathrm{L}}^x < 0 < D_{\mathrm{U}}^x$. However, these works also require that the noise distribution is bounded. When the noise distribution $\mathcal{D}$ is a bounded distribution (i.e. $\mathcal{D}$ satisfies $\bar{w}_{\mathrm{L}} := \inf_{w\sim\mathcal{D}} w > -\infty$ and $\bar{w}_{\mathrm{U}} := \sup_{w\sim\mathcal{D}} w < \infty$), then there is a one-to-one mapping between Definition 1 and strict state constraints. Formally, when $\mathcal{D}$ is bounded, the expected-state safety constraints in Definition 1 are equivalent to the strict state constraints that $D_{\mathrm{L}}^{\mathbb{E}[x]} - \bar{w}_{\mathrm{L}} \le x_t \le D_{\mathrm{U}}^{\mathbb{E}[x]} - \bar{w}_{\mathrm{U}}$ for all $t \in [T]$. Therefore, the expected-state constraint formulation is *strictly more general* than the safety formulation studied in previous works.

We study expected state constraints because they allow for more general noise distributions. For example, if $\mathcal{D}$ is normally distributed with mean 0 and variance 1, then $D_{\mathrm{L}}^x \le x_t \le D_{\mathrm{U}}^x$ is impossible to satisfy with probability $1 - o_T(1)$ for any constant $D_{\mathrm{L}}^x, D_{\mathrm{U}}^x$. Therefore, for distributions with unbounded support, the expected state constraints are a natural way to make the problem feasible. For notational simplicity, we use $D = (D_{\mathrm{L}}, D_{\mathrm{U}}) = (D_{\mathrm{L}}^{\mathbb{E}[x]}, D_{\mathrm{U}}^{\mathbb{E}[x]})$ to represent the bounds for the expected-state constraints.

2.3. Baseline Class

In both Li et al. (2021b) and Dean et al. (2019), the regret baseline for the $\tilde{O}_T(T^{2/3})$ results is the total cost of the best stationary linear controller of the form $u_t = -Kx_t$ that is safe for θ^* with probability 1. We will refer to the class of stationary linear controllers that are safe for θ^* with probability 1 as the class of safe linear controllers. Since not all linear controllers are safe for dynamics θ^*, this is restricted to K that will maintain safety for θ^* for any realization of the noise, and therefore can be a very weak baseline. For example, when D_{U} and D_{L} are not symmetric, the best linear controller must still behave symmetrically. Symmetric behavior may be far from optimal for D_{U} and D_{L} that are not symmetric, yet linear controllers lack the flexibility to behave asymmetrically. As another example, when the noise distribution is unbounded, there only exists a single safe linear controller (the $K = \frac{a^*}{b^*}$ controller).

To evaluate our algorithm, we use the baseline of the class of *truncated linear controllers*. The class of truncated linear controllers for $\theta = (a, b) \in \Theta$ is defined as $\mathcal{C}_{\mathrm{tr}}^\theta = \{C_K^\theta\}_{K\in[\frac{a-1}{b}, \frac{a}{b}]}$, where C_K^θ is defined as

$$C_K^\theta(x) = \begin{cases} -Kx & \text{if } D_{\mathrm{L}}^{\mathbb{E}[x]} \le (a-bK)x \le D_{\mathrm{U}}^{\mathbb{E}[x]} \\ \frac{D_{\mathrm{U}}^{\mathbb{E}[x]} - ax}{b} & \text{if } (a-bK)x > D_{\mathrm{U}}^{\mathbb{E}[x]} \\ \frac{D_{\mathrm{L}}^{\mathbb{E}[x]} - ax}{b} & \text{if } (a-bK)x < D_{\mathrm{L}}^{\mathbb{E}[x]}. \end{cases} \tag{3}$$

Note that every controller in the class of truncated linear controllers for dynamics θ is safe with probability 1 for dynamics θ. Furthermore, the class of truncated linear controllers for dynamics θ contains every linear controller that is with probability 1 safe for dynamics θ. Therefore, the class of truncated linear controllers is a strict superset of the class of safe linear controllers. We use the class of truncated linear controllers as a baseline because these controllers are computationally tractable while also being better suited for constrained LQR than standard linear controllers. For example, truncated linear controllers can effectively handle asymmetric constraints. As noted above, every controller in the baseline class $\mathcal{C}_{\mathrm{tr}}^{\theta^*}$ is safe, and therefore this is a fair baseline for our safe algorithm.

To evaluate our algorithm, we compare the total cost of the algorithm to the expected total cost of the best truncated linear controller when the dynamics are known. Define $K_{\mathrm{opt}}(\theta, T) := \arg\min_{K\in[\frac{a-1}{b}, \frac{a}{b}]} J^*(\theta, C_K^\theta, T)$. Then the expected total cost of the best truncated linear controller for dynamics θ^* is $\min_{C\in\mathcal{C}_{\mathrm{tr}}^{\theta^*}} T\cdot J^*(\theta^*, C, T) = T \cdot J^*(\theta^*, C_{K_{\mathrm{opt}}(\theta^*,T)}^{\theta^*}, T)$. Therefore, the regret of an algorithm with controller C_{alg} is

$$\mathrm{Regret}(C_{\mathrm{alg}}) := T \cdot J(\theta, C_{\mathrm{alg}}, T, 0, W) - T \cdot J^*(\theta^*, C_{K_{\mathrm{opt}}(\theta^*,T)}^{\theta^*}, T). \tag{4}$$

2.4. Initial Uncertainty

Without any prior knowledge about the unknown dynamics θ^*, it is impossible for any algorithm to satisfy Definition 1 for all possible $\theta^* \in \mathbb{R}^2$ for any non-trivial noise distribution. For example, if the noise is normally distributed, then with probability 1 any choice of control at time $t = 1$ will violate Definition 1 for some $\theta^* \in \mathbb{R}^2$. Therefore, we must make some assumptions about the initial uncertainty in θ^* in order for the problem to be feasible. As is standard in LQR learning problems (Abbasi-Yadkori and Szepesvári, 2011; Li et al., 2021b), we assume that there exists a known initial uncertainty set $\Theta \subseteq \mathbb{R}^2$ such that $\theta^* \in \Theta$.

Assumption 2 *There exists* $\Theta = \Theta_a \times \Theta_b = [\underline{a}, \bar{a}] \times [\underline{b}, \bar{b}]$ *such that* $\theta^* \in \Theta$ *and* $\bar{b} \geq \underline{b} > 0$ *and* $\bar{a} \geq \underline{a} > 0$.

We define the size of the initial uncertainty set Θ as $\operatorname{size}(\Theta) = \max(\bar{a} - \underline{a}, \bar{b} - \underline{b})$. Note that the assumption that $a^*, b^* > 0$ is made only to simplify the proofs, and the same results hold for general θ^* such that $b^* \neq 0$ ($b^* = 0$ corresponds to a degenerate case). In addition to assuming knowledge of Θ, we also assume access to a controller C^{init} that allows for some amount of initial safe exploration. As shown in Schiffer and Janson (2024), this assumption is asymptotically only slightly stronger than assuming that the problem is feasible. Furthermore, if the noise distribution $\mathcal{D}$ is bounded (with bound $\bar{w}$), then Assumption 3 holds for a simple linear controller C^{init} as long as Θ satisfies $\operatorname{size}(\Theta) \leq \frac{\min(D_{\text{U}}^{\mathbb{E}[x]}, D_{\text{L}}^{E[x]})}{2(1+\frac{\bar{a}}{\underline{b}}(\|D^{\mathbb{E}[x]}\|_\infty + \bar{w})}$.

Assumption 3 *There exists a known controller* C^{init} *such that* $\forall x \in \left[D_{\text{L}}^{\mathbb{E}[x]} + F_{\mathcal{D}}^{-1}(\frac{1}{T^4}), D_{\text{U}}^{\mathbb{E}[x]} + F_{\mathcal{D}}^{-1}(1 - \frac{1}{T^4})\right]$,

$$D_{\text{L}}^{\mathbb{E}[x]} + \frac{b^*}{\log(T)} \leq a^* x + b^* C^{\text{init}}(x) \leq D_{\text{U}}^{\mathbb{E}[x]} - \frac{b^*}{\log(T)}. \tag{5}$$

2.5. Problem Statement

Problem 1 (Safe LQR Learning) *Find an algorithm* C^{alg} *that takes as input* $D, \mathcal{D}, \Theta$, *and* T *that satisfy Assumptions 1–3, and achieves regret under linear dynamics with respect to baseline* $\mathcal{C}_{\text{tr}}^{\theta^*}$ *that is as low as possible, while also satisfying* $\sup_{\theta \in \Theta} \mathbb{P}\left(C^{\text{alg}} \text{ is safe with respect to } \theta\right) = 1 - o_T(1/T)$.

Informally, $\sup_{\theta \in \Theta} \mathbb{P}\left(C^{\text{alg}} \text{ is safe with respect to } \theta\right) = 1 - o_T(1/T)$ is equivalent to saying that for any $\theta \in \Theta$, if $\theta^* = \theta$ then using C^{alg} will result in a series of controls that satisfy Definition 1 with high probability. Note that in Problem 1, we only require that C_{alg} is safe with high probability rather than safe with probability 1. The reason for this is that requiring safety with probability 1 would mean that C^{alg} is unable to use any conclusions about θ^* learned from the history that do not hold with probability 1. For example in the case of unbounded noise distributions, making any statement about θ^* from historical data that holds with probability 1 is impossible. Therefore, we allow a vanishing $o_T(1/T)$ probability of the algorithm not being safe to allow the algorithm to use historical information when choosing safe controls. Note that the choice of $o_T(1/T)$ is made for expositional purposes, and an equivalent result holds when $1 - o_T(1/T)$ is replaced with $1 - \delta$ for $\delta < 1$. Throughout this paper, we use $O_T(\cdot)$ and other big-O notation to represent equations that hold for sufficiently large T, where equations with $O_T(\cdot)$ hold for sufficiently large T and contain unwritten constants that are independent of T and any other variables included in the parentheses. For expositional purposes in the proofs, we will also assume that $\log(T^{1/12})$ is an integer, however simple modifications to the algorithm allow the same result to hold for all T. More discussion of notation and definitions can be found in Appendix A.

3. Theoretical Results

We now formally state our main result and provide some general intuition for the proof and algorithm. We present a more detailed proof sketch of Theorem 2 in Section 4 and full proof in Appendix E.

Theorem 2 *Algorithm 2 with probability* $1 - o_T(1/T)$ *achieves* $\tilde{O}_T(\sqrt{T})$ *regret with respect to baseline* $\mathcal{C}_{\text{tr}}^{\theta^*}$ *while also satisfying* $\sup_{\theta \in \Theta} \mathbb{P}\left(C^{\text{alg}} \text{ is safe with respect to } \theta\right) = 1 - o_T(1/T)$.

The intuition of Algorithm 2 is outlined in Algorithm 1. The algorithm first explores for $\tilde{\Theta}_T(\sqrt{T})$ steps using C^{init} from Assumption 3. Using the data from this exploration, the algorithm calculates a regularized least-squares estimate of θ^* (denoted $\hat{\theta}_{\text{wu}}$) that is accurate up to $\tilde{O}_T(T^{-1/4})$. Based on this least-squares estimate, the algorithm then decides if the support of the noise distribution $\mathcal{D}$ is small or large relative to the constraint boundary D. In the small noise case, the algorithm uses the best unconstrained controller for dynamics $\hat{\theta}_{\text{wu}}$ with small modifications to the control as needed to guarantee constraint satisfaction with high probability. Because the noise is small in this case, the modification is only needed a small fraction of the time. Therefore, in this case the regret of the algorithm is only slightly more than the regret of the optimal unconstrained controller for $\hat{\theta}_{\text{wu}}$, which can be shown to be $\tilde{O}_T(\sqrt{T})$ using standard certainty equivalence results. In the large noise case, the algorithm uses a truncated certainty equivalence approach that guarantees $\tilde{O}_T(\sqrt{T})$ regret with high probability. Intuitively, in this case the noise is large enough to force the algorithm to a constant fraction of the time be non-linear by a constant amount. This non-linearity allows the algorithm to learn the unknown dynamics at a faster rate of $1/\sqrt{t}$, which leads to regret of $\tilde{O}_T(\sqrt{T})$.

Algorithm 1 Outline of Algorithm 2 for proof of Theorem 2

Explore for $\tilde{\Theta}_T(\sqrt{T})$ steps using controller C^{init} from Assumption 3.
$\hat{\theta}_{\text{wu}} \leftarrow$ regularized least-squares estimate of θ^*.
Using $\hat{\theta}_{\text{wu}}$, determine if support of noise distribution $\mathcal{D}$ is large or small relative to boundary D.
if *support of* $\mathcal{D}$ *is* ***small*** *relative to* D **then**
  For the rest of the steps, use the optimal unconstrained linear controller for dynamics $\hat{\theta}_{\text{wu}}$ with small modifications to the control as necessary to enforce constraint satisfaction w.h.p.
if *support of* $\mathcal{D}$ *is* ***large*** *relative to* D **then**
1 **for** $s \in [0 : \log(\sqrt{T}) - 1]$ **do**
    $\hat{\theta}_s \leftarrow$ regularized least-squares estimate of θ^* using data seen so far
    $\epsilon_s \leftarrow$ high probability bound on $\|\theta^* - \hat{\theta}_s\|_\infty$
    $C_s^{\text{alg}} \leftarrow$ optimal truncated linear controller for dynamics $\hat{\theta}_s$
2 For $\sqrt{T}2^s$ steps, use controller C_s^{alg} truncated to be safe for all θ satisfying $\|\theta - \hat{\theta}_s\|_\infty \le \epsilon_s$

In proving Theorem 2, we show that the class of truncated linear controllers satisfies two natural assumptions of continuity, formalized in the following two lemmas. Informally, Lemma 2 says that the cost of the optimal truncated linear controller is Lipschitz continuous in the dynamics. Therefore, using the optimal controller for dynamics θ that are close to the true dynamics θ^* does not incur significantly higher cost.

Lemma 2 *There exists* $\epsilon_{\text{L2}} = \tilde{\Omega}_T(1)$ *such that for any* $\|\theta - \theta^*\|_\infty \le \epsilon_{\text{L2}}$ *and* $t \le T$,

$$|J^*(\theta^*, C^{\theta}_{K_{\text{opt}}(\theta,t)}, t) - J^*(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}, t)| \le \tilde{O}_T\left(\|\theta - \theta^*\|_\infty + \frac{1}{T^2}\right).$$

The proof of Lemma 2 can be found in Appendix D. Next, informally, Lemma 3 says that the cost of using a truncated linear controller is Lipschitz continuous in the starting state. Therefore, if $|x - y|$ is sufficiently small, then the difference in total cost of starting at x versus y is linear in $|x - y|$.

Lemma 3 *There exist* $\epsilon_{\text{L3}}, \delta_{\text{L3}} = \tilde{\Omega}_T(1)$ *such that for any* θ *satisfying* $\|\theta - \theta^*\|_\infty \le \epsilon_{\text{L3}}$ *the following holds. For* $t < T$, *let* $W' = \{w_i\}_{i=0}^{t-1}$. *Then for any* $K \in [\frac{a-1}{b}, \frac{a}{b}]$, *there exists a set* $\mathcal{Y}_{\text{L3}} \in \mathbb{R}^t$ *that depends only on* C_K^θ *such that the following holds. Define* $E_{\text{L3}}\left(C_K^\theta, W'\right)$ *as the event that* $W' \in \mathcal{Y}_{\text{L3}}$. *Then*

$\mathbb{P}(E_{\mathrm{L3}}(C_K^\theta, W')) \geq 1 - o_T(1/T^{10})$ *and for any* $|x|, |y| \leq 4\log^2(T)$ *such that* $|x-y| \leq \delta_{\mathrm{L3}}$*, conditional on event* $E_{\mathrm{L3}}(C_K^\theta, W')$*,*

$$\left| t \cdot J(\theta^*, C_K^\theta, t, x, W') - t \cdot J(\theta^*, C_K^\theta, t, y, W') \right| \leq \tilde{O}_T(|x-y| + \|\theta - \theta^*\|_\infty). \tag{6}$$

The proof of Lemma 3 can be found in Appendix C. As discussed above, truncated linear controllers are a natural extension of linear controllers better suited for problems with safety constraints. Because truncated linear controllers are not linear, the analysis of this class requires new theoretical tools (see Appendices D and C). These results may be independently interesting in that non-linear controllers have not been well-studied in this setting and therefore little was previously known about properties of such controller classes.

4. Proof Sketch of Theorem 2

The full proof of Theorem 2 can be found in Appendix E. Before presenting the algorithm, we need additional notation. Define $\mathcal{C}^{\mathrm{unc}} = \{C_K^{\mathrm{unc}}\}_{K \in \mathbb{R}}$ as the class of untruncated linear controllers, so $C_K^{\mathrm{unc}}(x) = -Kx$. For any controller C and dynamics θ, define $J^*(\theta, C) = \lim_{T \to \infty} J^*(\theta, C, T)$. Define $K_{\mathrm{opt}}(\theta) = \arg\max_K J^*(\theta, C_K^\theta)$ and $F_{\mathrm{opt}}(\theta) = \arg\max_K J^*(\theta, C_K^{\mathrm{unc}})$. Finally, define $C_{\mathrm{switch}} = \frac{c_{\mathrm{E81}} D_{\mathrm{U}}}{c_{\mathrm{L39}}^2}$ where c_{E81} is defined in Equation (81) and satisfies $c_{\mathrm{E81}} = \tilde{O}_T(1)$, and c_{L39} is defined in Lemma 39 and satisfies $c_{\mathrm{L39}} = \Omega_T(1)$. Knowing the exact values of these constants is not crucial for understanding the theoretical ideas presented in the algorithm, as they depend only on known problem parameters and give a resulting constant C_{switch} that is $\tilde{O}_T(1)$. The algorithm that achieves the regret bound of Theorem 2 is Algorithm 2.

Algorithm 2 Truncated Linear Controller Safe LQR

Input: $D, \mathcal{D}, \Theta, C^{\mathrm{init}}, T, \lambda$
for $t \leftarrow 0$ *to* $\sqrt{T} - 1$ **do**
  $\phi_t \sim \mathrm{Rademacher}(0.5)$ Use control $u_t = C^{\mathrm{init}}(x_t) + \frac{\phi_t}{\log(T)}$
$\hat{\theta}_{\mathrm{wu}} \leftarrow (Z_{\sqrt{T}}^\top Z_{\sqrt{T}} + \lambda I)^{-1} Z_{\sqrt{T}}^\top X_{\sqrt{T}}$
for $s \leftarrow 0$ *to* $\log_2(\sqrt{T}) - 1$ **do**
  $T_s \leftarrow 2^s \sqrt{T}$
  $\epsilon_s \leftarrow B_{T_s} \sqrt{\frac{\max(V_{T_s}^{22}, V_{T_s}^{11})}{V_{T_s}^{11} V_{T_s}^{22} - (V_{T_s}^{12})^2}}$
  $\hat{\theta}_s^{\mathrm{pre}} \leftarrow (Z_{T_s}^\top Z_{T_s} + \lambda I)^{-1} Z_{T_s}^\top X_{T_s}$
  $\hat{\theta}_s \leftarrow \arg\max_{\|\theta - \hat{\theta}_s^{\mathrm{pre}}\| \leq \epsilon_s} a - bK_{\mathrm{opt}}(\theta)$
  $C_s^{\mathrm{alg}} \leftarrow \begin{cases} C_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}^{\mathrm{unc}} & \text{if } \bar{w} + D_{\mathrm{U}} - \frac{D_{\mathrm{U}}}{\hat{a}_{\mathrm{wu}} - \hat{b}_{\mathrm{wu}} F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})} \leq C_{\mathrm{switch}} T^{-1/4} \\ C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s} & \text{otherwise} \end{cases}$
  for $t \leftarrow T_s$ *to* $2T_s - 1$ **do**
    if $\bar{w} + D_{\mathrm{U}} - \frac{D_{\mathrm{U}}}{\hat{a}_{\mathrm{wu}} - \hat{b}_{\mathrm{wu}} F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})} \leq C_{\mathrm{switch}} T^{-1/4}$ **then**
      $u_t^{\mathrm{safeU}}, u_t^{\mathrm{safeL}} \leftarrow \max\left\{u : \max_{\|\theta - \hat{\theta}^{\mathrm{wu}}\|_\infty \leq \epsilon_0} ax_t + bu \leq D_{\mathrm{U}}\right\}, \min\left\{u : \min_{\|\theta - \hat{\theta}^{\mathrm{wu}}\|_\infty \leq \epsilon_0} ax_t + bu \geq D_{\mathrm{L}}\right\}$
    else
      $u_t^{\mathrm{safeU}}, u_t^{\mathrm{safeL}} \leftarrow \max\left\{u : \max_{\|\theta - \hat{\theta}_s\|_\infty \leq \epsilon_s} ax_t + bu \leq D_{\mathrm{U}}\right\}, \min\left\{u : \min_{\|\theta - \hat{\theta}_s\|_\infty \leq \epsilon_s} ax_t + bu \geq D_{\mathrm{L}}\right\}$
      Use control $u_t = \max\left(\min\left(C_s^{\mathrm{alg}}(x_t), u_t^{\mathrm{safeU}}\right), u_t^{\mathrm{safeL}}\right)$

Algorithm 2 High-Level Intuition The main intuition behind the proof of Theorem 2 is to design an algorithm that combines the faster learning rates under tight constraints from Schiffer and Janson (2024) with the observation that $\tilde{O}_T(\sqrt{T})$ regret is possible in unconstrained LQR learning with unknown dynamics. Algorithm 2 is broken into two phases. The first phase is a warm-up exploration phase that allows the algorithm to learn about the unknown dynamics quickly but potentially incurs high per-step cost. The second phase splits the choice of C_s^{alg} into two cases (Line 6) depending on the estimated dynamics ($\hat{\theta}_{\text{wu}}$) at the end of the first phase. The first case in Line 6 corresponds to when the support of the noise is sufficiently small so that we can bound the regret of the algorithm using the observation that $\tilde{O}_T(\sqrt{T})$ regret is possible in the unconstrained setting. More specifically, this case is when the boundaries are far enough away from the origin compared to the magnitude of the noise, and therefore the algorithm can use a controller very close to the optimal unconstrained controller. The second case in Line 6 corresponds to when the support of the noise is sufficiently large so that learn the unknown dynamics at a faster rate. More specifically, in this case we argue that the uncertainty bound ϵ_s will decrease at a rate of $\tilde{O}_T(1/\sqrt{T_s})$. We give more details on the $\tilde{O}_T(\sqrt{T})$ regret of these two cases separately below.

Sufficiently small noise case In this case, we let $C_s^{\text{alg}} = C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$, i.e. the optimal unconstrained controller based on the data in the warm-up period. First, we show that the controller $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ has $\tilde{O}_T(\sqrt{T})$ more expected total cost for T_s steps than the baseline controller $C^{\theta^*}_{K_{\text{opt}}(\theta^*,T_s)}$. Intuitively, this follows from the fact that $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ has similar expected cost to the best infinite-time unconstrained controller for θ^*, and the best infinite-time controller and the best finite-time controller for T_s steps have similar expected cost. Because $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ is an unconstrained linear controller, we can also show that the realized total cost of using this controller concentrates to within $\tilde{O}_T(\sqrt{T})$ of the expected total cost with high probability.

The last (and most subtle) part of this case is to show that enforcing safety in Line 8 only contributes $\tilde{O}_T(\sqrt{T})$ regret. This is where we use the fact that $\bar{w} + D_{\text{U}} - \frac{D_{\text{U}}}{\hat{a}_{\text{wu}} - \hat{b}_{\text{wu}} F_{\text{opt}}(\hat{\theta}_{\text{wu}})} \leq C_{\text{switch}} T^{-1/4}$. When this equation holds, the probability that the algorithm uses control $u_t = u_t^{\text{safeU}}$ or $u_t = u_t^{\text{safeL}}$ is at most $\tilde{O}_T(T^{-1/4})$ for any t. Furthermore, each time these controls are used, the extra cost compared to using control $u_t = C_s^{\text{alg}}(x_t)$ is $\tilde{O}_T(T^{-1/4})$. Combining these two facts, the total extra regret from using controls u_t^{safeU} or u_t^{safeL} is $\tilde{O}_T(\sqrt{T})$ with probability $1 - o_T(1/T)$. The warm-up period has regret of $\tilde{O}_T(\sqrt{T})$ with probability $1 - o_T(1/T)$ because the algorithm is safe with high probability and the length of warm-up is $\sqrt{T}$. Putting this all together, with probability $1 - o_T(1/T)$, the total regret of the algorithm in this case is $\tilde{O}_T(\sqrt{T})$. For more details on this case and a formal breakdown of each source of regret, see Appendix H.

Sufficiently large noise case In this case, we have that $C_s^{\text{alg}} = C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}$. To prove that the regret is $\tilde{O}_T(\sqrt{T})$ in this case, we will show that with probability $1 - o_T(1/T)$, the uncertainty bound satisfies $\epsilon_s = \tilde{O}_T(1/\sqrt{T_s})$ for every s. We use a variant of Lemma 21 in Schiffer and Janson (2024), which informally says that ϵ_s is upper bounded by $\tilde{O}_T(1/\sqrt{|S_{T_s}|})$ with probability $1 - o_T(1/T)$, where $|S_{T_s}|$ is roughly the number of times $t < T_s$ that the algorithm uses control u_t^{safeU}. To bound regret with this lemma, we show that with probability $1 - o_T(1/T)$, we have $|S_{T_s}| \geq \Omega_T(T_s)$ for all s. When using the controller $C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}$, there exist constants $\epsilon, d_\epsilon > 0$ such that at every time step t when the control is not u_t^{safeU}, there is an ϵ probability that the state increases by d_ϵ. Informally, this says that at every step, either $u_t = u_t^{\text{safeU}}$ or the state will increase by a constant amount with a constant probability. Therefore, because D is a constant relative to T, we have that with high probability, every $\Omega(1)$ steps the state will exceed $P(\theta^*, K_{\text{opt}}(\hat{\theta}_s), D_{\text{U}})$ or there will be a t such that $u_t = u_t^{\text{safeU}}$. The control at any time t where $x_t \geq P(\theta^*, K_{\text{opt}}(\hat{\theta}_s), D_{\text{U}})$ is $u_t = u_t^{\text{safeU}}$. Therefore, with high probability every $\Omega(1)$ steps there will exist a t such that the algorithm uses control $u_t = u_t^{\text{safeU}}$. This implies that $|S_{T_s}| \geq \Omega(T_s)$ for every s with high probability and therefore with probability $1 - o_T(1/T)$, $\epsilon_s \leq \tilde{O}_T(1/\sqrt{T_s})$.

Once we have the above bound on ϵ_s, all that is left is bounding each source of regret. The first source of regret is the regret from using certainty equivalence, i.e. using $\hat{\theta}_s$ instead of using θ^* in finding C_s^{alg}. Applying Lemma 2 gives that the expected cost of using $C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}$ instead of $C^{\theta^*}_{K_{\text{opt}}(\theta^*)}$ is $\tilde{O}_T(T_s\|\hat{\theta}_s - \theta^*\|_\infty + 1/T)$. Because $\|\hat{\theta}_s - \theta^*\|_\infty \leq \epsilon_s \leq \tilde{O}_T(1/\sqrt{T_s})$ with high probability, this source of regret is $\tilde{O}_T(\sqrt{T})$ with high probability. The second source of regret is the regret from randomness in the regret random variable, which can be bounded by $\tilde{O}_T(\sqrt{T})$ by a variant of McDiarmids Inequality. The third source of regret is the regret of enforcing safety with u_t^{safeU} and u_t^{safeL} in the choice of u_t. By construction u_t differs from C_s^{alg} by $\tilde{O}_T(\epsilon_s) = \tilde{O}_T(1/\sqrt{T_s})$ at every time step. Therefore by Lemma 3, the regret of enforcing safety by using u_t is $\tilde{O}_T(\sqrt{T})$ with high probability. The warm-up period has regret $\tilde{O}_T(\sqrt{T})$ as in the small noise case. Finally, there is one additional component of regret in this proof, as we are using the best infinite time controller rather than the best T_s-step controller in round s. However, we can show that this only adds at most $\tilde{O}_T(\sqrt{T})$ extra cost, and therefore the total regret is with probability $1 - o_T(1/T)$ still $\tilde{O}_T(\sqrt{T})$. For more details on this case and a formal breakdown of each source of regret, see Appendix F.

Proof Sketch of Lemmas 2 and 3 While both of these properties are easy to show for the class of linear controllers, proving them for the class of truncated linear controllers is significantly more complicated. We first outline the proof of Lemma 3. Lemma 3 compares the cost of two trajectories when using truncated linear controller $C^{\theta}_{K_{\text{opt}}(\theta,t)}$, one trajectory starting at state x and the other trajectory starting at state $x+\delta$. In the proof of Lemma 3, we show that the difference in states of the two trajectories will decrease at most (but not all) time steps. The difference does not decrease at all time steps because the difference between $\hat{\theta}$ and θ^* leads to low probability events where the difference between the states of the two trajectories increases. We are able to bound the probability of the event that the difference in state increases, and this gives the desired result. For Lemma 2, we first show that the truncated linear controller $C^{\theta}_{K_{\text{opt}}(\theta,t)}$ under dynamics θ has only $\tilde{O}_T(\|\theta - \theta^*\|_\infty)$ more cost than the truncated linear controller $C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}$ under dynamics θ^*. We then show that for any K, the truncated linear controller C_K^{θ} under dynamics θ^* for t steps has only $\tilde{O}_T(\|\theta - \theta^*\|_\infty)$ more cost than C_K^{θ} under dynamics θ for t steps. Combining these two results directly gives the desired result of Lemma 2.

5. Discussion

In this section we discuss a few limitations of our results and some open questions. To prove Theorem 2, we introduced several new tools for technical analysis of non-linear controllers. We believe that many of our results and techniques will generalize to higher dimensions, but due to the already very complex nature of the proofs in one-dimension (and already very long appendix), this is beyond the scope of the current paper. That being said, in Appendix J we give a high-level discussion on how we believe that Theorem 2 can extend to higher dimensions with a series of unproven conjectures. We first discuss how to extend truncated linear controllers and Algorithm 2 and then provide a simple symmetrical setting in which we believe the ideas of Theorem 2 will easily extend. See Appendix J for more details.

In this work, we also focus on state constraints rather than constraints on the actions. We expect that very minor modifications to Algorithm 2 will naturally extend these results to also apply to the setting where the controls u_t must satisfy constraints. More specifically, we would need to choose C_s^{alg} in Algorithm 2 to only choose controls that satisfy control constraints with an extra buffer of $\tilde{\Theta}_T(\epsilon_s)$. See Schiffer and Janson (2024) for more details on how results regarding state constraints can generalize to problems with control constraints as well. As discussed above, the question of whether $\tilde{O}_T(\sqrt{T})$ regret is possible for all noise distributions in higher dimensions is an open question for future work. In this paper, we also introduced the class of truncated linear controllers and proved some desirable properties of this class of controllers. We expect these properties to still hold in higher dimensions, but we leave formal study of this to future work.

Acknowledgments

The authors would like to thank Na Li and Shahriar Talebi for helpful discussions. B.S. and L.J. received funding from NSF grant CBET-2112085 and B.S. received funding from the National Science Foundation Graduate Research Fellowship grant DGE 2140743.

References

Yasin Abbasi-Yadkori and Csaba Szepesvári. Regret bounds for the adaptive control of linear quadratic systems. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 1–26. JMLR Workshop and Conference Proceedings, 2011.

Marc Abeille and Alessandro Lazaric. Thompson sampling for linear-quadratic control problems. In *Artificial intelligence and statistics*, pages 1246–1254. PMLR, 2017.

Naman Agarwal, Brian Bullins, Elad Hazan, Sham Kakade, and Karan Singh. Online control with adversarial disturbances. In *International Conference on Machine Learning*, pages 111–119. PMLR, 2019.

Brian DO Anderson and John B Moore. *Optimal control: linear quadratic methods*. Courier Corporation, 2007.

Archith Athrey, Othmane Mazhar, Meichen Guo, Bart De Schutter, and Shengling Shi. Regret analysis of learning-based linear quadratic gaussian control with additive exploration. In *2024 European Control Conference (ECC)*, pages 1795–1801. IEEE, 2024.

Alberto Bemporad and Manfred Morari. Robust model predictive control: A survey. In *Robustness in identification and control*, pages 207–226. Springer, 2007.

Alberto Bemporad, Manfred Morari, Vivek Dua, and Efstratios N Pistikopoulos. The explicit linear quadratic regulator for constrained systems. *Automatica*, 38(1):3–20, 2002.

Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3387–3395, 2019.

Alon Cohen, Tomer Koren, and Yishay Mansour. Learning linear-quadratic regulators efficiently with only sqrtt regret. pages 1300–1309, 2019.

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. Regret bounds for robust adaptive control of the linear quadratic regulator. *Advances in Neural Information Processing Systems*, 31, 2018.

Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht. Safely learning to control the constrained linear quadratic regulator. In *2019 American Control Conference (ACC)*, pages 5582–5588. IEEE, 2019.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. On optimality of adaptive linear-quadratic regulators. *arXiv preprint arXiv:1806.10749*, 2018.

Charles Fefferman, Bernat Guillén Pegueroles, Clarence W Rowley, and Melanie Weber. Optimal control with learning on the fly: a toy problem. *Revista matemática iberoamericana*, 38(1):175–187, 2021.

Jaime F Fisac, Anayo K Akametalu, Melanie N Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Transactions on Automatic Control*, 64(7):2737–2752, 2018.

Nathan Fulton and André Platzer. Safe reinforcement learning via formal methods: Toward safe control through proof and learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Milan Ganai, Zheng Gong, Chenning Yu, Sylvia Herbert, and Sicun Gao. Iterative reachability estimation for safe reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.

Kunal Garg, Songyuan Zhang, Oswin So, Charles Dawson, and Chuchu Fan. Learning safe control for multi-robot systems: Methods, verification, and open challenges. *Annual Reviews in Control*, 57:100948, 2024.

Shangding Gu, Long Yang, Yali Du, Guang Chen, Florian Walter, Jun Wang, and Alois Knoll. A review of safe reinforcement learning: Methods, theory and applications. *arXiv preprint arXiv:2205.10330*, 2022.

Elad Hazan and Karan Singh. Introduction to online nonstochastic control. *arXiv preprint arXiv:2211.09619*, 2022.

Mohammad Khosravi and Roy S Smith. Nonlinear system identification with prior knowledge on the region of attraction. *IEEE Control Systems Letters*, 5(3):1091–1096, 2020.

Johannes Köhler, Elisa Andina, Raffaele Soloperto, Matthias A Müller, and Frank Allgöwer. Linear robust adaptive model predictive control: Computational complexity and conservatism. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 1383–1388. IEEE, 2019.

Bruce Lee, Anders Rantzer, and Nikolai Matni. Nonasymptotic regret analysis of adaptive linear quadratic control with model misspecification. In *6th Annual Learning for Dynamics & Control Conference*, pages 980–992. PMLR, 2024.

Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1):1334–1373, 2016.

Yingying Li, Subhro Das, and Na Li. Online optimal control with affine constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 8527–8537, 2021a.

Yingying Li, Subhro Das, Jeff Shamma, and Na Li. Safe adaptive learning-based control for constrained linear quadratic regulators with regret guarantees. *arXiv preprint arXiv:2111.00411*, 2021b.

Yingying Li, Tianpeng Zhang, Subhro Das, Jeff Shamma, and Na Li. Non-asymptotic system identification for linear systems with nonlinear policies. *arXiv preprint arXiv:2306.10369*, 2023.

Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Matthias Lorenzen, Mark Cannon, and Frank Allgöwer. Robust mpc with recursive model update. *Automatica*, 103:461–471, 2019.

Xiaonan Lu, Mark Cannon, and Denis Koksal-Rivet. Robust adaptive model predictive control: Performance and parameter estimation. *International Journal of Robust and Nonlinear Control*, 31(18):8703–8724, 2021.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. *Advances in Neural Information Processing Systems*, 32, 2019.

Horia Mania, Michael I Jordan, and Benjamin Recht. Active learning for nonlinear system identification with guarantees. *arXiv preprint arXiv:2006.10277*, 2020.

Zahra Marvi and Bahare Kiumarsi. Safe reinforcement learning: A control barrier function optimization approach. *International Journal of Robust and Nonlinear Control*, 31(6):1923–1940, 2021.

Colin McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.

Ali Mesbah. Stochastic model predictive control: An overview and perspectives for future research. *IEEE Control Systems Magazine*, 36(6):30–44, 2016.

Teodor Mihai Moldovan and Pieter Abbeel. Safe exploration in markov decision processes. *arXiv preprint arXiv:1205.4810*, 2012.

Deepan Muthirayan, Jianjun Yuan, Dileep Kalathil, and Pramod P Khargonekar. Online learning for predictive control with provable regret guarantees. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 6666–6671. IEEE, 2022.

Frauke Oldewurtel, Colin N Jones, and Manfred Morari. A tractable approximation of chance constrained stochastic mpc based on affine disturbance feedback. In *2008 47th IEEE conference on decision and control*, pages 4731–4736. IEEE, 2008.

J.B. Rawlings and D.Q. Mayne. *Model Predictive Control: Theory and Design*. Nob Hill Pub., 2009. ISBN 9780975937709.

Alicia Arce Rubio, Alexandre Seuret, Yassine Ariba, and Alessio Mannisi. Optimal control strategies for load carrying drones. *Delays and Networked Control Systems*, pages 183–197, 2016.

Yahya Sattar and Samet Oymak. Non-asymptotic and accurate learning of nonlinear dynamical systems. *The Journal of Machine Learning Research*, 23(1):6248–6296, 2022.

Benjamin Schiffer and Lucas Janson. Foundations of safe online reinforcement learning in the linear quadratic regulator: Generalized baselines. *arXiv preprint arXiv:2410.21081*, 2024.

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

Max Simchowitz, Horia Mania, Stephen Tu, Michael I Jordan, and Benjamin Recht. Learning without mixing: Towards a sharp analysis of linear system identification. In *Conference On Learning Theory*, pages 439–473. PMLR, 2018.

Yue Sun, Samet Oymak, and Maryam Fazel. Finite sample system identification: Optimal rates and the role of regularization. In *Learning for dynamics and control*, pages 16–25. PMLR, 2020.

Ambuj Tewari and Susan A Murphy. From ads to interventions: Contextual bandits in mobile health. *Mobile health: sensors, analytic methods, and applications*, pages 495–517, 2017.

Akifumi Wachi, Yanan Sui, Yisong Yue, and Masahiro Ono. Safe exploration and optimization of constrained mdps using gaussian processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Akifumi Wachi, Xun Shen, and Yanan Sui. A survey of constraint formulations in safe reinforcement learning. *arXiv preprint arXiv:2402.02025*, 2024.

Feicheng Wang and Lucas Janson. Exact asymptotics for linear quadratic adaptive control. *The Journal of Machine Learning Research*, 22(1):12136–12247, 2021.

Feicheng Wang and Lucas Janson. Rate-matching the regret lower-bound in the linear quadratic regulator with unknown dynamics. *arXiv preprint arXiv:2202.05799*, 2022.

Yihang Yao, Zuxin Liu, Zhepeng Cen, Jiacheng Zhu, Wenhao Yu, Tingnan Zhang, and Ding Zhao. Constraint-conditioned policy optimization for versatile safe reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.

Lintao Ye, Ming Chi, Zhi-Wei Liu, and Vijay Gupta. Online actuator selection and controller design for linear quadratic regulation with unknown system model. *IEEE Transactions on Automatic Control*, 2024.

Zichen Zhao and Qianxiao Li. Adaptive sampling methods for learning dynamical systems. In *Mathematical and Scientific Machine Learning*, pages 335–350. PMLR, 2022.

Yang Zheng and Na Li. Non-asymptotic identification of linear dynamical systems using multiple trajectories. *IEEE Control Systems Letters*, 5(5):1693–1698, 2020.

Ingvar Ziemann and Henrik Sandberg. Regret lower bounds for learning linear quadratic gaussian systems. *IEEE Transactions on Automatic Control*, 2024.

Outline of Appendix

In this section, we will present a brief outline of the organization of the appendix. While the appendix is relatively long, the length is in part due to the fact that we chose to provide detailed line-by-line proofs of some algebraically complicated results. Many of the lemmas and theorems should be intuitively true to the reader, but we wanted to provide rigorous details of all of the results.

Because we are analyzing non-linear controllers, many of the proofs are extremely algebra and notation heavy due to the fact that these controllers cannot be analyzed with standard approaches. We tried as hard as possible to simplify the proofs and make them more readable. In each lemma, we also include line-by-line logic for every algebraic derivation, with the hope of making each individual proof as easy to follow as possible.

The proofs of the main results are written in a completely modular way – so the main proofs are relatively short and easy to read, and then key ideas are presented as lemmas which are proven in later sections. For example, the actual proof of Lemma 2 can be found in the "Main Proof" subsection of Section D, and then the two key ideas are broken down in the following two subsections of Section D. Note that any appendix section that ends with (Lemma/Proposition x) contains the proof of the corresponding Lemma or Proposition.

Contents

Appendix A. Notation

A.1. Equation Notation

Throughout this paper, we use notation such as $o_T(\cdot)$, $O_T(\cdot)$, $\omega_T(\cdot)$, $\Omega_T(\cdot)$, where the subscript T highlights that these equations hold for sufficiently large T. The following ways we use O-notation are relatively standard, and we include them here for completeness. We also use Ω-notation that is defined equivalently in the other direction. When using this notation, the functions $f(T)$ and $g(t)$ will always be non-negative.

- $f(T) = O_T(g(T))$ if there exists T_0 and $M \in \mathbb{R}$ such that for $T \geq T_0$, $f(T) \leq M \cdot g(T)$.
- $f(T) = \Omega_T(g(T))$ if there exists T_0 and $M \in \mathbb{R}$ such that for $T \geq T_0$, $f(T) \geq M \cdot g(T)$.

- $f(T) = o_T(g(T))$ if for every constant $\epsilon > 0$ there exists T_0 such that for all $T \geq T_0$, $f(T) \leq \epsilon \cdot g(T)$.
- $f(T) = \omega_T(g(T))$ if for every constant $\epsilon > 0$ there exists T_0 such that for all $T \geq T_0$, $f(T) \geq \epsilon \cdot g(T)$.
- $f(T) = \tilde{O}_T(g(T))$ if there exists T_0 and $k, M \in \mathbb{R}$ such that for $T \geq T_0$, $f(T) \leq M \cdot g(T) \cdot \log^k(T)$.

Whenever equations or inequalities involve random variables, the results hold with almost surely unless specified otherwise.

A.2. Problem Specifications

Below is a (non-exhaustive) list of notation used throughout the appendix, much of which is similar to Schiffer and Janson (2024).

- q, r : coefficients for the cost at time t of $qx_t^2 + ru_t^2$.
- $W = \{w_t\}_{t=0}^{T-1}$: The noise random variables for the T-length trajectory.
- $\mathcal{D}$: Distribution of w_t
  - B_P : Upper bound on the density of $\mathcal{D}$
  - $F_{\mathcal{D}}$: Cumulative Density Function (CDF) of $\mathcal{D}$
  - $\bar{w}$: the bound of $\mathcal{D}$ when the distribution is bounded.
- $\Theta = [\underline{a}, \bar{a}] \times [\underline{b}, \bar{b}]$: The given initial set of dynamics such that $\theta^* \in \Theta$ and $\text{size}(\Theta) = \min(\bar{a} - \underline{a}, \bar{b} - \underline{b})$
- $\theta^* = (a^*, b^*)$: The true (unknown) dynamics.
- C^{init} : The initial safe controller satisfying Assumption 2.
- $D = (D_{\text{L}}, D_{\text{U}})$: the expected-state boundary for the safety constraint.
- A set of controls $\{u_t\}$ are safe for dynamics $\{\theta_t\}$ if for all t, $D_{\text{L}} \leq a_t x_t + b_t u_t \leq D_{\text{U}}$.
- $H_t = (x_0, u_0, x_1, u_1, ..., u_{t-1}, x_t)$ and $\mathcal{F}_t = \sigma(H_t)$.
- $J(\theta, C, T, x, W)$: The random variable cost of using controller C starting at state $x_0 = x$ for T time steps under dynamics θ with noise random variables W.
- $J^*(\theta, C, T) = J^*(\theta, C, T, 0) = \mathbb{E}[J(\theta, C, T, x, W) \mid \theta, C, T, x]$ and $J^*(\theta, C, T) = J^*(\theta, C, T, 0)$.
- $J^*(\theta, C) = J^*(\theta, C, 0) = \lim_{T \to \infty} J^*(\theta, C, T, 0)$.
- $\mathcal{C}^\theta = \{C_K^\theta\}_{K \in [K_{\text{L}}^\theta, K_{\text{U}}^\theta]}$: a class of controllers that are safe for dynamics θ that are parameterized by $K \in [K_{\text{L}}^\theta, K_{\text{U}}^\theta]$
- $K_{\text{opt}}(\theta, T)$: The K that maximizes $J^*(\theta, C_K^\theta, T, 0)$ for $K \in [K_{\text{L}}^\theta, K_{\text{U}}^\theta]$.
- $K_{\text{opt}}(\theta)$: The K that maximizes $J^*(\theta, C_K^\theta)$ for $K \in [K_{\text{L}}^\theta, K_{\text{U}}^\theta]$.
- C_K^{unc} : The unconstrained linear controller with parameter K, i.e. such that $C_K^{\text{unc}}(x) = -Kx$.
- $F_{\text{opt}}(\theta)$: The K that maximizes $J^*(\theta, C_K^{\text{unc}})$.

A.3. Algorithm Notation

- s_e : The number of rounds of the safe exploitation loop.
- $T_s = 2^s\sqrt{T}$: The length **and** starting time of round s of the safe exploitation phase. Note that $T_0 = \sqrt{T}$.
- ϵ_s : Uncertainty bound for θ^* in round s of the for loop.
- $\hat{\theta}_s$: An estimate of θ^* that is with high probability within ϵ_s distance of θ^*
- $C_s^{\text{alg}}(x_t)$: the controller that the algorithm uses in round s of the safe exploitation phase before safety adjustments
- $u_t^{\text{safeL}}, u_t^{\text{safeU}}$: bounds to enforce safety on the chosen control, which is $u_t = \max\left(\min\left(C_s^{\text{alg}}(x_t), u_t^{\text{safeU}}\right), u_t^{\text{safeL}}\right)$.
- C^{alg} : The controller of the algorithm.

A.4. Proof Notation

- $W_s = \{w_i\}_{i=T_s}^{T_{s+1}-1}$: Noise random variables in the round s of the safe exploitation phase.
- $(x'_0, x'_1, ...)$ and $(u'_0, u'_1, ...)$: Unless otherwise specified, these are the states and controls of the algorithm C^{alg}.
- $(\hat{x}_{T_0}, \hat{x}_{T_0+1}, ...)$: Unless otherwise defined in the theorem/lemma statement, $\hat{x}_{T_0}, \hat{x}_{T_0+1}, ...$ is the sequence of states if the control at each time $t \geq T_0$ is $C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s, T_s)}(x_t)$ for $s = \lfloor \log_2\left(\sqrt{T}\right) \rfloor$ and starting at $\hat{x}_{T_0} = x'_{T_0}$.
- $E_{\text{safe}} = \{\forall t < T : D_{\text{L}} \leq a^* x'_t + b^* u'_t \leq D_{\text{U}}\}$: The event that all of the controls satisfy the safety constraints.
- $E_1 = \left\{\forall t < T : |w_t| \leq \log^2(T)\right\}$: Event that all noise values have magnitude less than $\log^2(T)$
- $E_0 = \left\{\forall s \leq s_e : \|\theta^* - \hat{\theta}_s\|_\infty \leq \epsilon_s\right\}$: The event that all of the estimates of θ^* are within ϵ_s of θ^*.
- $E_2 = E_0 \bigcap \left\{\max_{s \in [0:s_e]} \epsilon_s \leq \tilde{O}_T(T^{-1/4})\right\}$.
- $E_2^s = \left\{\|\hat{\theta}_s - \theta^*\|_\infty \leq \epsilon_s \leq c_T \cdot T^{-1/4}\right\}$, where c_T is the coefficient in the $\tilde{O}_T(T^{-1/4})$ of the definition of event E_2.
- $E = E_{\text{safe}} \cap E_1 \cap E_2$
- $B_x = \log^3(T)$: Used throughout the appendix to simplify notation.
- $K^\theta_{D_{\text{U}}}$: the value of K that satisfies the equation $\frac{D_{\text{U}}}{a - bK^\theta_{D_{\text{U}}}} - D_{\text{U}} = \bar{w}$.

Appendix B. Additional Related Work

Another general line of work that is related but less directly comparable to our results is the area of model predictive control and system identification (Bemporad and Morari, 2007; Köhler et al., 2019; Lu et al., 2021; Oldewurtel et al., 2008; Mesbah, 2016; Bemporad et al., 2002; Muthirayan et al., 2022; Lorenzen et al., 2019; Simchowitz et al., 2018; Zhao and Li, 2022; Mania et al., 2019; Li et al., 2023). The results in these areas tend to focus more on feasibility and empirical performance rather than theoretical regret bounds, and therefore are less directly related to our work. Certainty equivalence algorithms consist of estimating the true dynamics and finding the optimal controller for these estimated dynamics. Our main algorithm uses a certainty equivalence approach to achieve the same rate of regret in the safety-constrained LQR setting. Less closely related to this paper, there is also a line of work studying optimal control with adversarial disturbances, where the goal is still to minimize regret but the system dynamics are known (see e.g. (Agarwal et al., 2019; Hazan and Singh, 2022)). Li et al. (2021a) also study optimal constrained control but again assume that the dynamics are known. The techniques and results of these lines of work with known dynamics are substantially different from our paper. This is because the key difficulty of our problem is that we do not know how to be safe apriori and must be safe while learning, which is not an issue with known dynamics.

Appendix C. Total Cost is Continuous in Starting Position (Lemma 3)

C.1. Main Proof

Proof

Let $\delta_{\mathrm{L3}} = \frac{1}{\log^{10}(T)}$ and $\epsilon_{\mathrm{L3}} = \frac{1}{\log^{46}(T)}$

Define $\epsilon = \|\theta - \theta^*\|_\infty$ and $\delta = |x - y|$. In order to bound the cost difference of the two trajectories, we will first bound the differences in states and controls of the two trajectories. We begin with the following lemma bounding the difference in future states when starting at two different initial states.

Lemma 4 *In the setting of Problem 1, for any $\theta \in \Theta$ such that $\epsilon := \|\theta - \theta^*\|_\infty \leq \frac{1}{\log^{46}(T)}$, $t \leq T$, $W' = \{w_i\}_{i=0}^{t-1}$, and any $K \in [\frac{a-1}{b}, \frac{a}{b}]$, there exists $\mathcal{Y}_{\mathrm{L4}} \in \mathbb{R}^t$ that only depends on K and θ such that the event $E_{\mathrm{L4}}(K, \theta, W') := \{W' \in \mathcal{Y}_{\mathrm{L4}}\}$ satisfies $\mathbb{P}(E_{\mathrm{L4}}(K, \theta, W')) = 1 - o_T(1/T^{10})$ and the following holds. Suppose that $|x|, |y| \leq 4\log^2(T)$ and $d := |x - y| \leq \frac{1}{\log^{10}(T)}$. Define d_i as the difference in state at time i when starting at $x_0 = x$ versus starting at $x_0 = y$ and using controller $C_K^\theta \in \mathcal{C}_{\mathrm{tr}}^\theta$ with noise variables W'. Then there exists an $L = \tilde{O}_T(1)$ such that for sufficiently large T, conditional on $E_{\mathrm{L4}}(K, \theta, W')$,*

$$d_i \leq \begin{cases} 2\xi^i \cdot d, & \text{for } L < i \leq t \\ 4d + \tilde{O}_T(\epsilon) & \text{for } 0 \leq i \leq L, \end{cases} \tag{7}$$

where $\xi := \left(1 - \frac{1}{\log^{10}(T)}\right)$.

The proof of Lemma 4 can be found in Appendix C.2.

We can also bound the difference in control in terms of the difference in state.

Lemma 5 *In the setting of Problem 1, for any $\theta \in \Theta$ such that $\|\theta - \theta^*\|_\infty \leq \frac{1}{\log^{46}(T)}$, any $K \in [\frac{a-1}{b}, \frac{a}{b}]$, and any x, y such that $d := |y - x| \leq \frac{1}{\log^{10}(T)}$,*

$$|C_K^\theta(x) - C_K^\theta(y)| = O_T(d). \tag{8}$$

The proof of Lemma 5 can be found in Appendix C.4.

We also will need the following event, which is a subset of the event E_1 applied only to times $i < t$.

Definition 6 *Define the event E_1^t as the event that for all $i \leq t - 1$, $|w_i| \leq \log^2(T)$.*

We can proceed by bounding the difference in total costs conditional on the event $E_{\mathrm{L4}}(K, \theta, W') \cap E_1^t$. Let $d_0, d_1, ..., d_t$ and $d_0^u, ..., d_{t-1}^u$ respectively be the absolute difference in states and controls when starting at $x_0 = x$ versus starting at $x_0 = x + \delta$ and using controller C_K^θ with noise W'. Let $x_0, ..., x_t$ and $u_0, ..., u_{t-1}$ be the states and controls when using controller C_K^θ starting at $x_0 = x$ with noise W'. Then we have the

following result conditional on $E_{\mathrm{L4}}(K, \theta, W') \cap E_1^t$ for sufficiently large T:

$$
\begin{aligned}
&\left| t \cdot J(\theta^*, C_K^\theta, t, x, W') - t \cdot J(\theta^*, C_K^\theta, t, x+\delta, W') \right| \\
&\leq 2qd_t|x_t| + qd_t^2 + \sum_{i=0}^{t-1} 2qd_i|x_i| + qd_i^2 + 2r|u_i|d_i^u + r\left(d_i^u\right)^2 \\
&\leq 2qd_t|x_t| + qd_t^2 + \sum_{i=0}^{t-1} 2qd_i|x_i| + qd_i^2 + 2r|u_i|O_T(d_i) + rO_T\left(d_i\right)^2 && \text{Lemma 5} \\
&= O_T\left(\sum_{i=0}^{t}(d_i + d_i^2)\left(|x| + \|D\|_\infty + \max_{w \in W'}|w|\right)\right) && \text{Lemma 56} \\
&= \tilde{O}_T\left(\sum_{i=0}^{t}(d_i + d_i^2)\right) && [\text{Event } E_1^t,\ \|D\|_\infty \leq \log^2(T),\ |x| \leq 4\log^2(T)] \\
&= \tilde{O}_T\left(\sum_{i=0}^{L}\left((4\delta + \tilde{O}_T(\epsilon)) + (4\delta + \tilde{O}_T(\epsilon))^2\right) + \sum_{i=L+1}^{t}\left(2\xi^i\delta + 4\xi^{2i}\delta^2\right)\right) && \text{Eq (7)} \\
&= \tilde{O}_T\left(\delta + \epsilon + \delta\sum_{i=0}^{t}\xi^i + \delta^2\sum_{i=0}^{t}\xi^{2i}\right) \\
&= \tilde{O}_T(\delta + \epsilon).
\end{aligned}
$$

The last line comes from the fact that $\xi = 1 - \frac{1}{\log^{10}(T)}$ and the formula for the sum of a geometric series. The above result holds conditional on event $E_{\mathrm{L3}}(K, \theta, W') := E_{\mathrm{L4}}(K, \theta, W') \cap E_1^t$, and by a union bound and Equation (159),

$$\mathbb{P}(E_{\mathrm{L3}}(K, \theta, W')) = \mathbb{P}(E_{\mathrm{L4}}(K, \theta, W') \cap E_1^t) = 1 - o_T(1/T^{10}).$$

■

C.2. Bounding Position Offset (Lemma 4)

In order to prove Lemma 4, we will use the following lemma that has a similar result but holds conditional on an event that depends on x.

Lemma 7 *There exists an $L = \tilde{O}_T(1)$ such that the following holds. Suppose that $|x|, |y| \leq 4\log^2(T)$ and $d := |x - y| \leq \frac{1}{\log^{10}(T)}$. In the setting of Problem 1, for any $\theta \in \Theta$ such that $\epsilon := \|\theta - \theta^*\|_\infty \leq \frac{1}{\log^{46}(T)}$, $t \leq T$, $W' = \{w_i\}_{i=0}^{t-1}$, and any $K \in [\frac{a-1}{b}, \frac{a}{b}]$, there exists $\mathcal{Y}_{\mathrm{L7}} \in \mathbb{R}^t$ that only depends on x, K and θ such that the event $E_{\mathrm{L7}}(x, K, \theta, W') := \{W' \in \mathcal{Y}_{\mathrm{L7}}\}$ satisfies $\mathbb{P}(E_{\mathrm{L7}}(x, K, \theta, W')) = 1 - o_T(1/T^{20})$ and the following holds. Define d_i as the difference in state at time $i \leq t$ when starting at $x_0 = x$ versus starting at $x_0 = y$ and using controller C_K^θ with noise variables W'. Then for sufficiently large T, conditional on $E_{\mathrm{L7}}(x, K, \theta, W')$,*

$$d_i \leq \begin{cases} \left(1 - \frac{1}{\log^{10}(T)}\right)^i \cdot d, & \text{if } i > L \\ 2d + \tilde{O}_T(\epsilon) & \text{if } i \leq L. \end{cases} \tag{9}$$

The proof of Lemma 7 can be found in Appendix C.3.

Now we need to find a single event $E_{\mathrm{L4}}(K,\theta,W')$ such that Equation (7) holds for all $|x|,|y| \le 4\log^2(T)$ under this event. Define the set

$$G := \left\{-4\log^2(T) + \frac{i}{\log^{10}(T)}\right\}_{i\in[0:8\log^{12}(T)]},$$

i.e. G is a grid of points evenly spaced $\frac{1}{\log^{10}(T)}$ apart. Note that $|G| = \tilde{O}_T(1)$. Now, take

$$E_{\mathrm{L4}}(K,\theta,W') = \bigcap_{g\in G} E_{\mathrm{L7}}(g,K,\theta,W').$$

First, we note that because $\mathbb{P}(E_{\mathrm{L7}}(g,K,\theta,W')) = 1 - o_T(1/T^{20})$ for all g and because $|G| = \tilde{O}_T(1)$, by a union bound $\mathbb{P}(E_{\mathrm{L4}}) = 1 - o_T(1/T^{10})$.

Now, consider any $|x|,|y| \le 4\log^2(T)$ such that $|x-y| \le \frac{1}{\log^{10}(T)}$. Then there must exist some $g \in G$ such that $\max(|x-g|,|y-g|) \le \frac{1}{\log^{10}(T)}$. For this g, let $d_0^x, d_1^x, \ldots$ be the sequence of differences of states when starting at state g versus x and using controller C_K^θ with noise W', and likewise let $d_0^y, d_1^y, \ldots$, be the sequence of absolute differences of states when starting at state g versus y and using controller C_K^θ with noise W'. Conditional on event $E_{\mathrm{L4}}(K,\theta,W')$, we have by Lemma 7 that $\{d_i^x\}$ and $\{d_i^y\}$ both satisfy Equation (9). Since $\{d_i^x\}$ and $\{d_i^y\}$ are both distances comparing to the same set of states starting at state g, we have by the triangle inequality that

$$d_i \le d_i^x + d_i^y.$$

Therefore, for $i \le t$ we have the following, where L is from Lemma 7:

$$d_i \le \begin{cases} 2\left(1 - \frac{1}{\log^{10}(T)}\right)^i \cdot d, & \text{if } i > L \\ 4d + \tilde{O}_T(\epsilon) & \text{if } i \le L. \end{cases} \tag{10}$$

This is exactly the desired result, and therefore we are done.

C.3. Bounding Conditional Position Offset (Lemma 7)

Proof

The main tool we will use for this proof is the following lemma that bounds the difference in future states in three different cases.

Lemma 8 *For any x, y, define $d = |y-x|$. In the setting of Problem 1 and for sufficiently large T, suppose $\theta \in \Theta$, $K \in \left[\frac{a-1}{b}, \frac{a}{b}\right]$, and $\|\theta - \theta^*\|_\infty = \epsilon \le \frac{1}{\log^{46}(T)}$. Then for some $\rho := |a^* - b^*K| + O_T(\epsilon)$,*

$$|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| \le \begin{cases} \min\left(2\rho d, \rho d + O_T(\epsilon)(|x| + \|D\|_\infty)\right) & \text{if } \mathcal{Z} \\ O_T(\epsilon)d & \text{if } \mathcal{W} \\ \rho d & \text{otherwise} \end{cases} \tag{11}$$

$$\mathcal{Z} := \left\{\min(x,y) \le \frac{D_L}{a-bK} \le \max(x,y) \le \frac{D_U}{a-bK} \text{ or } \frac{D_L}{a-bK} \le \min(x,y) \le \frac{D_U}{a-bK} \le \max(x,y)\right\}$$

$$\mathcal{W} := \left\{\max(x,y) \le \frac{D_L}{a-bK} \text{ or } \frac{D_U}{a-bK} \le \min(x,y)\right\}.$$

The proof of Lemma 8 can be found in Appendix C.4.

The rest of this proof will be structured as follows. First, we will introduce some additional definitions that we will use to construct event $E_{\mathrm{L}7}(x, K, \theta, W')$. Then, we will prove Lemma 7 in two cases.

Define $x_0, x_1, ..., x_T$ as the sequence of states when starting at state $x_0 = x$ and using controller C_K^θ with noise W'. For $i \le t$, define the event

$$X(i, x, K, \theta, W') := \left\{ \min\left(\left| x_i - \frac{D_L}{a - bK} \right|, \left| x_i - \frac{D_U}{a - bK} \right| \right) \le \frac{3}{\log^{10}(T)} \right\}.$$

Note that whether the event $X(i, x, K, \theta, W')$ occurs depends on $w_0, ..., w_{i-1}$. For $0 \le j \le t$ and $x \in \mathbb{R}$, define the event $H(j, x, K, \theta, W')$ as

$$H(j, x, K, \theta, W') := \left\{ \left| \{0 \le i \le j : X(i, x, K\theta, W')\} \right| \le \log^{23}(T) + \frac{24 B_P j}{\log^{10}(T)} \right\}.$$

Define

$$E^*(x, K, \theta, W') := \bigcap_{0 \le j \le t} H(j, x, K, \theta, W').$$

Now we will show that $\mathbb{P}(E^*(x, K, \theta, W')) = 1 - o_T\left(\frac{1}{T^{20}}\right)$. Fix any $j \le t$. If $j \le \log^{23}(T)$, then $H(j, x, K, \theta, W')$ holds with probability 1 by definition. Now suppose $j > \log^{23}(T)$. Because $\mathcal{D}$ has a density bounded by B_P and $x_i = a^* x_{i-1} + b^* u_{i-1} + w_{i-1}$, we must have that $\mathbb{P}(X(i, x, K, \theta, W')) \le \frac{12 B_P}{\log^{10}(T)}$ for all i. Define $M_k = \sum_{i=0}^{k-1} 1_{X(i,x,K,\theta,W')} - \frac{12 B_P}{\log^{10}(T)}$. For sufficiently large T, M_k is a supermartingale with differences bounded in magnitude by 1. Therefore, by the Azuma–Hoeffding inequality, with probability $1 - o_T(1/T^{21})$,

$$\left| \{0 \le i \le j : X(i, x, K, \theta, W')\} \right| \le \frac{12 B_P (j+1)}{\log^{10}(T)} + \log(T)\sqrt{j} \le \frac{24 B_P j}{\log^{10}(T)},$$

where the last inequality holds for sufficiently large T and $j > \log^{23}(T)$. Therefore, $\mathbb{P}(H(j, x, K, \theta, W')) \ge 1 - o_T(1/T^{21})$. Taking a union bound over all $\log^{23}(T) < j \le t$ gives that $\mathbb{P}(E^*(x, K, \theta, W')) = 1 - o_T(1/T^{20})$.

To prove Lemma 7, will split the range of potential K into two parts, $K \in \left[\frac{a^* - 1 + \frac{1}{\log^9(T)}}{b^*}, \frac{a}{b} \right]$ and $K \in \left[\frac{a-1}{b}, \frac{a^* - 1 + \frac{1}{\log^9(T)}}{b^*} \right]$. We will also use the following bounds.

Lemma 9 *For any $\theta \in \Theta$ such that $\|\theta - \theta^*\|_\infty \le \frac{1}{\log^{10}(T)}$,*

$$\frac{a - 1}{b} = \frac{a^* - 1 - O_T\left(\frac{1}{\log^{10}(T)}\right)}{b^*}$$

and

$$\frac{a}{b} = \frac{a^* + O_T\left(\frac{1}{\log^{10}(T)}\right)}{b^*}.$$

The proof of Lemma 9 can be found in Appendix C.5.

Now we are ready to proceed with the two cases for K.

Case 1: $K \in \left[\frac{a^* - 1 + \frac{1}{\log^9(T)}}{b^*}, \frac{a}{b} \right]$

For $i \le t$, define

$$\mathcal{Z}_i := \left\{ \min(x_i, y_i) \le \frac{D_L}{a - bK} \le \max(x_i, y_i) \le \frac{D_U}{a - bK} \text{ or } \frac{D_L}{a - bK} \le \min(x_i, y_i) \le \frac{D_U}{a - bK} \le \max(x_i, y_i) \right\}$$

and define

$$\kappa(j) = |\{0 \le i \le j : \mathcal{Z}_i\}|.$$

Because Lemma 9 implies that $\frac{a}{b} = \frac{a^*}{b^*} + O_T\left(\frac{1}{\log^{10}(T)}\right)$, we have for $K \in \left[\frac{a^* - 1 + \frac{1}{\log^9(T)}}{b^*}, \frac{a}{b}\right]$ that $|a^* - b^*K| \le 1 - \frac{1}{\log^9(T)}$. Because $\epsilon \le \frac{1}{\log^{46}(T)}$, this implies that $|a^* - b^*K| + O_T(\epsilon) \le 1 - \frac{1}{2\log^9(T)}$. This will allow us to bound the ρ in Lemma 8 by $1 - \frac{2}{\log^9(T)}$. Combining this with Lemma 8, we have the following piece-wise upper bound (note that we combined the $\mathcal{W}$ and the "otherwise" case using that $O_T(\epsilon) \le 1 - \frac{1}{2\log^9(T)}$ for suff large T),

$$d_{j+1} \le \begin{cases} \min\left(2\left(1 - \frac{1}{2\log^9(T)}\right) d_j, d_j + O_T(\epsilon)(|x_j| + \|D\|_\infty)\right) & \text{if } \mathcal{Z}_j \\ \left(1 - \frac{1}{2\log^9(T)}\right) d_j & \text{otherwise.} \end{cases} \tag{12}$$

Conditional on event E_1^t, for all $j \le t$, $|x_j| \le O_T(\log^2(T))$ by Lemma 56 because $\|D\|_\infty \le \log^2(T)$ and $|x| \le 4\log^2(T)$. Starting with the base case that $d_0 = d$, this with Equation (12) implies the following two relationships both hold for d_{j+1} conditional on event E_1^t for sufficiently large T. Equation (13) holds because $\left(1 - \frac{1}{2\log^9(T)}\right) \le 1$ for sufficiently large T and using the second term in the min of Equation (12). Equation (14) holds using the first term in the min of Equation (12).

$$d_{j+1} \le d + \kappa(j) \cdot O_T(\epsilon \log^2(T)) \tag{13}$$

and

$$d_{j+1} \le \left(\left(1 - \frac{1}{2\log^9(T)}\right)^{j+1} 2^{\kappa(j)}\right) \cdot d. \tag{14}$$

Equations (13) and (14) look almost like the desired result, and the remaining step is to show that $\kappa(j)$ is sufficiently "small".

Next, define the event A_j as

$$A_j := \left\{\forall i \le \min(j, \log^{33}(T)) : d_i \le 2d + \epsilon \log^{36}(T)\right\} \bigcap \left\{\forall \log^{33}(T) < i \le j : d_i \le \left(1 - \frac{1}{\log^{10}(T)}\right)^i d\right\}.$$

By this construction, A_t is exactly what we are trying to show in Lemma 7 with $L = \log^{33}(T)$. We will now prove that A_t holds for sufficiently large T conditional on $E^*(x, K, \theta, W') \cap E_1^t$.

For sufficiently large T and any $j \le t$, by construction of A_j and because $d \le \frac{1}{\log^{10}(T)}$ and $\epsilon \le \frac{1}{\log^{46}(T)}$, we have that

$$A_j \subseteq \left\{\forall 0 \le i \le j : d_i \le \frac{3}{\log^{10}(T)}\right\}. \tag{15}$$

Note that for event $\mathcal{Z}_i$ to hold, it must be the case that x_i is within d_i of either $\frac{D_U}{a-bK}$ or $\frac{D_L}{a-bK}$. Therefore, conditional on $E^*(x, K, \theta, W') \cap A_j$, we have for $j \geq \log^{33}(T)$,

$$\begin{aligned}
\kappa(j) &= |\{0 \leq i \leq j : \mathcal{Z}_i\}| \\
&\leq \left|\left\{0 \leq i \leq j : \min\left(\left|x_i - \frac{D_U}{a-bK}\right|, \left|x_i - \frac{D_L}{a-bK}\right|\right) \leq d_i\right\}\right| \\
&\leq \left|\left\{0 \leq i \leq j : \min\left(\left|x_i - \frac{D_U}{a-bK}\right|, \left|x_i - \frac{D_L}{a-bK}\right|\right) \leq \frac{3}{\log^{10}(T)}\right\}\right| && \text{Equation (15)} \\
&= |\{0 \leq i \leq j : X(i, x, K, \theta, W')\}| \\
&\leq \log^{23}(T) + \frac{24 B_P j}{\log^{10}(T)}. && E^*(x, K, \theta, W') \\
&= O_T\left(\frac{j+1}{\log^{10}(T)}\right) && (16)
\end{aligned}$$

We will now use Equations (13) and (14) to show that A_{j+1} holds conditional on $E_1^t \cap E^*(x, K, \theta, W') \cap A_j$. In order to show that A_{j+1} holds conditional on A_j, we must show that d_{j+1} satisfies the necessary inequality in the definition of A_{j+1}. Consider the following two cases for $j \geq 0$.

If $j + 1 \leq \log^{33}(T)$, for sufficiently large T conditional on $A_j \cap E_1^t \cap E^*(x, K, \theta, W')$,

$$\begin{aligned}
d_{j+1} &\leq d + \kappa(j) \cdot O_T(\epsilon \log^2(T)) && \text{Equation (13)} \\
&= d + O_T(j\epsilon \log^2(T)) && \kappa(j) \leq j + 1 \\
&\leq d + O_T(\epsilon \log^{35}(T)) \\
&\leq d + \log^{36}(T)\epsilon \\
&\leq 2d + \log^{36}(T)\epsilon. && (17)
\end{aligned}$$

This is the necessary inequality that needs to be shown in order for A_{j+1} to hold given that A_j holds if $j + 1 \leq \log^{33}(T)$.

If $j+1 > \log^{33}(T)$, for sufficiently large T conditional on $A_j \cap E_1^t \cap E^*(x,K,\theta,W')$,

$$
\begin{aligned}
&d_{j+1}\\
&\le \left(1-\frac{1}{2\log^9(T)}\right)^{j+1} 2^{\kappa(j)}\cdot d && \text{Equation (14)}\\
&\le \left(1-\frac{1}{2\log^9(T)}\right)^{j+1} 2^{O_T(\frac{j+1}{\log^{10}(T)})}\cdot d && \text{Equation (16)}\\
&= \left(1-\frac{1}{2\log^9(T)}\right)^{j+1} e^{O_T(\frac{j+1}{\log^{10}(T)})}\cdot d\\
&\le \left(1-\frac{1}{2\log^9(T)}\right)^{j+1} \left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)+O_T\left(\frac{1}{\log^{20}(T)}\right)\right)^{j+1}\cdot d && [e^x \le 1+x+x^2 \text{ for } x\le 1]\\
&= \left(1-\frac{1}{2\log^9(T)}\right)^{j+1} \left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1}\cdot d\\
&= \left(1-\frac{1}{2\log^9(T)}+O_T\left(\frac{1}{\log^{10}(T)}\right)-\frac{1}{2\log^9(T)}\cdot O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1}\cdot d\\
&\le \left(1-\frac{1}{2\log^9(T)}+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1}\cdot d\\
&\le \left(1-\frac{1}{\log^{10}(T)}\right)^{j+1}\cdot d. && \text{for sufficiently large } T
\end{aligned}
\tag{18}
$$

This is the necessary inequality that needs to be shown in order for A_{j+1} to hold given that A_j holds if $j+1 \ge \log^{33}(T)$.

Equations (17) and (18) together imply that for sufficiently large T, A_{j+1} holds conditional on $A_j \cap E_1^t \cap E^*(x,K,\theta,W')$. Note that A_0 always holds by definition because $d_0 = d$. Therefore, we can conclude by induction that A_t must hold conditional on $E_1^t \cap E^*(x,K,\theta,W')$ for sufficiently large T. Finally, by definition of A_t, this implies that conditional on $E^*(x,K,\theta,W') \cap E_1^t$ for sufficiently large T, for all $0 \le j \le t$,

$$
d_j \le \begin{cases} \left(1-\frac{1}{\log^{10}(T)}\right)^j \cdot d, & \text{if } j > \log^{33}(T)\\ d+\tilde{O}_T(\epsilon), & \text{if } j \le \log^{33}(T). \end{cases}
\tag{19}
$$

Taking $E_{\mathrm{L}7}(x,K,\theta,W') = E^*(x,K,\theta,W') \cap E_1^t$, by a union bound we have that $\mathbb{P}(E_{\mathrm{L}7}(x,K,\theta,W') \ge 1 - o_T(1/T^{20})$. This completes the proof of Lemma 7 for Case 1.

Case 2: $K \in \left[\frac{a-1}{b}, \frac{a^*-1+\frac{1}{\log^9(T)}}{b^*}\right]$

Define

$$
\mathcal{W}_j := \left\{\min(x_j,y_j) \ge \frac{D_U}{a-bK} \text{ or } \max(x_j,y_j) \le \frac{D_L}{a-bK}\right\}
$$

and

$$
\lambda(j) := |\{0 \le i \le j : \mathcal{W}_i\}|.
$$

For any $K \in \left[\frac{a-1}{b}, \frac{a^*-1+\frac{1}{\log^9(T)}}{b^*}\right]$, we have that $|a^* - b^*K| - 1 \leq O_T\left(\frac{1}{\log^{10}(T)}\right)$ by Lemma 9. This with the fact that $\epsilon \leq \frac{1}{\log^{46}(T)}$ implies that $(a^* - b^*K) + O_T(\epsilon) \leq |a^* - b^*K| + O_T(\epsilon) \leq 1 + O_T\left(\frac{1}{\log^{10}(T)}\right)$. This allows us to bound the ρ in Lemma 8 to be $1 + O_T\left(\frac{1}{\log^{10}(T)}\right)$. By Lemma 8 and plugging this bound in for ρ, this gives the following bound.

$$d_{j+1} \leq \begin{cases} O_T(\epsilon)\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right) d_j & \text{If } \mathcal{W}_j \\ \min\left(2\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right) d_j, \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)(d_j + O_T(\epsilon)(|x_j| + \|D\|_\infty))\right) & \text{If } \mathcal{Z}_j \\ \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right) d_j & \text{Otherwise} \end{cases} \tag{20}$$

Similar to in the proof of Case 1 above, by Lemma 56 and the assumption that $\|D\|_\infty \leq \log^2(T)$, we have that conditional on event E_1^t, Equation (20) implies the following two relationships. The first relationship comes from using the first term in the min of Equation (20) and recursing.

$$d_{j+1} \leq \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1} \cdot 2^{\kappa(j)} \cdot O_T(\epsilon)^{\lambda(j)} \cdot d. \tag{21}$$

The second relationship comes from using the second term in the min of Equation (20) and bounding $\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)(|x_j| + \|D\|_\infty) = O_T(\log^2(T))$ under event E_1^t. This gives the recursive relationship of

$$d_{j+1} \leq \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)(O_T(\epsilon))^{1_{\mathcal{W}_j}} \cdot d_j + O_T(\epsilon \log^2(T)) 1_{\mathcal{Z}_j}. \tag{22}$$

In other words, at every step there is a multiplicative factor of $\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)$. When $\mathcal{W}_j$ holds, there is an additional multiplicative factor of $O_T(\epsilon)$. When $\mathcal{Z}_j$ holds, there is an additive factor of $O_T(\epsilon \log^2(T))$. Unwrapping Equation (22) gives that, at time $j+1$, any additive factor contributed at time $i \leq j$ will be scaled by $O_T(\epsilon)^{\lambda(j)-\lambda(i)}\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i}$. This gives that

$$d_{j+1} \leq \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1} O_T(\epsilon)^{\lambda(j)} \cdot d + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{j} 1_{\mathcal{Z}_i} O_T(\epsilon)^{\lambda(j)-\lambda(i)} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i}. \tag{23}$$

Again this almost looks like the desired result, except we need to show that the additional terms involving $\kappa(j)$ and $\lambda(j)$ are not "too large". We will use the following lemma that lower bounds $\lambda(j)$ using the same event A_j as defined above in the first case. Similar to Case 1, we will then use this to show that for sufficiently large T, A_t holds conditional on $E_1^t \cap E_{\text{L10}}(x, K, \theta, W') \cap E^*(x, K, \theta, W')$.

Lemma 10 *Suppose $|1 - (a^* - b^*K)| = O_T\left(\frac{1}{\log^9(T)}\right)$. Then in the setting of Problem 1 and using the notation and assumptions of Lemma 7, there exists a $\mathcal{Y}_{\text{L10}} \in \mathbb{R}^t$ that only depends on x, K, θ such that the event $E_{\text{L10}}(x, K, \theta, W') := \{W' \in \mathcal{Y}_{\text{L10}}\}$ satisfies $\mathbb{P}(E_{\text{L10}}(x, K, \theta, W')) = 1 - o_T(1/T^{20})$ and that for all $t_1 < t_2 \leq t$ satisfying $t_2 - t_1 \geq \log^8(T)$, the following is true conditional on event $A_{t_2} \cap E_{\text{L10}}(x, K, \theta, W')$ for sufficiently large T:*

$$\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{t_2+1-t_1} O_T(\epsilon)^{\lambda(t_2)-\lambda(t_1)} \leq \left(1 - \frac{1}{2\log^9(T)}\right)^{t_2+1-t_1}.$$

The proof of Lemma 10 can be found in Appendix C.6.

We will now show that the event A_{j+1} holds conditional on $E_{\mathrm{L10}}(x,K,\theta,W')\cap E^*(x,K,\theta,W')\cap E_1^t\cap A_j$.

For $j<\log^8(T)$, conditional on $E_1^t\cap E_{\mathrm{L10}}(x,K,\theta,W')\cap E^*(x,K,\theta,W')\cap A_j$ and for sufficiently large T,

$$\begin{aligned}
&d_{j+1}\\
&\le \left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1} O_T(\epsilon)^{\lambda(j)}\cdot d\\
&\quad + O_T(\epsilon\log^2(T))\cdot\sum_{i=0}^{j} 1_{\mathcal{Z}_j} O_T(\epsilon)^{\lambda(j)-\lambda(i)}\left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} && \text{Eq. (23)}\\
&\le \left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{\log^8(T)+1}\cdot d + O_T(\epsilon\log^2(T))\cdot\sum_{i=0}^{j}\left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j} && \epsilon\le O_T(1)\\
&\le \left(1+O_T\left(\frac{1}{\log^{2}(T)}\right)\right)\cdot d + O_T(\epsilon\log^2(T))\cdot(j+1)\cdot\left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j} && \text{Lemma 11}\\
&\le \left(1+O_T\left(\frac{1}{\log^{2}(T)}\right)\right)\cdot d + O_T(\epsilon\log^2(T))\cdot(\log^8(T)+1)\cdot\left(1+O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{\log^8(T)}\\
&\le \left(1+O_T\left(\frac{1}{\log^{2}(T)}\right)\right)\left(d+O_T\left(\epsilon\log^{10}(T)\right)\right) && \text{Lemma 11}\\
&\le 2d+O_T(\epsilon\log^{10}(T)) && \text{Suff. large } T\\
&\le 2d+\epsilon\log^{36}(T). && \text{Suff. large } T
\end{aligned} \tag{24}$$

Above, we used the following result:

Lemma 11 *Suppose $g(T)$ is a non-negative function of T such that $g(T)>1$ for sufficiently large T. Furthermore, suppose $f(T)$ is a non-negative function of T that satisfies $f(T)g(T)\le 1/2$ for sufficiently large T. Then we have that*

$$1+f(T)g(T)\le(1+f(T))^{g(T)}\le 1+2f(T)g(T).$$

This implies that

$$(1+f(T))^{g(T)}=1+\Theta_T(f(t)\cdot g(T)).$$

Proof First, we note that for any $x\ge 0$ and $r>1$, $(1+x)^r\ge 1+rx$. This implies that for sufficiently large T, we have that

$$(1+f(T))^{g(T)}\ge 1+f(T)g(T).$$

This proves one direction of the desired equation. For the other direction, note that for $r>0$ and $x\in[0,1/r)$, we have $(1+x)^r\le\frac{1}{1-rx}$. This implies that

$$\begin{aligned}
(1+f(T))^{g(T)} &\le \frac{1}{1-f(T)g(T)}\\
&= 1+\frac{f(T)g(T)}{1-f(T)g(T)}\\
&\le 1+2f(T)g(T).
\end{aligned}$$

This proves the other direction of the desired equation. Therefore we have that $(1+f(T))^{g(T)}=1+\Theta(f(T)g(T))$. ∎

For $\log^8(T) \le j < \log^{33}(T)$, conditional on event $E_1^t \cap E_{\text{L10}}(x, K, \theta, W') \cap E^*(x, K, \theta, W') \cap A_j$ and for sufficiently large T,

$$
\begin{aligned}
& d_{j+1} \\
& \le \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1} O_T(\epsilon)^{\lambda(j)} \cdot d \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{j} 1_{\mathcal{Z}_j} O_T(\epsilon)^{\lambda(j)-\lambda(i)} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} && \text{Eq (23)} \\
& \le \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1-0} O_T(\epsilon)^{\lambda(j)-\lambda(0)} \cdot d \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{j} 1_{\mathcal{Z}_j} O_T(\epsilon)^{\lambda(j)-\lambda(i)} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} \\
& \le \left(1 - \frac{1}{2\log^9(T)}\right)^{j+1} d \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{j} 1_{\mathcal{Z}_j} O_T(\epsilon)^{\lambda(j)-\lambda(i)} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} && \text{Lemma 10} \\
& \le d + O_T(\epsilon \log^2(T)) \sum_{i=0}^{\lceil j - \log^8(T) \rceil - 1} 1_{\mathcal{Z}_j} O_T(\epsilon)^{\lambda(j)-\lambda(i)} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1-i} \\
& \quad + O_T(\epsilon \log^2(T)) \sum_{i=\lceil j - \log^8(T) \rceil}^{j} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} \\
& \le d + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{\lceil j - \log^8(T) \rceil - 1} \left(1_{\mathcal{Z}_j} \cdot \left(1 - \frac{1}{2\log^9(T)}\right)^{j+1-i}\right) \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=\lceil j - \log^8(T) \rceil}^{j} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} && \text{Lemma 10} \\
& \le d + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{\lceil j - \log^8(T) \rceil - 1} \left(1_{\mathcal{Z}_j} \cdot O_T(1)\right) \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=\lceil j - \log^8(T) \rceil}^{j} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j-i} \\
& \le d + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{\lceil j - \log^8(T) \rceil - 1} \left(1_{\mathcal{Z}_j} \cdot O_T(1)\right) \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=\lceil j - \log^8(T) \rceil}^{j} \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{\log^8(T)} \\
& \le d + O_T(\epsilon \log^2(T)) \cdot \sum_{i=0}^{\lceil j - \log^8(T) \rceil - 1} \left(1_{\mathcal{Z}_j} \cdot O_T(1)\right) \\
& \quad + O_T(\epsilon \log^2(T)) \cdot \sum_{i=\lceil j - \log^8(T) \rceil}^{j} \left(1 + O_T\left(\frac{1}{\log^2(T)}\right)\right) && \text{Lemma 11} \\
& \le d + O_T(\epsilon \log^{35}(T)) + O_T\left(\epsilon \log^{10}(T)\right) \\
& \le d + O_T(\epsilon \log^{35}(T)) \\
& \le d + \epsilon \log^{36}(T). && \text{Suff large } T
\end{aligned}
$$

Finally, for $j \geq \log^{33}(T)$, conditional on event $E_1^t \cap E_{\mathrm{L10}}(x,K,\theta,W') \cap E^*(x,K,\theta,W') \cap A_j$ and for sufficiently large T,

$$\begin{aligned}
d_{j+1} &\leq \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1} \cdot 2^{\kappa(j)} \cdot O_T(\epsilon)^{\lambda(j)} \cdot d && \text{Equation (21)} \\
&\leq \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{j+1-0} \cdot O_T(\epsilon)^{\lambda(j)-\lambda(0)} \cdot 2^{\kappa(j)} \cdot d \\
&\leq \left(1 - \frac{1}{2\log^9(T)}\right)^{j+1} 2^{\kappa(j)} \cdot d && \text{Lemma 10} \\
&\leq \left(1 - \frac{1}{\log^{10}(T)}\right)^{j+1} \cdot d. && \text{As in Equation (18)}
\end{aligned}$$

Combining all three cases, we have that for all $j \geq 0$, conditional on $E_1^t \cap E_{\mathrm{L10}}(x,K,\theta,W') \cap E^*(x,K,\theta,W') \cap A_j$, A_{j+1} holds. As in Case 1, we can conclude by induction using A_0 as the base case to get that conditional on $E_{\mathrm{L10}}(x,K,\theta,W') \cap E^*(x,K,\theta,W') \cap E_1^t$, the event A_t holds, which implies that

$$d_j \leq \begin{cases} \left(1 - \frac{1}{\log^{10}(T)}\right)^j \cdot d, & \text{if } j > \log^{33}(T) \\ 2d + \tilde{O}_T(\epsilon), & \text{if } j \leq \log^{33}(T). \end{cases} \tag{25}$$

Taking $E_{\mathrm{L7}}(x,K,\theta,W') = E_{\mathrm{L10}}(x,K,\theta,W') \cap E^*(x,K,\theta,W') \cap E_1^t$, we have by a union bound that $\mathbb{P}(E_{\mathrm{L7}}(x,K,\theta,W')) = 1 - o_T(1/T^{20})$. This completes the proof of Lemma 7 for Case 2.

■

C.4. Bounding Continuity of Controls (Lemmas 5 and 8)

Proof We have four cases depending on the values of x, y. We will prove the results of Lemma 5 and Lemma 8 for each of these cases separately. WLOG assume that $x \leq y$.

Case 1: $\frac{D_L}{a-bK} \leq x \leq y \leq \frac{D_U}{a-bK}$.

In this case, $C_K^\theta(x) = -Kx$ and $C_K^\theta(y) = -Ky$, and therefore the following two equations hold. Case 1 Lemma 8:

$$|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| = |a^* - b^*K|d = (|a^* - b^*K| + O_T(\epsilon))d.$$

Case 1 Lemma 5:

$$|C_K^\theta(x) - C_K^\theta(y)| = Kd \leq \frac{a}{b} \cdot d \leq \frac{\bar{a}}{\underline{b}} \cdot d = O_T(d).$$

Case 2: $\frac{D_U}{a-bK} \leq x \leq y$ or $x \leq y \leq \frac{D_L}{a-bK}$ (which is $\mathcal{W}$ of Lemma 8).

First, assume the former is true. Then $C_K^\theta(x) = \frac{D_U - ax}{b}$ and likewise $C_K^\theta(y) = \frac{D_U - ay}{b}$. Therefore the following equations hold.

Case 2 Lemma 8:

$$\begin{aligned}
|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| &= d\left|a^* - \frac{a}{b}b^*\right| \\
&= d\left|\frac{a^*b - ab^*}{b}\right| \\
&\leq d\frac{\max\left((a+\epsilon)b - a(b-\epsilon), |(a-\epsilon)b - a(b+\epsilon)|\right)}{b} \\
&= d\frac{\epsilon b + \epsilon a}{b} \\
&\leq d\frac{\epsilon\bar{b} + \epsilon\bar{a}}{\underline{b}} \\
&\leq O_T(\epsilon)d.
\end{aligned}$$

Case 2 Lemma 5:

$$|C_K^\theta(y) - C_K^\theta(x)| = \frac{a}{b}\cdot d \leq \frac{\bar{a}}{\underline{b}}\cdot d = O_T(d).$$

The same logic holds for when $x \leq y \leq \frac{D_L}{a-bK}$.

Case 3: $x \leq \frac{D_L}{a-bK}$ and $y \geq \frac{D_U}{a-bK}$

In this case, $C_K^\theta(x) = \frac{D_L - ax}{b}$ and $C_K^\theta(y) = \frac{D_U - ay}{b}$. We will use the fact that $|a-bK|d = |a-bK||y-x| \geq |D_U - D_L|$ in this case.

Case 3 Lemma 8:

$$\begin{aligned}
&|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| \\
&= \left|\frac{b^*}{b}(D_L - D_U) + \left(a^* - \frac{a}{b}b^*\right)(x-y)\right| \\
&\leq \frac{b^*}{b}|D_U - D_L| + \left|a^* - \frac{a}{b}b^*\right| d \\
&\leq \frac{b^*}{b}|a - bK|d + \left|a^* - \frac{a}{b}b^*\right| d \\
&\leq \frac{b^*}{b}|a^* - b^*K|d + \frac{b^*}{b}|a - a^* + (b^* - b)K|\, d + \left|a^* - \frac{a}{b}b^*\right| d \\
&\leq |a^* - b^*K|d + \left|\frac{b^*}{b} - 1\right||a^* - b^*K|d + \frac{b^*}{b}|a - a^* + (b^* - b)K|d + \left|a^* - \frac{a}{b}b^*\right| d \\
&\leq \left(|a^* - b^*K| + O_T(\epsilon)\right) d. \qquad (26)
\end{aligned}$$

In the last line we used that $|a^* - b^*K| \leq a^* + b^*|K| \leq \bar{a} + \bar{b}\frac{\bar{a}+1}{\underline{b}} = O_T(1)$, that $|\frac{b^*}{b} - 1| \leq \frac{\epsilon}{\underline{b}} = O_T(\epsilon)$, that $|a - a^* + (b^* - b)K| \leq \epsilon(1 + |K|) \leq \epsilon(1 + \frac{\bar{a}+1}{\underline{b}}) = O_T(\epsilon)$, and that $|a^* - \frac{a}{b}b^*| \leq \epsilon + \frac{a}{b}\epsilon \leq \epsilon + \frac{\bar{a}}{\underline{b}}\epsilon = O_T(\epsilon)$.

Case 3 Lemma 5:

$$\begin{aligned}|C_K^\theta(y) - C_K^\theta(x)| &= \left|\frac{1}{b}(D_U - D_L) + \frac{a}{b}(x-y)\right| \\ &= \frac{1}{b}|D_U - D_L| + \frac{a}{b}|x-y| \\ &\le \frac{1}{b}|a-bK|d + \frac{a}{b}d \\ &\le \frac{1}{\underline{b}}d + \frac{\bar{a}}{\underline{b}}d \\ &= O_T(d).\end{aligned}$$

Case 4: If $\frac{D_L}{a-bK} \le x \le \frac{D_U}{a-bK}$ and $y \ge \frac{D_U}{a-bK}$. Note that by symmetry, this is equivalent to $\frac{D_L}{a-bK} \le y \le \frac{D_U}{a-bK}$ and $x \le \frac{D_L}{a-bK}$. We will first assume the former. For Lemma 8, this case is equivalent to $\mathcal{Z}$.

Case 4 Lemma 8:

In this case, $C_K^\theta(x) = -Kx$ and $C_K^\theta(y) = \frac{D_U - ay}{b}$. Furthermore, in this case $|y-x| \ge \left|y - \frac{D_U}{a-bK}\right|$. Therefore, in this case we have

$$\begin{aligned}&|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| \\ &= |a^*x + b^*C_K^\theta(x) - a^*y - b^*Ky + b^*Ky - b^*C_K^\theta(y)| \\ &\le |(a^* - b^*K)x - (a^* - b^*K)y| + b^*\left|-Ky - \frac{D_U - ay}{b}\right| \\ &\le |(a^* - b^*K)x - (a^* - b^*K)y| + b^*\left|\frac{(a-bK)y - D_U}{b}\right| \\ &\le |(a^* - b^*K)x - (a^* - b^*K)y| + \frac{b^*|a-bK|}{b}\left|y - \frac{D_U}{a-bK}\right| \\ &\le |(a^* - b^*K)x - (a^* - b^*K)y| + \frac{b^*|a-bK|}{b}|y-x| \\ &= |a^* - b^*K|d + |a-bK|\frac{b^*}{b}d \\ &\le |a^* - b^*K|d + |a-bK|d + \left|1 - \frac{b^*}{b}\right||a-bK|d \\ &\le 2|a^* - b^*K|d + |a - bK - (a^* - b^*K)|d + \left|1 - \frac{b^*}{b}\right||a-bK|d \\ &\le 2(|a^* - b^*K| + O_T(\epsilon))d. \qquad \text{As in Equation (26)}\end{aligned}$$

Alternatively, note that in this case,

$$(a^* - b^*K)|x| \le (a - bK)|x| + O_T(\epsilon)|x| \tag{27}$$

and

$$(a-bK)x \le D_U \le (a-bK)y. \tag{28}$$

Therefore,

$$\begin{aligned}
|(a^* - b^*K)x - D_U| &\le |(a - bK)x - D_U| + O_T(\epsilon)|x| && \text{Equation (27)} \\
&\le |(a - bK)x - (a - bK)y| + O_T(\epsilon)|x| && \text{Equation (28)} \\
&\le |a - bK|d + O_T(\epsilon)|x| \\
&\le |a^* - b^*K|d + O_T(\epsilon)(d + |x|). && (29)
\end{aligned}$$

Therefore we can find an alternative bound on $|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)|$, using Equation (29) and that $|y| \le |x| + d$.

$$\begin{aligned}
&|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| \\
&= \left|(a^* - b^*K)x - \frac{b^*}{b}D_U - \left(a^* - \frac{ab^*}{b}\right)y\right| \\
&\le |(a^* - b^*K)x - D_U| + \left|1 - \frac{b^*}{b}\right| D_U + \left|a^* - \frac{ab^*}{b}\right| |y| \\
&\le |a^* - b^*K|d + O_T(\epsilon)(d + |x|) + \left|1 - \frac{b^*}{b}\right| D_U + \left|a^* - \frac{ab^*}{b}\right| |y| && \text{Equation (29)} \\
&\le |a^* - b^*K|d + O_T(\epsilon)(d + |x|) + \left|1 - \frac{b^*}{b}\right| D_U + \left|a^* - \frac{ab^*}{b}\right| (|x| + d) \\
&\le (|a^* - b^*K| + O_T(\epsilon))d + O_T(\epsilon)(|x| + D_U) \\
&\le (|a^* - b^*K| + O_T(\epsilon))d + O_T(\epsilon)(|x| + \|D\|_\infty).
\end{aligned}$$

where in the last line we once again bounded $|1 - \frac{b^*}{b}| = O_T(\epsilon)$ and $|a^* - \frac{ab^*}{b}| = O_T(\epsilon)$. Therefore, we have shown in this case that

$$\begin{aligned}
&|a^*x + b^*C_K^\theta(x) - a^*y - b^*C_K^\theta(y)| \\
&\le \min\left(2(|a^* - b^*K| + O_T(\epsilon))d, (|a^* - b^*K| + O_T(\epsilon))d + O_T(\epsilon)(|x| + \|D\|_\infty)\right)
\end{aligned}$$

Case 4 Lemma 5:

$$\begin{aligned}
|C_K^\theta(x) - C_K^\theta(y)| &= \left|-Kx - \frac{D_U - ay}{b}\right| \\
&\le |K||x - y| + \left|-Ky - \frac{D_U - ay}{b}\right| \\
&\le |K||x - y| + \left|\frac{(a - bK)y - D_U}{b}\right| \\
&\le |K||x - y| + \frac{|a - bK|}{b}\left|y - \frac{D_U}{a - bK}\right| \\
&\le |K||x - y| + \frac{|a - bK|}{b}|y - x| && \text{Equation (28)} \\
&= |K|d + \frac{|a - bK|}{b}d \\
&\le \frac{\bar{a} + 1}{\underline{b}}d + \frac{1}{\underline{b}}d \\
&= O_T(d).
\end{aligned}$$

Because these four cases cover all possible situations, we have shown the desired two lemmas. ■

C.5. Approximating K Bounds (Lemma 9)

Proof For sufficiently large T we have the following two results, using that $\|\theta - \theta^*\|_\infty \le \frac{1}{\log^{10}(T)}$:

$$\begin{aligned}
\frac{a-1}{b} &\ge \frac{a^* - \frac{1}{\log^{10}(T)} - 1}{b^* + \frac{1}{\log^{10}(T)}} \\
&= \frac{a^*-1}{b^*} \cdot \frac{b^*}{b^* + \frac{1}{\log^{10}(T)}} - \frac{1}{\log^{10}(T)(b^* + \frac{1}{\log^{10}(T)})} \\
&= \frac{a^*-1}{b^*} \cdot \left(1 - \frac{1}{\log^{10}(T)\left(b^* + \frac{1}{\log^{10}(T)}\right)}\right) - \frac{1}{\log^{10}(T)(b^* + \frac{1}{\log^{10}(T)})} \\
&= \frac{a^*-1}{b^*} - \frac{a^*-1}{b^*\log^{10}(T)(b^* + \frac{1}{\log^{10}(T)})} - \frac{1}{\log^{10}(T)(b^* + \frac{1}{\log^{10}(T)})} \\
&= \frac{a^* - 1 - O_T\left(\frac{1}{\log^{10}(T)}\right)}{b^*}.
\end{aligned}$$

$$\begin{aligned}
\frac{a}{b} &\le \frac{a^* + \frac{1}{\log^{10}(T)}}{b^* - \frac{1}{\log^{10}(T)}} \\
&= \frac{a^*}{b^*} \cdot \frac{b^*}{b^* - \frac{1}{\log^{10}(T)}} + \frac{1}{\log^{10}(T)(b^* - \frac{1}{\log^{10}(T)})} \\
&= \frac{a^*}{b^*} \cdot \left(1 + \frac{1}{\log^{10}(T)\left(b^* - \frac{1}{\log^{10}(T)}\right)}\right) + \frac{1}{\log^{10}(T)(b^* - \frac{1}{\log^{10}(T)})} \\
&= \frac{a^*}{b^*} + \frac{a^*}{b^*(b^* - \frac{1}{\log^{10}(T)})\log^{10}(T)} + \frac{1}{\log^{10}(T)(b^* - \frac{1}{\log^{10}(T)})} \\
&= \frac{a^* + O_T(1/\log^{10}(T))}{b^*}.
\end{aligned}$$

■

C.6. Bounding Impact of Extreme Case (Lemma 10)

Proof The first step to this proof is to construct event $E_{\mathrm{L10}}(x, K, \theta, W')$. For any $t_2 > t_1$ and $t_2 - t_1 \ge \log^8(T)$, define the event $E_{\mathrm{L10}}^{t_1,t_2}$ as

$$E_{\mathrm{L10}}^{t_1,t_2} = \left\{ \exists j \in [t_1 : t_2 - \lceil \log^5(T) \rceil - 1] : \left| \sum_{i=j}^{j+\lceil \log^5(T) \rceil} w_i \right| \ge 7\log^2(T) \right\}.$$

Define

$$E_{\mathrm{L10}}(x, K, \theta, W') := E_1^t \cap \bigcap_{t_1 < t_2 \le t, t_2 - t_1 \ge \log^8(T)} E_{\mathrm{L10}}^{t_1,t_2}.$$

First we will show that $\mathbb{P}(E_{\mathrm{L10}}(x, K, \theta, W')) = 1 - o_T(1/T^{20})$. Consider any pair $t_2 > t_1$ such that $t_2 - t_1 \ge \log^8(T)$. Divide the interval $[t_1 : t_2 - 1]$ into $\lfloor \frac{t_2 - t_1}{\lceil \log^5(T) \rceil + 1} \rfloor$ consecutive disjoint intervals of

length $\lceil \log^5(T) \rceil + 1$. Consider one such interval $[s_1, s_2]$. Then the distribution of $\frac{1}{\sqrt{\lceil \log^5(T) \rceil + 1}} \sum_{i=s_1}^{s_2} w_i$ converges in distribution to $N(0, \sigma_{\mathcal{D}}^2)$ as T grows, where we recall $\sigma_{\mathcal{D}}^2$ is the variance of distribution $\mathcal{D}$. The rate of this convergence depends on $\mathcal{D}$. Therefore, for sufficiently large T, we have that

$$\left| \mathbb{P}\left(\left| \frac{1}{\sqrt{\lceil \log^5(T) \rceil + 1}} \sum_{i=s_1}^{s_2} w_i \right| \geq \sigma_{\mathcal{D}}/2 \right) - \mathbb{P}\left(|N(0, \sigma_{\mathcal{D}}^2)| \geq \sigma_{\mathcal{D}}/2 \right) \right| \leq 0.1. \tag{30}$$

This implies that

$$\mathbb{P}\left(\left| \frac{1}{\sqrt{\lceil \log^5(T) \rceil + 1}} \sum_{i=s_1}^{s_2} w_i \right| \geq \sigma_{\mathcal{D}}/2 \right) \geq \mathbb{P}\left(|N(0, \sigma_{\mathcal{D}}^2)| \geq \sigma_{\mathcal{D}}/2 \right) - 0.1 \geq 0.5. \tag{31}$$

For sufficiently large T, we have that $\frac{\sqrt{\lceil \log^5(T) \rceil + 1}\sigma_{\mathcal{D}}}{2} \geq 7\log^2(T)$, and therefore this implies that for sufficiently large T,

$$\mathbb{P}\left(\left| \sum_{i=s_1}^{s_2} w_i \right| \geq 7\log^2(T) \right) \geq 0.5. \tag{32}$$

Because the random variables in each disjoint interval are independent, we have that each interval independently satisfies Equation (32) with probability at least $1/2$. Therefore, for sufficiently large T, the probability that Equation (32) fails to hold for all $\lfloor \frac{|t_2 - t_1|}{\lceil \log^5(T) \rceil + 1} \rfloor \geq \log^2(T)$ intervals is at most $(1/2)^{\lfloor \frac{|t_2 - t_1|}{\lceil \log^5(T) \rceil + 1} \rfloor} \leq 0.5^{\log^2(T)} = o_T(1/T^{22})$. Therefore, we have shown that

$$\mathbb{P}(E_{\text{L10}}^{t_1, t_2}) \geq 1 - o_T(1/T^{22}).$$

Since there are less than T^2 pairs (t_1, t_2) and $\mathbb{P}(E_1^t) \geq \mathbb{P}(E_1) = 1 - o_T(1/T^{20})$ by Equation (159), we have by a union bound that

$$\mathbb{P}(E_{\text{L10}}(x, K, \theta, W')) \geq 1 - o_T(T^2/T^{22}) - o_T(1/T^{20}) = 1 - o_T(1/T^{20}).$$

Lemma 12 *Using the assumptions and notation of the proof of Lemma 10, for all pairs t_1, t_2 such that $t_2 - t_1 \geq \log^8(T)$, conditional on event $A_{t_2} \cap E_{\text{L10}}(x, K, \theta, W')$,*

$$\lambda(t_2) - \lambda(t_1) = \Omega_T\left(\frac{|t_2 - t_1|}{\log^8(T)} \right). \tag{33}$$

The proof of Lemma 12 can be found in Section C.7.

By Lemma 12, conditional on $A_{t_2} \cap E_{\mathrm{L}10}(x, K, \theta, W')$, we that:

$$\begin{aligned}
&\left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{t_2+1-t_1} O_T(\epsilon)^{\lambda(t_2)-\lambda(t_1)} \\
&= \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{t_2+1-t_1} O_T\left(1/\log(T)\right)^{\lambda(t_2)-\lambda(t_1)} && \epsilon = O_T(1/\log(T)) \\
&\le \left(1 + O_T\left(\frac{1}{\log^{10}(T)}\right)\right)^{t_2+1-t_1} \cdot O_T\left(\frac{1}{\log(T)}\right)^{\Omega_T\left(\frac{|t_2-t_1|}{\log^8(T)}\right)} && \text{Equation (33)} \\
&\le \left(1 + O_T\left(\frac{1}{\log^{2}(T)}\right)\right)^{(t_2+1-t_1)/\log^8(T)} \cdot O_T\left(\frac{1}{\log(T)}\right)^{\Omega_T\left(\frac{|t_2-t_1|}{\log^8(T)}\right)} && \text{Lemma 11} \\
&\le O_T\left(\left(\frac{1}{\log(T)}\left(1 + \frac{1}{\log^2(T)}\right)\right)\right)^{\Omega_T\left(\frac{|t_2-t_1|}{\log^8(T)}\right)} \\
&\le \left(O_T\left(\frac{1}{\log(T)}\right)\right)^{\Omega_T\left(\frac{|t_2-t_1|}{\log^8(T)}\right)} \\
&\le \left(\left(O_T\left(\frac{1}{\log(T)}\right)\right)^{\Omega_T\left(\frac{1}{\log^8(T)}\right)}\right)^{(t_2+1-t_1)} \\
&\le \left(1 - \frac{2}{\log^9(T)}\right)^{t_2+1-t_1}. && (34)
\end{aligned}$$

This is the desired result. In the last line we used that for sufficiently large T,

$$\begin{aligned}
\left(O_T\left(\frac{1}{\log(T)}\right)\right)^{\Omega_T\left(\frac{1}{\log^8(T)}\right)} &\le \left(\frac{1}{2}\right)^{\Omega_T\left(\frac{1}{\log^8(T)}\right)} \\
&\le \left(\frac{1}{2}\right)^{\frac{4}{\log^9(T)}} \\
&\le \left(1 - \frac{2}{\log^9(T))}\right) && \text{Lemma 11}
\end{aligned}$$

Note that the first inequality above is a very loose bound, however it is what we need to prove the desired lemma. ∎

C.7. Bounding Frequency of Extreme Case (Lemma 12)

To show Equation (33), we will show that for all $t_2 \ge \lceil \log^8(T) \rceil$, conditional on event $A_{t_2} \cap E_{\mathrm{L}10}(x, k, \theta, W')$, for every $j \le t_2 - \lceil \log^8(T) \rceil + 1$ there exists some $i \in [j : j + \lceil \log^8(T) \rceil)$ such that $\mathcal{W}_i$ holds, where we recall that

$$\mathcal{W}_i = \left\{ \min(x_i, y_i) \ge \frac{D_U}{a - bK} \text{ or } \max(x_i, y_i) \le \frac{D_L}{a - bK} \right\}.$$

This in turn implies Equation (33) because we can divide $[t_1 + 1 : t_2]$ into $\Omega_T(\frac{|t_2-t_1|}{\log^8(T)})$ disjoint intervals of the form $[j : j + \lceil \log^8(T) \rceil)$ where each interval contains an i such that $\mathcal{W}_i$ holds.

For the rest of the proof, we will prove by contradiction that conditional on event $A_{t_2} \cap E_{\mathrm{L}10}(x, k, \theta, W')$, for every $j \le t_2 - \lceil \log^8(T) \rceil$ there exists some $i \in [j : j + \lceil \log^8(T) \rceil)$ such that $\mathcal{W}_i$ holds. Assume that this is not the case, and there exists j such that there are no $i \in [j : j + \lceil \log^8(T) \rceil)$ such that $\mathcal{W}_i$ holds.

By definition of $\mathcal{W}_i$, if $y_i \notin \left[\frac{D_L}{a-bK} - d_i, \frac{D_U}{a-bK} + d_i\right]$, then $\mathcal{W}_i$ must hold. Recall that conditional on event A_{t_2}, $d_i \leq \frac{3}{\log^{10}(T)}$ for all $i \leq t_2$. Therefore, conditional on event A_{t_2}, if $y_i \notin \left[\frac{D_L}{a-bK} - \frac{3}{\log^{10}(T)}, \frac{D_U}{a-bK} + \frac{3}{\log^{10}(T)}\right]$ then $\mathcal{W}_i$ must hold. Because we assumed that there are no $i \in [j : j + \lceil \log^8(T) \rceil)$ such that $\mathcal{W}_i$ holds, this implies that for all $i \in [j : j + \lceil \log^8(T) \rceil)$,

$$y_i \in \left[\frac{D_L}{a - bK} - \frac{3}{\log^{10}(T)}, \frac{D_U}{a - bK} + \frac{3}{\log^{10}(T)}\right]. \tag{35}$$

We also have that for sufficiently large T,

$$\begin{aligned}
\frac{\|D\|_\infty}{a - bK} &\leq \frac{\|D\|_\infty}{a^* - b^*K - O_T\left(\frac{1}{\log^{10} T}\right)} && \|\theta - \theta^*\|_\infty \leq 1/\log^{10}(T) \\
&\leq \frac{\|D\|_\infty}{1 - O_T\left(\frac{1}{\log^9 T}\right)} && |1 - (a^* - b^*K)| \leq \frac{1}{\log^9(T)} \\
&\leq 2\|D\|_\infty \\
&\leq 2\log^2(T).
\end{aligned} \tag{36}$$

Therefore, if $|y_i| \geq \log^2(T) \geq 2\log^2(T) + \frac{3}{\log^{10}(T)}$ for sufficiently large T, then $\mathcal{W}_i$ must hold. For the rest of the proof, we will show that if Equation (35) holds for all $i \in [j : j + \lceil \log^8(T) \rceil)$, then at least one such i must satisfy $|y_i| \geq 3\log^2(T)$, which implies that $\mathcal{W}_i$ will hold which is a contradiction.

Lemma 13 *Using the notation and assumptions of Lemma 12, conditional on $A_{t_2} \cap E_{\text{L10}}(x, k, \theta, W')$, if $y_i \in \left[\frac{D_L}{a-bK} - \frac{3}{\log^{10}(T)}, \frac{D_U}{a-bK} + \frac{3}{\log^{10}(T)}\right]$, then $y_{i+1} - y_i \in [w_i - O_T(1/\log^7(T)), w_i + O_T(1/\log^7(T))]$.*

Proof The control at time i is either $-Ky_i$, $\frac{D_U - ay_i}{b}$, or $\frac{D_L - ay_i}{b}$. If the control is $-Ky_i$, then under event E_1^t,

$$\begin{aligned}
|y_{i+1} - y_i - w_i| &= |(a^* - b^*K)y_i - y_i| \\
&= |y_i||1 - (a^* - b^*K)| \\
&= O_T\left(\frac{|y_i|}{\log^9(T)}\right) && \text{Assumed in Lemmas 10, 12, and 13} \\
&= O_T\left(\frac{1}{\log^7(T)}\right). && \text{Under event } E_1^t \text{ by Lemma 56.}
\end{aligned}$$

The control at state y_i is $\frac{D_U - ay_i}{b}$ only when $y_i \geq \frac{D_U}{a-bK}$. Because $y_i \leq \frac{D_U}{a-bK} + \frac{3}{\log^{10}(T)}$, this implies that $\left|y_i - \frac{D_U}{a-bK}\right| \leq \frac{3}{\log^{10}(T)}$, and because $(a - bK) \leq 1$ this implies that $|D_U - (a-bK)y_i| = O_T(1/\log^{10}(T))$.

Therefore, under event E_1^t, when the control at state y_i is $\frac{D_U - ay_i}{b}$,

$$\begin{aligned}
|y_{i+1} - y_i - w_i| &= \left|a^* y_i + b^* \frac{D_U - a y_i}{b} - y_i\right| \\
&\le |(a^* - b^* K) y_i - y_i| + b^* \left|K y_i + \frac{D_U - a y_i}{b}\right| \\
&\le |(a^* - b^* K) - 1||y_i| + \frac{b^*}{b} |D_U - (a - bK) y_i| \\
&\le O_T\left(\frac{|y_i|}{\log^9(T)}\right) + O_T\left(\frac{1}{\log^{10}(T)}\right) \\
&\le O_T\left(\frac{1}{\log^7(T)}\right). && \text{Under event } E_1^t \text{ by Lemma 56}
\end{aligned}$$

A symmetric result holds if the control at state y_i is $\frac{D_L - a y_i}{b}$ (which happens when $y_i \le \frac{D_L}{a - bK}$). This exactly implies the desired result.

■

Using Lemma 13, for $j \le i_1 < i_2 \le j + \lceil \log^8(T) \rceil$ such that $i_2 - i_1 \le \lceil \log^5(T) \rceil$ and sufficiently large T, if $y_i \in \left[\frac{D_L}{a-bK} - \frac{3}{\log^{10}(T)}, \frac{D_U}{a-bK} + \frac{3}{\log^{10}(T)}\right]$ for all $i \in [j : j + \lceil \log^8(T) \rceil])$, then

$$\begin{aligned}
|y_{i_2+1} - y_{i_1}| &\ge \left|\sum_{j=i_1}^{i_2} w_j\right| - O_T\left(\frac{|i_2 - i_1|}{\log^7(T)}\right) \\
&\ge \left|\sum_{j=i_1}^{i_2} w_j\right| - \frac{1}{\log(T)}. && i_2 - i_1 \le \lceil \log^5(T) \rceil \qquad (37)
\end{aligned}$$

By construction, event $E_{\mathrm{L}10}(x, K, \theta, W')$ directly implies that for sufficiently large T, there exists some $i \in [j : j + \lceil \log^8(T) \rceil - \lceil \log^5(T) \rceil - 1]$ such that

$$\left|\sum_{j=i}^{i + \lceil \log^5(T) \rceil} w_j\right| \ge 7 \log^2(T) \ge 2 \cdot 3 \log^2(T) + \frac{1}{\log(T)}. \qquad (38)$$

Combining this with Equation (37) for $i_1 = i$ and $i_2 = i + \lceil \log^5(T) \rceil$, conditional on $A_{t_2} \cap E_{\mathrm{L}10}(x, k, \theta, W')$,

$$|y_{i_2+1} - y_{i_1}| \ge 6 \log^2(T).$$

This implies that either $|y_i|$ or $|y_{i + \lceil \log^5(T) \rceil + 1}|$ is greater than $3 \log^2(T)$. However, as argued above this implies that $\mathcal{W}_i$ or $\mathcal{W}_{i + \lceil \log^5(T) \rceil + 1}$ holds, which is a contradiction. This completes the proof by contradiction.

Appendix D. Total Cost is Continuous in Truncation Parameter (Lemma 2)

D.1. Main Proof

Proof

Let $\epsilon_{\mathrm{L}2} = \frac{1}{\log^{46}(T)}$. We will combine the following two results.

Lemma 14 *Under Assumptions 1–3, for any θ such that $\|\theta - \theta^*\|_\infty = \epsilon \le \frac{1}{\log^{46}(T)}$, the following holds for the class of truncated linear controllers for $t \le T$:*

$$\bar{J}(\theta, C^{\theta}_{K_{\mathrm{opt}}(\theta,t)}, t) - \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,t)}, t) = \tilde{O}_T(\epsilon).$$

The proof of Lemma 14 can be found in Appendix D.2.

Lemma 15 *Under Assumptions 1–3, for any* $\|\theta - \theta^*\|_\infty = \epsilon \leq \frac{1}{\log^{46}(T)}$, $t \leq T$, *and* $K \in [\frac{a-1}{b}, \frac{a}{b}]$,

$$|\bar{J}(\theta^*, C_K^\theta, t) - \bar{J}(\theta, C_K^\theta, t)| = \tilde{O}_T\left(\epsilon + \frac{1}{T^2}\right). \tag{39}$$

The proof of Lemma 15 can be found in Appendix D.3.

Putting together Lemma 14 and Lemma 15 with $K = K_{\text{opt}}(\theta, t)$, we have the desired result that

$$\bar{J}(\theta^*, C^\theta_{K_{\text{opt}}(\theta,t)}, t) - \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}, t) = \tilde{O}_T\left(\epsilon + \frac{1}{T^2}\right).$$

■

D.2. Total Cost Continuous in Dynamics+Parameter (Lemma 14)

Proof First, we will prove some results about $a^*, b^*, K_{\text{opt}}(\theta^*, t)$. Because $b, b^* \geq \underline{b}$ and $\|\theta - \theta^*\|_\infty = \epsilon \leq \frac{1}{\log^{46}(T)} < b/2$ for large enough T, we have that

$$\left|\left(\frac{a^*}{b^*}\right)^2 - \left(\frac{a}{b}\right)^2\right| = \left|\frac{(a^*)^2b^2 - (b^*)^2a^2}{b^2(b^*)^2}\right| \leq \frac{\epsilon^2b^2 + 2\epsilon ab^2 + 2\epsilon ba^2 + \epsilon^2a^2}{b^2(b-\epsilon)^2} = O_T(\epsilon). \tag{40}$$

$$\left|\frac{a}{b} - \frac{a^*}{b^*}\right| = \left|\frac{a^*b - b^*a}{bb^*}\right| \leq \left|\frac{\epsilon b + \epsilon a}{b(b-\epsilon)}\right| = O_T(\epsilon). \tag{41}$$

Let K' be the solution to $a^* - b^*K_{\text{opt}}(\theta^*, t) = a - bK'$. Then

$$K' = \frac{(a-a^*) + b^*K_{\text{opt}}(\theta^*, t)}{b} = K_{\text{opt}}(\theta^*, t) + \frac{(b^*-b)K_{\text{opt}}(\theta^*, t)}{b} + \frac{a-a^*}{b}.$$

Since $K_{\text{opt}}(\theta^*, t) \leq \frac{a^*}{b^*}$ by definition, we have the following two equations:

$$|K' - K_{\text{opt}}(\theta^*, t)| = \left|\frac{(b^*-b)K_{\text{opt}}(\theta^*, t)}{b} + \frac{a-a^*}{b}\right| \leq \left(\frac{a^*}{bb^*} + \frac{1}{b}\right)\epsilon = O_T(\epsilon). \tag{42}$$

$$|(K')^2 - (K_{\text{opt}}(\theta^*, t))^2| \leq |K' - K_{\text{opt}}(\theta^*, t)| \cdot |K' + K_{\text{opt}}(\theta^*, t)| = O_T(\epsilon). \tag{43}$$

By the choice of K', using the controller $C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}$ under dynamics θ^* results in the exact same sequence of states as using the controller $C^\theta_{K'}$ under dynamics θ. This is because $a - bK' = a^* - b^*K_{\text{opt}}(\theta^*, t)$, which by construction of truncated linear controllers implies that $ax + bC^\theta_{K'}(x) = a^* + b^*C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}$ for all x. The controls will however be different, and we will now bound that difference in controls.

Define $x_0, x_1, ..., x_t$ as the sequence of states when using controller $C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}$ under dynamics θ^* starting at state $x_0 = 0$. Then we have the following result.

$$\left|rC^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}(x_i)^2 - rC^\theta_{K'}(x_i)^2\right| = \begin{cases} \left|rx_i^2\left((K_{\text{opt}}(\theta^*, t))^2 - (K')^2\right)\right| & \text{if } x_i \in [\frac{D_L}{a^*-b^*K_{\text{opt}}(\theta^*,t)}, \frac{D_U}{a^*-b^*K_{\text{opt}}(\theta^*,t)}] \\ \left|r\left(\frac{D_U - a^*x_i}{b^*}\right)^2 - r\left(\frac{D_U - ax_i}{b}\right)^2\right| & \text{if } x_i > \frac{D_U}{a^*-b^*K_{\text{opt}}(\theta^*,t)} \\ \left|r\left(\frac{D_L - a^*x_i}{b^*}\right)^2 - r\left(\frac{D_L - ax_i}{b}\right)^2\right| & \text{if } x_i < \frac{D_L}{a^*-b^*K_{\text{opt}}(\theta^*,t)} \end{cases} \tag{44}$$

By Equation (43), this implies the following.

$$\begin{aligned}
&\left|rC^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}(x_i)^2 - rC^{\theta}_{K'}(x_i)^2\right| \\
&\leq \begin{cases}
O_T(x_i^2\epsilon) & \text{if } x_i \in [\frac{D_L}{a^*-b^*K_{\text{opt}}(\theta^*,t)}, \frac{D_U}{a^*-b^*K_{\text{opt}}(\theta^*,t)}] \\
rD_U^2\left|\left(\frac{1}{b^*}\right)^2 - \left(\frac{1}{b}\right)^2\right| + 2D_U r|x_i|\left|\frac{a}{b} - \frac{a^*}{b^*}\right| + rx_i^2\left|\left(\frac{a^*}{b^*}\right)^2 - \left(\frac{a}{b}\right)^2\right| & \text{if } x_i > \frac{D_U}{a^*-b^*K_{\text{opt}}(\theta^*,t)} \\
rD_L^2\left|\left(\frac{1}{b^*}\right)^2 - \left(\frac{1}{b}\right)^2\right| + 2|D_L| r|x_i|\left|\frac{a}{b} - \frac{a^*}{b^*}\right| + rx_i^2\left|\left(\frac{a^*}{b^*}\right)^2 - \left(\frac{a}{b}\right)^2\right| & \text{if } x_i < \frac{D_L}{a^*-b^*K_{\text{opt}}(\theta^*,t)}
\end{cases}
\end{aligned}$$

By Equations (40) and (41), we get the following result.

$$\left|rC^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}(x_i)^2 - rC^{\theta}_{K'}(x_i)^2\right| \leq \begin{cases}
O_T(x_i^2)\epsilon & \text{if } x_i \in [\frac{D_L}{a^*-b^*K_{\text{opt}}(\theta^*,t)}, \frac{D_U}{a^*-b^*K_{\text{opt}}(\theta^*,t)}] \\
O_T(D_U^2\epsilon + D_U|x_i|\epsilon) + O_T(x_i^2\epsilon) & \text{if } x_i > \frac{D_U}{a^*-b^*K_{\text{opt}}(\theta^*,t)} \\
O_T(D_L^2\epsilon + |D_L||x_i|\epsilon) + O_T(x_i^2\epsilon) & \text{if } x_i < \frac{D_L}{a^*-b^*K_{\text{opt}}(\theta^*,t)}
\end{cases} \tag{45}$$

Using that $\|D\|_\infty \leq \log^2(T)$, in all three cases we have that

$$\left|rC^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}(x_i)^2 - rC^{\theta}_{K'}(x_i)^2\right| = \tilde{O}_T\left(1 + |x_i| + |x_i|^2\right)\epsilon. \tag{46}$$

The last fact we need is to note that x_i is a sequence of states for the controller $C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}$ under dynamics θ^*, which by construction will always satisfy that $D_L \leq a^*x_i + b^*C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}(x^*) \leq D_U$. Therefore, since $\mathbb{E}[|w_{i-1}|]$ and $\mathbb{E}[w_{i-1}^2]$ are constants relative to T that depend on $\mathcal{D}$, for all i,

$$\mathbb{E}[|x_i|] \leq \|D\|_\infty + \mathbb{E}[|w_{i-1}|] = O_T(\log^2(T)).$$

$$\mathbb{E}[|x_i|^2] \leq \|D\|_\infty^2 + \mathbb{E}[w_{i-1}^2] + 2\|D\|_\infty \mathbb{E}[|w_{i-1}|] = O_T(\log^4(T)).$$

Therefore, we can upper bound the difference in cost as follows:

$$\begin{aligned}
\bar{J}(\theta, C^{\theta}_{K'}, t) - \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}, t) &\leq \mathbb{E}\left[\frac{1}{t}\sum_{i=0}^{t-1}\left|rC^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}(x_i)^2 - rC^{\theta}_{K'}(x_i)^2\right|\right] \\
&\leq \frac{1}{t}\sum_{i=0}^{t-1}\tilde{O}_T\left(1 + \mathbb{E}[|x_i|] + \mathbb{E}[|x_i|^2]\right)\epsilon \qquad \text{Equation (46)} \\
&\leq \frac{1}{t}\sum_{i=0}^{t-1}\tilde{O}_T\left(\log^2(T) + \log^4(T)\right)\epsilon \\
&= \tilde{O}_T\left(\epsilon\right).
\end{aligned}$$

Finally, by definition of K_{opt} we know that

$$\bar{J}(\theta, C^{\theta}_{K_{\text{opt}}(\theta,t)}, t) \leq \bar{J}(\theta, C^{\theta}_{K'}, t),$$

therefore we can conclude that

$$\bar{J}(\theta, C^{\theta}_{K_{\text{opt}}(\theta,t)}, t) - \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*,t)}, t) = \tilde{O}_T(\epsilon).$$

■

D.3. Total Cost Continuous in Parameter (Lemma 15)

Proof For a set of time varying dynamics $\{\theta_j\}_{j=0}^{t-1}$ where $\theta_j \in \Theta$ for all j, we define the expected total cost for varying dynamics as

$$\bar{J}(\{\theta_j\}_{j=0}^{t-1}, C_K^\theta, t) := qx_t^2 + \sum_{j=0}^{t-1} qx_j^2 + rC_K^\theta(x_{j-1})^2,$$

where $x_0 = 0$ and $x_j = a_{j-1}x_{j-1} + b_{j-1}C_K^\theta(x_{j-1}) + w_{j-1}$. In other words, this is the total cost if the dynamics at time $j < t$ are θ_j.

For $i \in [0:t]$, let $\{\theta_j^i\}_{j=0}^{t-1}$ be a time varying dynamics with $\theta_j^i = \theta$ for all $j < i$ and $\theta_j^i = \theta^*$ for $j \geq i$. We will now compare the costs under dynamics $\{\theta_j^i\}_{j=0}^{t-1}$ versus under $\{\theta_j^{i+1}\}_{j=0}^{t-1}$. Let $x_0, x_1, ...x_t$ be the states when using controller C_K^θ under time-varying dynamics $\{\theta_j^i\}_{j=0}^{t-1}$ and $x_0^*, ...x_t^*$ be the states when using controller C_K^θ under time-varying dynamics $\{\theta_j^{i+1}\}_{j=0}^{t-1}$ (both starting at $x_0 = x_0^* = 0$). Up until time i, the dynamics of these two trajectories are the same (both equal to θ), and therefore the states and controls of the two trajectories are equivalent up until time i. Because C_K^θ is safe with respect to dynamics θ, $|x_i^*| = |x_i| \leq \|D\|_\infty + |w_{i-1}|$. Because $\|D\|_\infty \leq \log^2(T)$, this implies that

$$\mathbb{E}[|x_i^*|] = \mathbb{E}[|x_i|] = \tilde{O}_T(1). \tag{47}$$

Also note that by construction of the truncated linear controller, $|C_K^\theta(x_i)| \leq K|x_i| + \frac{\|D\|_\infty + a|x_i|}{b}$. Therefore, we have that

$$|x_{i+1} - x_{i+1}^*| = |ax_i + bC_K^\theta(x_i) - a^*x_i - b^*C_K^\theta(x_i)| \leq \epsilon|x_i| + \epsilon|C_K^\theta(x_i)| \leq \epsilon\left(|x_i| + K|x_i| + \frac{\|D\|_\infty + a|x_i|}{b}\right). \tag{48}$$

Combining Equations (47) and (48) gives that

$$\mathbb{E}[|x_{i+1} - x_{i+1}^*|] = \tilde{O}_T(\epsilon). \tag{49}$$

Consider x_{i+1}. Define the event $F = \{|x_{i+1}| < \log^3(T)\}$. As argued above, $|x_i| \leq \|D\|_\infty + |w_{i-1}| \leq 2\log^2(T)$ under event E_1. Furthermore, the control $C_K^\theta(x_i)$ is safe with respect to dynamics θ_i^i and $\|\theta_i^i - \theta^*\|_\infty = \|\theta - \theta^*\|_\infty \leq 1/\log^{46}(T) \leq 1/\log(T)$ for sufficiently large T. Therefore, we can apply Lemma 57 for one step to get that for sufficiently large T, $|x_{i+1}| \leq 4\log^2(T)$ under event E_1. Therefore, for sufficiently large T, $\mathbb{P}(F) \geq \mathbb{P}(E_1) = 1 - o_T(1/T^{11})$. By Lemma 54 (using the same logic as in Equation (56) in Schiffer and Janson (2024)), this implies that

$$\mathbb{P}(|x_{i+1}| \geq \log^3(T))\,\mathbb{E}[|x_{i+1}|^2 \mid |x_{i+1}| \geq \log^3(T)] = o_T(1/T^{10}).$$

The same logic holds for x_{i+1}^*. We showed above that $\mathbb{P}(|x_{i+1}| \leq 4\log^2(T)) = 1 - o_T(1/T^{11})$ (and the same equation holds for x_{i+1}^*). Therefore, we can apply Lemma 55 to get that

$$\begin{aligned}
&|t \cdot J^*(\{\theta_j^i\}_{j=0}^{t-1}, C_K^\theta, t, 0) - t \cdot J^*(\{\theta_j^{i+1}\}_{j=0}^{t-1}, C_K^\theta, t, 0)| \\
&= \mathbb{E}\left[\left|(t-i)\bar{J}(\theta^*, C_K^\theta, t-i, x_{i+1}) - (t-i)\bar{J}(\theta^*, C_K^\theta, t-i, x_{i+1}^*)\right|\right] \\
&= \tilde{O}_T\left(\mathbb{E}\left[\left|x_{i+1} - x_{i+1}^*\right|\right] + \epsilon + \frac{1}{T^2}\right) && \text{Lemma 55} \\
&= \tilde{O}_T\left(\epsilon + \frac{1}{T^2}\right). && \text{Equation (49)}
\end{aligned} \tag{50}$$

Now, we conclude by noting that

$$|t \cdot J^*(\theta^*, C_K^\theta, t) - t \cdot J^*(\theta, C_K^\theta, t)| = \left| \sum_{i=0}^{t} t \cdot J^*(\{\theta_j^{i+1}\}_{j=0}^{t-1}, C_K^\theta, t, 0) - t \cdot J^*(\{\theta_j^i\}_{j=0}^{t-1}, C_K^\theta, t, 0) \right|$$
$$= \tilde{O}_T\left(t\left(\epsilon + \frac{1}{T^2}\right)\right),$$

and dividing both sides of the equation by t gives the desired result. ■

Appendix E. Algorithm 2 Guarantees $\tilde{O}_T(\sqrt{T})$ Regret (Theorem 2)

E.1. Main Proof

For the proof of Theorem 2, recall the following notation (which was also defined in the proof sketch of Theorem 2). Define $\mathcal{C}^{\text{unc}} = \{C_K^{\text{unc}}\}_{K\in\mathbb{R}}$ as the class of untruncated linear controllers, where $C_K^{\text{unc}}(x) = -Kx$. For any controller C and dynamics θ, define $\bar{J}(\theta, C) = \lim_{T\longrightarrow\infty} \bar{J}(\theta, C, T)$. Define $K_{\text{opt}}(\theta) = \arg\sup_K \bar{J}(\theta, C_K^\theta)$ and $F_{\text{opt}}(\theta) = \arg\sup_K \bar{J}(\theta, C_K^{\text{unc}})$.

First we note that Lemmas 2 and 3 combined with Schiffer and Janson (2024) directly implies the desired result when $\mathcal{D}$ has infinite support. More formally, by Lemmas 2 and 3, the class of truncated linear controllers satisfies the assumptions of Theorem 1 in Schiffer and Janson (2024). If $\mathcal{D}$ has infinite support and $\|D\|_\infty = O_T(1)$, then Assumption 9 in Schiffer and Janson (2024) is satisfied. Furthermore, for noise distribution with infinite support, Algorithm 2 will choose the exact same controls as Algorithm 3 in Schiffer and Janson (2024). Therefore, under Assumptions 1–3, if $\mathcal{D}$ has infinite support, then Algorithm 2 with the baseline class of truncated linear controllers has regret of $\tilde{O}_T(\sqrt{T})$ by Theorem 1 in Schiffer and Janson (2024). Therefore, Theorem 1 in Schiffer and Janson (2024) directly proves Theorem 2 in the case when $\mathcal{D}$ has infinite support. For the rest of this proof, we will focus on proving Theorem 2 when $\mathcal{D}$ has bounded support, therefore making the following assumption.

Assumption 4 *The distribution $\mathcal{D}$ has bounded support, i.e. there exists $\bar{w} > 0$ such that $\mathbb{P}_{w\sim\mathcal{D}}(|w| \le \bar{w}) = 1$.*

For the rest of the proof of Theorem 2, we will also assume WLOG that $D_U \le |D_L|$.

Definition 16 *Define $K_{D_U}^\theta$ as the value that satisfies the equation*

$$\frac{D_U}{a - bK_{D_U}^\theta} - D_U = \bar{w}.$$

For the rest of Appendix E, let C^{alg} be the controller of Algorithm 2 and $\mathcal{C}_{\text{tr}}^\theta$ be the class of truncated linear controllers for dynamics θ as in Equation (160).

Let $s_e = \log_2(\sqrt{T}) - 1$, and let

$$E_0 := \left\{\forall s \in [0 : s_e] : \|\theta^* - \hat{\theta}_s^{\text{pre}}\|_\infty \le \epsilon_s\right\}. \tag{51}$$

We will use the following lemma (Lemma 23 in Schiffer and Janson (2024)) which bounds the uncertainty in θ^* from regularized least squares estimation.

Lemma 17 (Lemma 23 in Schiffer and Janson (2024), Theorem 1 in Abbasi-Yadkori and Szepesvári (2011)) *Suppose x_t and u_t are respectively the state and control at time t when using an arbitrary controller C starting at state $x_0 = 0$. Define $z_t = (x_t, u_t)$ and let $\lambda > 0$. Let $Z_t \in \mathbb{R}^{t\times 2}$ where the ith row is z_{i-1}, let $X_t \in \mathbb{R}^{t\times 1}$ where the ith element is x_i, and let $I \in \mathbb{R}^{2\times 2}$ be the identity matrix. Then under Assumptions 1–3, with probability $1 - o_T\left(\frac{1}{T^2}\right)$ the following holds for all $1 \le t \le T-1$ and for any $S \subseteq [0 : t-1]$:*

$$\|\theta^* - (Z_t^\top Z_t + \lambda I)^{-1} Z_t^\top X_t\|_\infty \le \sqrt{\frac{\max((V_t^S)_{11}, (V_t^S)_{22})}{\det(V_t^S)}} B_t, \tag{52}$$

where $V_t^S = \lambda I + \sum_{s=0}^{t-1} z_s z_s^\top 1_{s\in S}$, $B_t = \alpha\sqrt{\log\left(\det\left(V_t^{[0:t-1]}\right)\right) + \log(\lambda^2) + 2\log(T^2)} + \sqrt{\lambda}(\bar{a}^2 + \bar{b}^2)$, and α is from the subgaussian assumption on the noise distribution $\mathcal{D}$, which implies that there exists an α such that $\mathbb{E}_{w\sim\mathcal{D}}[\exp(\gamma w)] \le \exp(\gamma^2\alpha^2/2)$ for any $\gamma \in \mathbb{R}$.

By Lemma 17 we have that with probability $1 - o_T(1/T^2)$, for all s, $\|\theta^* - \hat{\theta}_s^{\text{pre}}\|_\infty \leq \epsilon_s$.Therefore,

$$\mathbb{P}(E_0) = 1 - o_T(1/T^2).$$

By construction we also have that $\|\hat{\theta}_s - \hat{\theta}_s^{\text{pre}}\|_\infty \leq \epsilon_s$. This implies by the triangle inequality that under event E_0, $\|\hat{\theta}_s - \theta^*\|_\infty \leq 2\epsilon_s$.

We also use the following uncertainty result (Lemma 2 in Schiffer and Janson (2024) rewritten for our algorithm):

Lemma 18 *Under Assumptions 1–3, there exists a $c_{\text{L18}} = \tilde{O}_T(1)$ such that with probability $1 - o_T(1/T^2)$*

$$\max_{s \in [0:s_e]} \epsilon_s \leq c_{\text{L18}} T^{-1/4} = \tilde{O}_T(T^{-1/4}).$$

Note that we explicitly named the constant in Lemma 18 as we will use this constant later in the proof. For the rest of this section, define

$$E_2 := E_0 \bigcap \left\{ \max_{s \in [0:s_e]} \epsilon_s \leq c_{\text{L18}} T^{-1/4} = \tilde{O}_T(T^{-1/4}) \right\}. \tag{53}$$

Lemma 18 implies that we have

$$\mathbb{P}(E_2) = 1 - o_T(1/T^2).$$

Define

$$E_2^0 := \{\epsilon_0 \leq c_{\text{L18}} T^{-1/4}\} \cap \{\|\theta^* - \hat{\theta}_0^{\text{pre}}\|_\infty \leq \epsilon_0\} \subseteq E_2.$$

Recall $\hat{\theta}_{\text{wu}}$, which is defined in Line 3 of Algorithm 2. Because $\hat{\theta}_{\text{wu}} = \hat{\theta}_0^{\text{pre}}$, by the same logic as above, under E_2^0 we have that $\|\theta^* - \hat{\theta}_{\text{wu}}\|_\infty \leq 2\epsilon_0 \leq 2c_{\text{L18}} T^{-1/4}$.

Define E_1 as

$$E_1 = \left\{ \forall t < T : |w_t| \leq \log^2(T) \right\}. \tag{54}$$

and E_{safe} as the following, where x'_t and u'_t are the states and controls respectively of the algorithm:

$$E_{\text{safe}} = \left\{ \forall t < T : D_{\text{L}} \leq a^* x'_t + b^* u'_t \leq D_{\text{U}} \right\}, \tag{55}$$

Finally, we define the event

$$E = E_1 \cap E_2 \cap E_{\text{safe}}.$$

By a union bound we have that $\mathbb{P}(E) = 1 - o_T(1/T^2)$. Using this new notation and Lemma 18, we can proceed to the main proof.

The desired safety of C^{alg} follows from the following lemma:

Lemma 19 *Under Assumptions 1–3 , Algorithm 2 is safe for T steps for dynamics θ^* with probability $1 - o_T(1/T^2)$.*

The proof of Lemma 19 follows directly from the construction of the events E_0, E_1, E_2 and the bounds on their probability (see Lemma 1 in Schiffer and Janson (2024) for more details).

The rest of this section will focus on proving that the regret of Algorithm 2 is $\tilde{O}_T(\sqrt{T})$ with probability $1 - o_T(1/T)$.

Let $C_{\text{switch}} = \frac{c_{\text{E81}} D_{\text{U}}}{c_{\text{L39}}^2} = \tilde{O}_T(1)$ where $c_{\text{E81}} = \tilde{O}_T(1)$ and is defined in Equation (81) and $c_{\text{L39}} = \Omega(1)$ defined in Lemma 39; Equation (81) and Lemma 39 will both appear in Appendix G.1. Note that C_{switch} is used in Line 6 of Algorithm 2. Define the event E_{E56} as

$$E_{\text{E56}} := \left\{ \bar{w} + D_U - \frac{D_U}{\hat{a}_{\text{wu}} - \hat{b}_{\text{wu}} F_{\text{opt}}(\hat{\theta}_{\text{wu}})} \leq C_{\text{switch}} T^{-1/4} \right\}. \tag{56}$$

We will study the regret of Algorithm 2 separately under event E_{E56} and under event $\neg E_{\text{E56}}$. Informally, if E_{E56} holds then the optimal linear controller is close to being safe for dynamics θ^*. If $\neg E_{\text{E56}}$, then the magnitude of the noise is large relative to the constraints, and therefore an argument similar to that of Theorem 1 in Schiffer and Janson (2024) will bound the regret.

Proposition 20 *Under Assumptions 1–3 and 4, there exists an event E_{P20} such that $E_{\text{P20}} \subseteq \neg E_{\text{E56}}$, such that $\mathbb{P}(E_{\text{P20}}) \geq \mathbb{P}(\neg E_{\text{E56}}) - o_T(1/T)$, and such that conditional on event E_{P20}, Algorithm 2 has $\tilde{O}_T(\sqrt{T})$ regret.*

The proof of Proposition 20 can be found in Appendix E.2.

Proposition 21 *Under Assumptions 1–3 and 4, there exists an event E_{P21} such that $E_{\text{P21}} \subseteq E_{\text{E56}}$, such that $\mathbb{P}(E_{\text{P21}}) \geq \mathbb{P}(E_{\text{E56}}) - o_T(1/T)$, and such that conditional on event E_{P21}, Algorithm 2 has $\tilde{O}_T(\sqrt{T})$ regret.*

The proof of Proposition 21 can be found in Appendix E.3.

Combining these two propositions gives that the regret of Algorithm 2 is $\tilde{O}_T(\sqrt{T})$ conditional on $E_{\text{P20}} \cup E_{\text{P21}}$. Because $E_{\text{P20}} \cap E_{\text{P21}} = \emptyset$ by construction, we have that

$$\mathbb{P}(E_{\text{P20}} \cup E_{\text{P21}}) = \mathbb{P}(E_{\text{P20}}) + \mathbb{P}(E_{\text{P21}}) \geq \mathbb{P}(E_{\neg 56}) - o_T(1/T) + \mathbb{P}(E_{\text{E56}}) - o_T(1/T) = 1 - o_T(1/T).$$

Therefore the desired result holds with unconditional probability $1 - o_T(1/T)$, completing the proof of Theorem 2.

E.2. Bounding Regret when Noise is Large (Proposition 20)

Proof We can decompose the regret in the following manner. As in Schiffer and Janson (2024), for any $(K, \{K_s\}_{0 \leq s \leq s_e})$ where $K, K_s \in (\frac{a-1}{b}, \frac{a}{b})$, define $x_0^{(K, \{K_s\}_{0 \leq s \leq s_e})}, x_1^{(K, \{K_s\}_{0 \leq s \leq s_e})}, \ldots$ as the states that result from starting at $x_0 = 0$ and at each time $t < T_0$ using controller $C_K^{\theta^*}$ and at $t \geq T_0$ uses controller $C_{K_s}^{\theta^*}$, where $s = \lfloor \log_2 \left(tT^{-1/2} \right) \rfloor$. Define $(K^*, \{K_s^*\}_{0 \leq s \leq s_e})$ as:

$$\begin{aligned}
&(K^*, \{K_s^*\}_{0 \leq s \leq s_e}) \\
&:= \underset{(K, \{K_s\}_{0 \leq s \leq s_e})}{\arg\min} \mathbb{E}\left[\sqrt{T} J\left(\theta^*, C_K^{\theta^*}, \sqrt{T}, 0, \{w_t\}_{t=0}^{T_0 - 1} \right) + \sum_{s=0}^{s_e} T_s J(\theta^*, C_{K_s}^{\theta^*}, T_s, x_{T_s}^{(K, \{K_s\}_{0 \leq s \leq s_e})}, W_s) \right].
\end{aligned}$$

Define x'_t as the state of the controller of Algorithm 2 at time t. Define $\hat{x}_{T_0}, \hat{x}_{T_0+1}, \ldots$ as the sequence of random variables representing the sequence of states if the control at each time $t \geq T_0$ is $C_{K_{\text{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}(\hat{x}_t)$ for $s = \lfloor \log_2 \left(tT^{-1/2} \right) \rfloor$ and starting at $\hat{x}_{T_0} = x'_{T_0}$.

$$
\begin{aligned}
&T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - T \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,T)}, T) \\
&\leq T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - \mathbb{E}\left[\sqrt{T} J\left(\theta^*, C^{\theta^*}_{K^*}, \sqrt{T}, 0, \{w_t\}_{t=0}^{\sqrt{T}-1}\right) + \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\theta^*}_{K^*_s}, T_s, x^*_{T_s}, W_s)\right] \\
&\leq T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - \mathbb{E}\left[\sum_{s=0}^{s_e} T_s \bar{J}(\theta^*, C^{\theta^*}_{K^*_s}, T_s, x^*_{T_s}, W_s)\right] \\
&= \underbrace{\sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s\right] - \mathbb{E}\left[\sum_{s=0}^{s_e} T_s \bar{J}(\theta^*, C^{\theta^*}_{K^*_s}, T_s, x^*_{T_s}, W_s)\right]}_{R_1} \\
&\qquad + \underbrace{\sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s\right] - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s\right]}_{R_{1b}} \\
&\qquad + \underbrace{\sum_{s=0}^{s_e} T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s, \hat{x}_{T_s}, W_s) - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s\right]}_{R_2} \\
&\qquad + \underbrace{\sum_{s=0}^{s_e} T_s J(\theta^*, C^{\mathrm{alg}}_s, T_s, x'_{T_s}, W_s) - \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s, \hat{x}_{T_s}, W_s)}_{R_3} \\
&\qquad + \underbrace{T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\mathrm{alg}}_s, T_s, x'_{T_s}, W_s)}_{R_0}. \qquad (57)
\end{aligned}
$$

Informally, we will show that with high probability $\epsilon_s = \tilde{O}_T(1/\sqrt{T_s})$ for all s.

Lemma 22 *Under Assumptions 1–3 and 4, there exists event E_{L22} such that $\mathbb{P}(E_{\mathrm{L22}}) = 1 - o_T(1/T)$ and such that conditional on $\neg E_{\mathrm{E56}} \cap E \cap E_{\mathrm{L22}}$,*

$$\max_{s \in [0:s_e]} \epsilon_s \sqrt{T_s} = \tilde{O}_T(1).$$

The proof of Lemma 22 can be found in Appendix F.1. Define event E_3 as

$$E_3 = \left\{ \max_{s \in [0:s_e]} \epsilon_s \sqrt{T_s} = \tilde{O}_T(1) \right\}.$$

Lemma 22 implies that $\neg E_{\mathrm{E56}} \cap E \cap E_{\mathrm{L22}} \subseteq E_3$. Note that compared to the regret decomposition in Schiffer and Janson (2024), there is an extra regret term R_{1b}. This extra regret term can be thought of as the extra regret caused by choosing the best infinite horizon controller instead of the best finite horizon controller. The following lemma bounds the regret of this term by $\tilde{O}_T(\sqrt{T})$.

Proposition 23 *Define R_{1b} as*

$$R_{1b} = \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s\right] - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s\right]. \qquad (58)$$

Under Assumptions 1–3 and 4, conditional on event $E \cap E_3$,

$$R_{1b} = \tilde{O}_T\left(\sqrt{T}\right).$$

The proof of Proposition 23 can be found in Appendix F.2. The following propositions bound the remaining regret terms.

Proposition 24 (Regret from Randomness) *Define R_2 as*

$$R_2 := \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s, \hat{x}_{T_s}, W_s) - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s, 0, W_s) \;\middle|\; \hat{\theta}_s\right].$$

Then under Assumptions 1–3 and 4 there exists an event $E_{\text{P}24}$ such that $\mathbb{P}(E_{\text{P}24}) = 1 - o_T(1/T)$ and conditional on $E_{\text{P}24} \cap \neg E_{\text{E}56} \cap E$,

$$R_2 = \tilde{O}_T(\sqrt{T}). \tag{59}$$

The proof of Proposition 24 can be found in Appendix F.3.

Proposition 25 *Define R_1 as*

$$R_1 := \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s, T_s)}, T_s, 0, W_s) \;\middle|\; \hat{\theta}_s\right] - \mathbb{E}\left[\sum_{s=0}^{s_e} T_s J(\theta^*, C^{\theta^*}_{K^*_s}, T_s, x^*_{T_s}, W_s)\right].$$

Under Assumptions 1–3 and 4, conditional on event $E_3 \cap E$,

$$R_1 = \tilde{O}_T\left(\sqrt{T}\right). \tag{60}$$

The proof of Proposition 25 can be found in Appendix F.4.

Proposition 26 *Define R_3 as (the random variable)*

$$R_3 := \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\text{alg}}_s, T_s, x'_{T_s}, W_s) - \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s, \hat{x}_{T_s}, W_s).$$

Then under Assumptions 1–3 and 4,there exists an event $E_{\text{P}26}$ such that $\mathbb{P}(E_{\text{P}26}) = 1 - o_T(1/T)$ and conditional on $E_{\text{P}26} \cap \neg E_{\text{E}56} \cap E \cap E_3$,

$$R_3 = \tilde{O}_T(\sqrt{T}). \tag{61}$$

The proof of Proposition 26 can be found in Appendix F.5.

Proposition 27 *Under Assumptions 1–3 and 4, conditional on event E,*

$$T \cdot J(\theta^*, C^{\text{alg}}, T, 0, W) - \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\text{alg}}_s, T_s, x'_{T_s}, W_s) = \tilde{O}_T(\sqrt{T}). \tag{62}$$

The proof of Proposition 27 can be found in Appendix F.6.

Using Equation (57) combined with Propositions 27, 23, 24, 25 and 26, conditional on event $\neg E_{\mathrm{E56}} \cap E_3 \cap E \cap E_{\mathrm{P26}} \cap E_{\mathrm{P24}}$ the total regret is upper bounded by

$$T \cdot J(\theta^*, C^{\mathrm{alg}}, T) - T \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,T)}, T) \leq R_0 + R_1 + R_{1b} + R_2 + R_3 = \tilde{O}_T\left(\sqrt{T}\right).$$

Combining Propositions 24 and 26, $\mathbb{P}(E_{\mathrm{P26}} \cap E_{\mathrm{P24}}) = 1 - o_T(1/T)$. Therefore, we have that

$$\begin{aligned}
&\mathbb{P}(E_3 \cap E \cap \neg E_{\mathrm{E56}} \cap E_{\mathrm{P26}} \cap E_{\mathrm{P24}}) \\
&= \mathbb{P}(E_3 \cap E \cap \neg E_{\mathrm{E56}}) - o_T(1/T) && \text{Remark 28} \\
&\geq \mathbb{P}(E_{\mathrm{L22}} \cap E \cap \neg E_{\mathrm{E56}}) - o_T(1/T) && \text{Lemma 22} \\
&\geq \mathbb{P}(\neg E_{\mathrm{E56}}) - o_T(1/T). && \text{Remark 28}
\end{aligned}$$

Above, we twice used the following remark:

Remark 28 *If two events $\mathcal{E}_1$ and $\mathcal{E}_2$ satisfy that $\mathbb{P}(\mathcal{E}_1) = 1 - o_T(1/T)$, then*

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2) = \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) - \mathbb{P}(\mathcal{E}_1 \cup \mathcal{E}_2) \geq \mathbb{P}(\mathcal{E}_2) - o_T(1/T)$$

Taking $E_{\mathrm{P20}} = E_3 \cap E \cap \neg E_{\mathrm{E56}} \cap E_{\mathrm{P26}} \cap E_{\mathrm{P24}}$ gives the desired result. ■

E.3. Bounding Regret when Noise is Small (Proposition 21)

Informally, E_{E56} implies that the optimal linear controller for θ^* is close to satisfying the constraints. Therefore, we will bound the regret by approximating both the best constrained controller and the controller of Algorithm 2 by the optimal unconstrained linear controller.

We will decompose the regret as follows. Define $C^{\mathrm{alg}'}$ to be the controller of Algorithm 2 after the warm-up period, i.e. starting at time $t = T_0$. Therefore, $C^{\mathrm{alg}'}_t = C^{\mathrm{alg}}_{t+T_0}$. Define $x'_0, x'_1, \ldots$ as the series of states when using algorithm C^{alg}. Define $W' = \{w_i\}_{i=T_0}^{T-1}$. Recall that C^{unc}_K is the linear controller such that $C^{\mathrm{unc}}_K(x) = -Kx$. We can decompose the regret as follows:

$$\begin{aligned}
&T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - T \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,T)}, T) \\
&\leq T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - (T - T_0) \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,T)}, T - T_0) \\
&= \underbrace{(T - T_0) \cdot \bar{J}(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T - T_0) - (T - T_0) \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,T)}, T - T_0)}_{R'_1} \\
&\quad + \underbrace{(T - T_0) \cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T - T_0, 0, W') - (T - T_0) \cdot \bar{J}(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T - T_0)}_{R'_2} \\
&\quad + \underbrace{(T - T_0) \cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T - T_0, x'_{T_0}, W') - (T - T_0) \cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T - T_0, 0, W')}_{R'_3} \\
&\quad + \underbrace{(T - T_0) \cdot J(\theta^*, C^{\mathrm{alg}'}, T - T_0, x'_{T_0}, W') - (T - T_0) \cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T - T_0, x'_{T_0}, W')}_{R'_4} \\
&\quad + \underbrace{T \cdot J(\theta^*, C^{\mathrm{alg}}, T, 0, W) - (T - T_0) \cdot J(\theta^*, C^{\mathrm{alg}'}, T - T_0, x'_{T_0}, W')}_{R'_5}. \qquad (63)
\end{aligned}$$

We will now individually analyze each of these components of regret. The first component of regret (R_1') is the extra expected cost of using $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ versus $C^{\theta^*}_{K_{\text{opt}}(\theta^*,T)}$. We will bound that regret with the following proposition.

Proposition 29 *Under Assumptions 1–3 and 4, conditional on event E_2^0,*

$$(T-T_0)\cdot \bar{J}(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0) - (T-T_0)\cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*,T)}, T-T_0) = \tilde{O}_T(\sqrt{T}). \tag{64}$$

The proof of Proposition 29 can be found in Appendix H.1.

The next source of regret (R_2') is the variation in the realization of the $T-T_0$ time step cost versus the expected cost. We will bound this regret with Proposition 30.

Proposition 30 *Under Assumptions 1–3 and 4, there exists an event E_{P30} such that $\mathbb{P}(E_{\text{P30}}) = 1 - o_T(1/T)$ and such that conditional on event E_{P30},*

$$\left|(T-T_0)\cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, 0, W') - (T-T_0)\cdot \bar{J}(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0)\right| = \tilde{O}_T(\sqrt{T}). \tag{65}$$

The proof of Proposition 30 can be found in Appendix H.2.

The next source of regret (R_3') comes from the starting state of the controller $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$. We will bound this regret with Proposition 31.

Proposition 31 *Under Assumptions 1–3 and 4, conditional on event E,*

$$\left|(T-T_0)\cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, x'_{T_0}, W') - (T-T_0)\cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, 0, W')\right| = \tilde{O}_T(1). \tag{66}$$

The proof of Proposition 31 can be found in Appendix H.3.

The next component of regret (R_4') is the additional cost of enforcing safety on top of the controller $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$. Define event $E^{\text{wu}}_{\text{safe}}$ as the event that the first $\sqrt{T}$ controls used by controller C^{alg} are safe for dynamics θ^*.

Proposition 32 *Under Assumptions 1–3 and 4, there exists an event E_{P32} such that $\mathbb{P}(E_{\text{P32}} \mid E_{\text{E56}} \cap E_2^0 \cap E^{\text{wu}}_{\text{safe}}) = 1 - o_T(1/T)$ and such that conditional on $E_{\text{E56}} \cap E_2^0 \cap E^{\text{wu}}_{\text{safe}} \cap E_{\text{P32}}$,*

$$\left|(T-T_0)\cdot J(\theta^*, C^{\text{alg}'}, T-T_0, x'_{T_0}, W') - (T-T_0)\cdot \bar{J}(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, x'_{T_0}, W')\right| = \tilde{O}_T(\sqrt{T}). \tag{67}$$

The proof of Proposition 32 can be found in Appendix H.4.

The last source of regret is the regret from the warm-up period. By Proposition 27, this source of regret is $\tilde{O}(\sqrt{T})$ conditional on event E, because by definition $T\cdot J(\theta^*, C^{\text{alg}}, T, 0, W) - (T-T_0)\cdot J(\theta^*, C^{\text{alg}'}, T-T_0, x'_{T_0}, W') = T\cdot J(\theta^*, C^{\text{alg}}, T, 0, W) - \sum_{s=0}^{s_e} T_s J(\theta^*, C^{\text{alg}}_s, T_s, x'_{T_s}, W_s)$.

Recall that $E \subseteq E_2^0 \cap E^{\text{wu}}_{\text{safe}}$. Therefore, conditional on $E_{\text{P32}} \cap E_{\text{P30}} \cap E \cap E_{\text{E56}}$, by Equation (63) and Propositions 29, 30, 31, 32, and 27, we have that

$$T\cdot J(\theta^*, C^{\text{alg}}, T, 0, W) - T\cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*,T)}, T) = \tilde{O}_T(\sqrt{T}).$$

Furthermore, because $\mathbb{P}(E_2^0 \cap E_{\text{safe}}^{\text{wu}}) \geq \mathbb{P}(E) \geq 1 - o_T(1/T)$, we have that

$$\begin{aligned}
&\mathbb{P}(E_{\text{P32}} \cap E_{\text{P30}} \cap E \cap E_{\text{E56}}) \\
&= \mathbb{P}(E_{\text{P32}} \cap E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E \cap E_{\text{E56}}) - o_T(1/T) && \text{Remark 28} \\
&= \mathbb{P}(E_{\text{P32}} \cap E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}) - o_T(1/T) && \text{Remark 28} \\
&= \mathbb{P}(E_{\text{P32}} \mid E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}) \mathbb{P}(E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}) - o_T(1/T) \\
&\geq (1 - o_T(1/T)) \mathbb{P}(E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}) - o_T(1/T) \\
&= \mathbb{P}(E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}) - o_T(1/T) \\
&= \mathbb{P}(E_{\text{E56}}) - o_T(1/T). && \text{Remark 28}
\end{aligned}$$

Taking $E_{\text{P21}} = E_{\text{P32}} \cap E_{\text{P30}} \cap E \cap E_{\text{E56}}$ gives the desired result.

Appendix F. Bounding Sources of Regret for Large Noise Case

F.1. Bounding the Error ϵ_s for Large Noise (Lemma 22)

Proof We will use the following equivalent version of Lemma 26 in Schiffer and Janson (2024) for Algorithm 2.

Lemma 33 *Let x_t, u_t respectively be the state and control of C^{alg} (the controller of Algorithm 2) at time t starting at $x_0 = 0$. Define $G_i = (x_0, u_0, ..., x_{i-1}, u_{i-1})$. For constant $\gamma > 0$, define S_t as*

$$S_t = \left\{ i < t : u_i = u_i^{\text{safeU}} \text{ and } \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, E) \geq \gamma \right\}. \tag{68}$$

Then under Assumptions 1–3 and for sufficiently large T, with probability $1 - o_T(1/T)$,

$$\max_{s \in [0:s_e]} \epsilon_s \sqrt{|S_{T_s}|} = \tilde{O}_T(1). \tag{69}$$

Proof By Lemmas 2 and 3, the class of truncated linear controllers satisfy all of the assumptions of Lemma 26 in Schiffer and Janson (2024). Therefore, Lemma 33 follows directly from the proof of Lemma 26 in Schiffer and Janson (2024). ■

While we have not yet explained the significance of Lemma 34, we state it here because the definition of ϵ^* is needed for other definitions below.

Lemma 34 *Define*

$$\epsilon^* := \bar{w} - \left(\frac{D_U}{a^* - b^* K_{\text{opt}}(\theta^*)} - D_U \right). \tag{70}$$

Then event $\neg E_{\text{E56}} \cap E$ can only hold if $\epsilon^* > 0$.

The proof of Lemma 34 can be found in Appendix G.1.

Define $\gamma_\epsilon = \frac{\mathbb{P}_{w \sim \mathcal{D}}(w \geq \bar{w} - 3\epsilon^*/8)}{2}$ (which is a constant) and define S_t' as

$$S_t' := \left\{ i < t : u_i = u_i^{\text{safeU}} \text{ and } \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, E) \geq \gamma_\epsilon \right\}. \tag{71}$$

Note that this is the same as the definition of S_t in Lemma 33 except with $\gamma = \gamma_\epsilon$.

Lemma 35 *Under Assumptions 1–3 and 4, there exists an event E_{L35} such that $\mathbb{P}(E_{\mathrm{L35}}) \geq 1 - o_T(1/T)$ and such that conditional on event $E_{\mathrm{L35}} \cap \neg E_{\mathrm{E56}}$,*

$$\max_{s \in [1:s_e]} \frac{T_s}{\left|S'_{T_s}\right|} = \tilde{O}_T(1).$$

The proof of Lemma 35 can be found in Appendix G.2.

Define E_{L33} as the event that Equation (69) holds for $S_{T_s} = S'_{T_s}$. Then $\mathbb{P}(E_{\mathrm{L33}}) = 1 - o_T(1/T)$ by Lemma 33. By Lemma 35, conditional on event $E_{\mathrm{L33}} \cap E_{\mathrm{L35}} \cap \neg E_{\mathrm{E56}}$,

$$\max_{s \in [1:s_e]} \epsilon_s \sqrt{T_s} \leq \sqrt{\max_{s \in [1:s_e]} \frac{T_s}{\left|S'_{T_s}\right|}} \left(\max_{s \in [1:s_e]} \epsilon_s \sqrt{\left|S'_{T_s}\right|} \right) = \tilde{O}_T(1).$$

Under event E_2, we also have that $\epsilon_0 \sqrt{T_0} = \tilde{O}_T(T^{-1/4}) T^{1/4} = \tilde{O}_T(1)$. Because $E \subseteq E_2$ this implies that conditional on E, we have $\epsilon_0 \sqrt{T_0} = \tilde{O}_T(1)$.

Therefore, conditional on $E_{\mathrm{L33}} \cap E_{\mathrm{L35}} \cap \neg E_{\mathrm{E56}} \cap E$,

$$\max_{s \in [0:s_e]} \epsilon_s \sqrt{T_s} = \tilde{O}_T(1).$$

Taking $E_{\mathrm{L22}} = E_{\mathrm{L33}} \cap E_{\mathrm{L35}}$ gives the desired result because $\mathbb{P}(E_{\mathrm{L22}}) = 1 - o_T(1/T)$ by a union bound. ■

F.2. Regret from Choosing Best Infinite-Time Controller (Proposition 23

Proof The goal of this proposition is to show that using the infinite horizon controller is not significantly worse than using the finite horizon controller. This proof will use the following lemma.

Lemma 36 *Under Assumptions 1–3 and 4, for any $\theta \in \Theta$ and $K \in [\frac{a-1}{b}, \frac{a}{b}]$,*

$$|\bar{J}(\theta, C_K^\theta, T) - \bar{J}(\theta, C_K^\theta)| = \tilde{O}_T\left(\frac{1}{T}\right).$$

The proof of Lemma 36 can be found in Appendix G.3.

We can apply Lemma 36 to get the following two equations:

$$\left|\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s) - \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)})\right| = \tilde{O}_T\left(\frac{1}{T_s}\right) \tag{72}$$

$$\left|\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, T_s) - \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)})\right| = \tilde{O}_T\left(\frac{1}{T_s}\right). \tag{73}$$

By definition, we also also have the following two inequalities.

$$\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, T_s) \leq \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s) \tag{74}$$

$$\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}) \leq \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}). \tag{75}$$

Combining Equations (72)–(75), we have that

$$\begin{aligned}
\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, T_s) &\geq \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}) - \tilde{O}_T\left(\frac{1}{T_s}\right) && \text{Equation (73)} \\
&\geq \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}) - \tilde{O}_T\left(\frac{1}{T_s}\right) && \text{Equation (75)} \\
&\geq \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s) - \tilde{O}_T\left(\frac{1}{T_s}\right). && \text{Equation (72).}
\end{aligned}$$

Combining this with Equation (74) gives that

$$\left|\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s) - \bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s)\right| = \tilde{O}_T\left(\frac{1}{T_s}\right). \tag{76}$$

This is almost the desired result, but to bound the regret term R_{1b} we need to bound the difference under dynamics θ^*, not under $\hat{\theta}_s$. Conditional on event E, $\|\hat{\theta}_s - \theta^*\|_\infty = \tilde{O}_T(T^{-1/4}) \leq \frac{1}{\log^{46}(T)}$ for sufficiently large T, and therefore Lemma 15 implies the following inequalities for sufficiently large T:

$$\left|\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s) - \bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s)\right| = \tilde{O}_T\left(\|\hat{\theta}_s - \theta^*\|_\infty + \frac{1}{T^2}\right) \tag{77}$$

$$\left|\bar{J}(\hat{\theta}_s, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s) - \bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s)\right| = \tilde{O}_T\left(\|\hat{\theta}_s - \theta^*\|_\infty + \frac{1}{T^2}\right). \tag{78}$$

Putting together Equations (76), (77), (78), and the fact that $T_s \leq T^2$, we have

$$\left|\bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s) - \bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s)\right| \leq \tilde{O}_T\left(\|\hat{\theta}_s - \theta^*\|_\infty + \frac{1}{T_s}\right). \tag{79}$$

Now we are ready to use Equation (79) to bound R_{1b} conditional on event $E \cap E_3$:

$$\begin{aligned}
&R_{1b} \\
&= \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s, 0, W_s) \;\middle|\; \hat{\theta}_s\right] - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s, W_s) \;\middle|\; \hat{\theta}_s\right] \\
&= \sum_{s=0}^{s_e} T_s \bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s) - \sum_{s=0}^{s_e} T_s \bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s) \\
&\leq \sum_{s=0}^{s_e} T_s \left|\bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s)}, T_s) - \bar{J}(\theta^*, C^{\hat{\theta}_s}_{K_{\text{opt}}(\hat{\theta}_s,T_s)}, T_s)\right| \\
&= \tilde{O}_T\left(\sum_{s=0}^{s_e} T_s \left(\|\hat{\theta}_s - \theta^*\|_\infty + \frac{1}{T_s}\right)\right). && \text{Eq (79)} \\
&= \tilde{O}_T\left(s_e + \sum_{s=0}^{s_e} T_s \epsilon_s\right) && \text{Event } E \\
&= \tilde{O}_T(\sqrt{T}) && \text{Event } E_3
\end{aligned}$$

The last line follows from the fact that $s_e = \tilde{O}_T(1)$ and that under event E_3, $T_s\epsilon_s = \sqrt{T_s}\left(\epsilon_s\sqrt{T_s}\right) = \tilde{O}_T(\sqrt{T_s}) = \tilde{O}_T(\sqrt{T})$, ■

F.3. Regret From Randomness (Proposition 24)

Because the events E and E_3 are defined equivalently to the events in Appendix F in Schiffer and Janson (2024), this proof is very similar to the proof of Proposition 8 in Schiffer and Janson (2024) with the events and variables with respect to Algorithm 2 in this paper instead of Algorithm 3 in Schiffer and Janson (2024). There are two differences between this proof and that of Proposition 8 in Schiffer and Janson (2024). The first difference is that the subscript on the controller is $K_{\text{opt}}(\hat{\theta}_s)$ rather than $K_{\text{opt}}(\hat{\theta}_s, T_s)$. The proof of Proposition 8 in Schiffer and Janson (2024) follows the proof of Proposition 5 in Schiffer and Janson

(2024), and primarily relies on analogous versions of Lemma 6 in Schiffer and Janson (2024) and Lemma 7 in Schiffer and Janson (2024). Examining the proofs of these lemmas, the proofs (and analogous results) hold for any controller $C_K^{\hat{\theta}_s}$ where $K \in [K_{\mathrm{L}}^{\hat{\theta}_s}, K_{\mathrm{U}}^{\hat{\theta}_s}]$. This is because the value of K is not used anywhere in the proof. Therefore, analogous versions of these lemmas hold for Algorithm 2 with $K_{\mathrm{opt}}(\hat{\theta}_s)$ instead of $K_{\mathrm{opt}}(\hat{\theta}_s, T_s)$.

The second major difference is that Proposition 8 in Schiffer and Janson (2024) state that the result holds conditional on E with high probability, while Proposition 24 holds conditional on $E \cap E_{\mathrm{P24}}$. In the proof of Proposition 5 in Schiffer and Janson (2024) (specifically Equation (45) in Schiffer and Janson (2024)), we can define the event

$$E_{\mathrm{E45}} := \left\{ \sum_{s=0}^{s_e} T_s J(\theta^*, C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, T_s, 0, W_s) - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s \right] \geq \tilde{O}_T(\sqrt{T}) \right\}.$$

Note that we replaced $K_{\mathrm{opt}}(\hat{\theta}_s, T_s)$ with $K_{\mathrm{opt}}(\hat{\theta}_s)$ for reasons discussed in the previous paragraph. Equation (45) in Schiffer and Janson (2024) implies that $\mathbb{P}(E_{\mathrm{E45}}) = 1 - o_T(1/T)$. Looking at the last sentence of the proof of Proposition 5 in Schiffer and Janson (2024), we have that conditional on $E \cap E_{\mathrm{E45}} \cap \bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, W_s)$,

$$\sum_{s=0}^{s_e} T_s J(\theta^*, C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, T_s, \hat{x}_{T_s}, W_s) - \sum_{s=0}^{s_e} \mathbb{E}\left[T_s J(\theta^*, C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, T_s, 0, W_s) \,\middle|\, \hat{\theta}_s \right] \leq \tilde{O}_T(\sqrt{T}). \tag{80}$$

Furthermore, because by construction $\mathbb{P}(E_{\mathrm{L3}}(C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, W_s)) = 1 - o_T(1/T^{10})$, we have by a union bound that $\mathbb{P}(E_{\mathrm{E45}} \cap \bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, W_s)) = 1 - o_T(1/T)$. Therefore, we can take $E_{\mathrm{P24}} = E_{\mathrm{E45}} \cap \bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C_{K_{\mathrm{opt}}(\hat{\theta}_s)}^{\hat{\theta}_s}, W_s)$ to get the desired result of Proposition 24.

F.4. Regret from Parameter Estimation (Proposition 25)

Because the event E and E_3 are defined equivalently to the events in Appendix F in Schiffer and Janson (2024), this proof is exactly identical to the proof of Proposition 9 in Schiffer and Janson (2024) with the events and variables with respect to Algorithm 2 in this paper instead of Algorithm 3 in Schiffer and Janson (2024).

F.5. Regret from Enforcing Safety (Proposition 26)

Because the events E and E_3 are defined analogously to the events in Appendix F in Schiffer and Janson (2024), this proof is very similar to the proof of Proposition 10 in Schiffer and Janson (2024) with the events and variables with respect to Algorithm 2 of this paper instead of Algorithm 3 in Schiffer and Janson (2024). Other than this redefining of events and variables, there are just two differences.

The first difference between Proposition 26 of this paper and Proposition 10 in Schiffer and Janson (2024) is that the subscript on the controller is $K_{\mathrm{opt}}(\hat{\theta}_s)$ rather than $K_{\mathrm{opt}}(\hat{\theta}_s, T_s)$. The proof of Proposition 10 in Schiffer and Janson (2024) follows the proof of Proposition 6 in Schiffer and Janson (2024) and analogous versions of Lemma 9 in Schiffer and Janson (2024), Lemma 10 in Schiffer and Janson (2024), and Lemma 16 in Schiffer and Janson (2024). These lemmas all hold when the controller $C_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}^{\hat{\theta}_s}$ is replaced with $C_K^{\hat{\theta}_s}$ for any $K \in [K_{\mathrm{L}}^{\hat{\theta}_s}, K_{\mathrm{U}}^{\hat{\theta}_s}]$ (because the proofs do not depend on the value of K). Therefore, analogous versions of these three lemmas hold for Algorithm 2 with $K_{\mathrm{opt}}(\hat{\theta}_s, T_s)$ replaced with $K_{\mathrm{opt}}(\hat{\theta}_s)$.

The second difference is that Proposition 10 in Schiffer and Janson (2024) shows a bound that holds with high probability conditional on $E \cap E_3$, while Proposition 26's bound holds conditional on $E \cap E_3 \cap E_{\mathrm{P26}}$. Examining the proof of Proposition 6 in Schiffer and Janson (2024) (which is the same as the proof of Proposition 10 in Schiffer and Janson (2024)), the high probability event comes from Lemma 9 in Schiffer and Janson (2024), and that high probability event comes from Lemma 16 in Schiffer and Janson (2024). Looking at the proof of Lemma 16 in Schiffer and Janson (2024), the final result is proven conditional on event E with conditional probability $1 - o_T(1/T^9)$. However, this "with conditional probability" is coming from the event $\bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s, T_s)}, W_s)$. Therefore, by Equation (75) in Schiffer and Janson (2024) and the last sentence in the proof of Lemma 16 in Schiffer and Janson (2024), for Algorithm 2,conditional on $E \cap \bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, W_s)$, for all s,

$$\begin{aligned}
&|T_s \cdot J(\theta^*, C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, T_s, x'_{T_s}, W_s) - T_s \cdot J(\theta^*, C^{\mathrm{alg}}_s, T_s, x'_{T_s}, W_s)| \\
&= \tilde{O}_T\left(\sum_{i=0}^{T_s-1} |C^{\mathrm{alg}}_s(x'_{T_s+i}) - C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}(x'_{T_s+i})| + T_s\epsilon_s\right).
\end{aligned}$$

Note that we replaced $K_{\mathrm{opt}}(\hat{\theta}_s, T_s)$ with $K_{\mathrm{opt}}(\hat{\theta}_s)$ for reasons discussed in the previous paragraph. Taking $E_{\mathrm{P26}} = \bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, W_s)$ gives the desired result because by a union round and Assumption 3, we have $\mathbb{P}\left(\bigcap_{s=0}^{s_e} E_{\mathrm{L3}}(C^{\hat{\theta}_s}_{K_{\mathrm{opt}}(\hat{\theta}_s)}, W_s)\right) = 1 - o_T(1/T)$.

F.6. Regret from Warm-Up (Proposition 27)

The proof of Proposition 27 follows exactly the same as the proof of Proposition 7 in Schiffer and Janson (2024). This is because the controller of Algorithm 2 is safe for dynamics θ^* under event E, and the result therefore follows directly.

Appendix G. Technical Tools for Large Noise Case

G.1. Necessary Condition for Large Noise Case (Lemma 34)

Proof The following lemma shows that $F_{\text{opt}}(\theta^*)$ and $F_{\text{opt}}(\hat{\theta}_{\text{wu}})$ are similar under event E.

Lemma 37 *Under Assumptions 1–3 and 4, conditional on event E_2^0, there exists $c_{\text{L37}} = \tilde{O}_T(1)$ such that for sufficiently large T,*

$$|F_{\text{opt}}(\theta^*) - F_{\text{opt}}(\hat{\theta}_{\text{wu}})| \leq c_{\text{L37}} T^{-1/4}.$$

The proof of Lemma 37 can be found in Appendix G.4.

Conditional on E (because $E \subseteq E_2^0$), we have that $\|\hat{\theta}_{\text{wu}} - \theta^*\|_\infty \leq 2\epsilon_0 \leq 2c_{\text{L18}} T^{-1/4}$. This combined with Lemma 37 implies that there exists $c_{\text{E81}} = \tilde{O}_T(1)$ such that under event E for sufficiently large T,

$$\hat{a} - \hat{b} F_{\text{opt}}(\hat{\theta}_{\text{wu}}) \leq a^* - b^* F_{\text{opt}}(\theta^*) + c_{\text{E81}} T^{-1/4}. \tag{81}$$

Now we will proceed with a proof by contradiction of Lemma 34. Assume event $\neg E_{\text{E56}} \cap E$ holds and $\epsilon^* \leq 0$, the latter of which implies

$$\frac{D_U}{a^* - b^* K_{\text{opt}}(\theta^*)} - D_U \geq \bar{w}, \tag{82}$$

which in turn implies that $K_{\text{opt}}(\theta^*) \geq K_{D_U}^{\theta^*}$ (recall $K_{D_U}^{\theta^*}$ was defined in Definition 16). A key result is the following relationship between $K_{\text{opt}}(\theta^*)$ and $F_{\text{opt}}(\theta^*)$.

Lemma 38 *Under Assumptions 1–3 and 4, for any $\theta \in \Theta$, if $K_{\text{opt}}(\theta) \geq K_{D_U}^{\theta}$, then $F_{\text{opt}}(\theta) \geq K_{D_U}^{\theta}$.*

The proof of Lemma 38 can be found in Appendix G.5.

We also will need the following result.

Lemma 39 *Under Assumptions 1–3 and 4, there exists $c_{\text{L39}} = O_T(1)$ such that $c_{\text{L39}} > 0$ and for all $\theta \in \Theta$,*

$$1 - c_{\text{L39}} > a - b F_{\text{opt}}(\theta) \geq c_{\text{L39}},$$
$$a - b K_{\text{opt}}(\theta) \geq c_{\text{L39}}.$$

The proof of Lemma 39 can be found in Appendix G.6.

Lemma 38 combined with Equation (82) give that $F_{\text{opt}}(\theta^*) \geq K_{D_U}^{\theta^*}$, or equivalently that $\bar{w} + D_U - \frac{D_U}{a^* - b^* F_{\text{opt}}(\theta^*)} \leq 0$. Therefore, we have that for sufficiently large T under event $\neg E_{\text{E56}} \cap E$,

$$\begin{aligned}
&\bar{w} + D_U - \frac{D_U}{\hat{a} - \hat{b} F_{\text{opt}}(\hat{\theta}_{\text{wu}})} \\
&\leq \bar{w} + D_U - \frac{D_U}{a^* - b^* F_{\text{opt}}(\theta^*) + c_{\text{E81}} T^{-1/4}} && \text{Equation (81)} \\
&= \frac{c_{\text{E81}} T^{-1/4} D_U}{(a^* - b^* F_{\text{opt}}(\theta^*))(a^* - b^* F_{\text{opt}}(\theta^*) + c_{\text{E81}} T^{-1/4})} + \bar{w} + D_U - \frac{D_U}{a^* - b^* F_{\text{opt}}(\theta^*)} \\
&\leq \frac{c_{\text{E81}} T^{-1/4} D_U}{(a^* - b^* F_{\text{opt}}(\theta^*))(a^* - b^* F_{\text{opt}}(\theta^*) + c_{\text{E81}} T^{-1/4})} && \text{Lemma 38, Eq (82)} \\
&\leq \left(\frac{c_{\text{E81}} D_U}{c_{\text{L39}}^2} \right) T^{-1/4} && \text{Lemma 39} \\
&= C_{\text{switch}} T^{-1/4}.
\end{aligned}$$

However, this contradicts event $\neg E_{\text{E56}}$ and therefore we have a contradiction. This implies the desired result that if $\neg E_{\text{E56}} \cap E$ holds, then $\epsilon^* > 0$.

■

G.2. Lower Bounding Frequency of Non-Linear Controls (Lemma 35)

Proof Define the event $E_2^s := \left\{\|\hat{\theta}_s^{\text{pre}} - \theta^*\|_\infty \le \epsilon_s = \tilde{O}_T(T^{-1/4})\right\}$. Define $G_i = (x_0, u_0, ..., x_{i-1}, u_{i-1})$ and define

$$S_t'' = \left\{i < t : u_i = u_i^{\text{safeU}} \text{ and } \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i) \ge \mathbb{P}_{w\sim\mathcal{D}}(w \ge \bar{w} - 3\epsilon^*/8)\right\}.$$

Lemma 40 *Under Assumptions 1–3 and 4 there exists a constant p_ϵ such that the following holds. For sufficiently large T and any $s \in [0 : s_e - 1]$ and any $T_s \le j < T_{s+1} - \lceil\log(T)\rceil$, there exists an event X_j that depends on $\{w_t\}_{t=j}^{j+\lceil\log(T)\rceil-1}$ such that $\mathbb{P}(X_j) \ge p_\epsilon$ and such that conditional on event $X_j \cap E_2^s \cap \neg E_{\text{E56}}$, there exists an $\ell \in [j : j + \lceil\log(T)\rceil)$ such that $\ell \in S''_{T_{s+1}}$.*

The proof of Lemma 40 can be found in Appendix G.7.

Define

$$\mathcal{E}^s := \left\{\sum_{\ell=0}^{\lfloor T_s/\lfloor\log(T)\rfloor\rfloor-1} 1_{X_{T_s+\ell\lfloor\log(T)\rfloor}} \ge p_\epsilon \left\lfloor\frac{T_s}{\lfloor\log(T)\rfloor}\right\rfloor - \sqrt{\left\lfloor\frac{T_s}{\lfloor\log(T)\rfloor}\right\rfloor}\log(T)\right\}.$$

Note that $\sum_{\ell=0}^{k}\left(1_{X_{T_s+\ell\lfloor\log(T)\rfloor}} - p_\epsilon\right)$ is a submartingale. Therefore, by the Azuma–Hoeffding inequality, we have that $\mathbb{P}(\mathcal{E}^s) = 1 - o_T(1/T^2)$. Define $\mathcal{E} = \cap_{s=0}^{s_e-1}\mathcal{E}^s$. Then by a union bound $\mathbb{P}(\mathcal{E}) = 1 - o_T(1/T)$.

Conditional on $\mathcal{E} \cap E \cap \neg E_{\text{E56}}$, we have that

$$\begin{aligned}
|S''_{T_{s+1}}| &\ge \sum_{\ell=0}^{\lfloor T_s/\lfloor\log(T)\rfloor\rfloor-1} 1_{X_{T_s+\ell\lfloor\log(T)\rfloor}} \\
&\ge p_\epsilon\left\lfloor\frac{T_s}{\lfloor\log(T)\rfloor}\right\rfloor - \sqrt{\left\lfloor\frac{T_s}{\lfloor\log(T)\rfloor}\right\rfloor}\log(T) && \text{Event } \mathcal{E} \\
&\ge \frac{p_\epsilon T_s}{2\log(T)} - \sqrt{\left\lfloor\frac{T_s}{\lfloor\log(T)\rfloor}\right\rfloor}\log(T) \\
&\ge \frac{p_\epsilon}{4\log(T)}\cdot T_s && \text{Suff. large } T \\
&= \frac{p_\epsilon}{8\log(T)}\cdot T_{s+1}. && T_{s+1} = 2T_s \qquad (83)
\end{aligned}$$

The following lemma is the same as Lemma 27 in Schiffer and Janson (2024). The proof is the same as the proof of that lemma, as the proof of Lemma 27 in Schiffer and Janson (2024) does not depend on the algorithm and only uses that $\mathbb{P}(E) = 1 - o_T(1/T^2)$.

Lemma 41 *Using the same notation and assumptions as in the proof of Lemma 35, for any constant $c < 1$,*

$$\mathbb{P}\Big(\forall i \in [0 : t-1], \mathbb{P}(E \mid G_i) \ge c\Big) = 1 - o_T(1/T).$$

Define $E_{\text{L41}} = \{\forall i \in [T_0 : T-1], \mathbb{P}(E \mid G_i) \ge 1 - \frac{\mathbb{P}_{w\sim\mathcal{D}}(w\ge\bar{w}-3\epsilon^*/8)}{2}\}$.

By Lemma 41, $\mathbb{P}(E_{\text{L41}}) = 1 - o_T(1/T)$. For any $i \in [T_0 : T-1]$, conditional on $E_{\text{L41}} \cap \{\mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i) \ge \mathbb{P}_{w\sim\mathcal{D}}(w \ge \bar{w} - 3\epsilon^*/8)\}$, by the law of total probability

$$\begin{aligned}
\mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i) &= \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, E)\mathbb{P}(E \mid G_i) + \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, \neg E)\mathbb{P}(\neg E \mid G_i) \\
&\le \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, E)\mathbb{P}(E \mid G_i) + \frac{\mathbb{P}_{w\sim\mathcal{D}}(w \ge \bar{w} - 3\epsilon^*/8)}{2}.
\end{aligned}$$

Rearranging terms gives

$$\begin{aligned}\mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, E) &\geq \frac{\mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i) - \frac{\mathbb{P}_{w\sim\mathcal{D}}(w \geq \bar{w} - 3\epsilon^*/8)}{2}}{\mathbb{P}(E \mid G_i)} \\ &\geq \mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i) - \frac{\mathbb{P}_{w\sim\mathcal{D}}(w \geq \bar{w} - 3\epsilon^*/8)}{2} \\ &\geq \frac{\mathbb{P}_{w\sim\mathcal{D}}(w \geq \bar{w} - 3\epsilon^*/8)}{2}.\end{aligned}$$

Therefore we have shown that conditional on $E_{\text{L41}} \cap \{\mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i) \geq \mathbb{P}_{w\sim\mathcal{D}}(w \geq \bar{w} - 3\epsilon^*/8)\}$, we also have $\mathbb{P}(u_i = u_i^{\text{safeU}} \mid G_i, E) \geq \frac{\mathbb{P}_{w\sim\mathcal{D}}(w\geq\bar{w}-3\epsilon^*/8)}{2}$. This implies that conditional on E_{L41}, for all $t \in [0:T]$,

$$S_t'' \subseteq S_t'.$$

Combining this with Equation (83), conditional on $E_{\text{L41}} \cap \mathcal{E} \cap E \cap \neg E_{\text{E56}}$,

$$\max_{s\in[1:s_e]} \frac{T_s}{|S'_{T_s}|} \leq \frac{8\log(T)}{p_\epsilon} = \tilde{O}_T(1).$$

We therefore take $E_{\text{L35}} = E_{\text{L41}} \cap \mathcal{E} \cap E$ to get the desired result because $\mathbb{P}(E_{\text{L35}}) = 1 - o_T(1/T)$ by a union bound. ■

G.3. Cost Difference of Optimal Finite vs Optimal Infinite Controller (Lemma 36)

Proof Let x_T be the state after starting at $x_0 = 0$ and using the controller C_K^θ for T steps under dynamics θ. Therefore, because C_K^θ is safe for dynamics θ, we must have that $|x_T| \leq \max(D_U, |D_L|) + \bar{w} \leq 2\log^2(T)$ for sufficiently large T. Therefore, there must exist an $L \leq 2\log^2(T)$ such that $\mathbb{P}(|x| \geq L)\,\mathbb{E}[x^2 \mid |x| \geq L] = o_T(1/T^{11})$. Define $W' = \{w_i\}_{i=0}^T$. We can apply Lemma 55 in the sixth line below to get that

$$\begin{aligned}&\left|\bar{J}(\theta, C_K^\theta, 2T) - \bar{J}(\theta, C_K^\theta, T)\right| \\ &= \left|\frac{T\cdot\bar{J}(\theta, C_K^\theta, T) + T\cdot\mathbb{E}\left[\bar{J}(\theta, C_K^\theta, T, x_T)\right]}{2T} - \bar{J}(\theta, C_K^\theta, T)\right| \\ &= \left|\frac{\mathbb{E}\left[\bar{J}(\theta, C_K^\theta, T, x_T)\right]}{2} - \frac{1}{2}\bar{J}(\theta, C_K^\theta, T)\right| \\ &= \frac{1}{2T}\left|\mathbb{E}\left[T\bar{J}(\theta, C_K^\theta, T, x_T)\right] - T\bar{J}(\theta, C_K^\theta, T)\right| \\ &= \frac{1}{2T}\left|\mathbb{E}\left[TJ(\theta, C_K^\theta, T, x_T, W') - TJ(\theta, C_K^\theta, T, 0, W')\right]\right| \\ &\leq \frac{1}{T}\tilde{O}_T\left(\mathbb{E}[|x_T|] + 0 + \frac{1}{T^2}\right) && \text{Lemma 55} \\ &\leq \tilde{O}_T\left(\frac{1}{T}\right). && |x_T| \leq \|D\|_\infty + \bar{w} = \tilde{O}_T(1)\end{aligned}$$

The last line follows from the fact that C_K^θ is safe for dynamics θ. Finally, we have that

$$\begin{aligned}
|\bar{J}(\theta, C_K^\theta, T) - \bar{J}(\theta, C_K^\theta)| &= \left| \sum_{i=0}^{\infty} \bar{J}(\theta, C_K^\theta, 2^i T) - \bar{J}(\theta, C_K^\theta, 2^{i+1} T) \right| \\
&\leq \sum_{i=0}^{\infty} \left| \bar{J}(\theta, C_K^\theta, 2^i T) - \bar{J}(\theta, C_K^\theta, 2^{i+1} T) \right| \\
&= \sum_{i=0}^{\infty} \tilde{O}_T\left(\frac{1}{T 2^i}\right) \\
&= \tilde{O}_T\left(\frac{1}{T}\right).
\end{aligned}$$

■

G.4. Estimating Optimal Linear Controller (Lemma 37)

Proof By Lemma 42, the optimal unconstrained controller for dynamics θ is $C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\theta)}$, where

$$F_{\mathrm{opt}}(\theta) = \arg\min_F T \cdot \bar{J}(\theta, C_F^{\mathrm{unc}}) = \arg\min_F \cdot \frac{q + rF^2}{1 - (a - bF)^2}. \tag{84}$$

We show in the proof of Lemma 39 that

$$F_{\mathrm{opt}}(\theta) = \frac{a^2 r - b^2 q - r + \sqrt{(b^2 q + r - a^2 r)^2 + 4a^2 b^2 q r}}{2abr}.$$

Note that this is a differentiable function in both a and b for $\theta \in \Theta$. Under event E_2^0, $\|\theta^* - \hat{\theta}_{\mathrm{wu}}\|_\infty = \tilde{O}_T(T^{-1/4})$ where $\hat{\theta}_{\mathrm{wu}}$ is the estimate from Line 3 of Algorithm 2. Therefore, a first order Taylor expansion of $F_{\mathrm{opt}}(\theta)$ around $\theta = \theta^*$ gives that for sufficiently large T, $|F_{\mathrm{opt}}(\theta^*) - F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})| = O_T(\|\theta^* - \hat{\theta}_{\mathrm{wu}}\|_\infty) = \tilde{O}_T(T^{-1/4}) = c_{\mathrm{L37}} T^{-1/4}$ for some $c_{\mathrm{L37}} = \tilde{O}_T(1)$.

■

G.5. Relating Optimal Linear to Optimal Truncated Linear (Lemma 38)

Proof We will prove the contrapositive, which is that if $F_{\mathrm{opt}}(\theta) < K^\theta_{D_U}$, then $K_{\mathrm{opt}}(\theta) < K^\theta_{D_U}$.

The first tool we need is the following result about $F_{\mathrm{opt}}(\theta)$.

Lemma 42 *For any $\theta \in \Theta$ and $K \in (\frac{a-1}{b}, \frac{a}{b}]$,*

$$\bar{J}(\theta, C_K^{\mathrm{unc}}) = \lim_{T \to \infty} \bar{J}(\theta, C_K^{\mathrm{unc}}, T) = \frac{\sigma_{\mathcal{D}}^2 (q + rK^2)}{1 - (a - bK)^2}.$$

This function is convex and twice differentiable for $K \in (\frac{a-1}{b}, \frac{a}{b}]$. Furthermore, if $1 - (a - bK) > 0$, then $\left|\frac{d}{dK} \bar{J}(\theta, C_K^{\mathrm{unc}})\right|$ and $\left|\frac{d^2}{dK^2} \bar{J}(\theta, C_K^{\mathrm{unc}})\right|$ are finite and $\frac{d^2}{dK^2} \bar{J}(\theta, C_K^{\mathrm{unc}}) > 0$.

Finally, if $K = \frac{a-1}{b}$, then $\bar{J}(\theta, C_K^{\mathrm{unc}}) = \infty$.

The proof of Lemma 42 can be found in Appendix G.8.

Lemma 42 implies that the function $\bar{J}(\theta, C_K^{\text{unc}})$ has a unique local minimum ($F_{\text{opt}}(\theta)$) and is convex. Therefore, if $F_{\text{opt}}(\theta) < K^\theta_{D_U}$, then for any $K' > K^\theta_{D_U}$,

$$\bar{J}(\theta, C^{\text{unc}}_{K^\theta_{D_U}}) \leq \bar{J}(\theta, C^{\text{unc}}_{K'}). \tag{85}$$

For any $K' \geq K^\theta_{D_U}$, the unconstrained and constrained controllers are the same, i.e. $C^{\text{unc}}_{K'} = C^\theta_{K'}$. This is because for $K' \geq K^\theta_{D_U}$ the unconstrained controller will always satisfy the state constraints because we assumed WLOG that $D_U \leq |D_L|$. This implies by Equation (85) that for any $K' > K^\theta_{D_U}$,

$$\bar{J}(\theta, C^\theta_{K^\theta_{D_U}}) \leq \bar{J}(\theta, C^\theta_{K'}).$$

Therefore, to prove that $K_{\text{opt}}(\theta) < K^\theta_{D_U}$ it is sufficient to find some $K' < K^\theta_{D_U}$ such that

$$\bar{J}(\theta, C^\theta_{K^\theta_{D_U}}) > \bar{J}(\theta, C^\theta_{K'}). \tag{86}$$

Let $K' = K^\theta_{D_U} - \epsilon$, where

$$0 < \epsilon \leq \min\left(\frac{4B_P}{(\bar{w} + D_U)^2}, \frac{\min(\bar{w}, D_U)/2}{(\bar{w} + D_U)}\right). \tag{87}$$

We will show that $\bar{J}(\theta, C^\theta_{K'}) < \bar{J}(\theta, C^\theta_{K^\theta_{D_U}})$ which proves the desired contrapositive result.

Because $a - bK^\theta_{D_U} = \frac{D_U}{D_U + \bar{w}} = 1 - \frac{\bar{w}}{D_U + \bar{w}}$, by Lemma 42 the function $\bar{J}(\theta, C_K^{\text{unc}})$ has a finite derivative at $K = K^\theta_{D_U}$. Furthermore, if $F_{\text{opt}}(\theta) < K^\theta_{D_U}$, then Lemma 42 implies that the derivative of $\bar{J}(\theta, C_F^{\text{unc}})$ is positive at $K = K^\theta_{D_U}$. Therefore, we can take a first order Taylor expansion around the point $K = K^\theta_{D_U}$ to get that for sufficiently small ϵ,

$$\bar{J}(\theta, C^{\text{unc}}_{K'}) - \bar{J}(\theta, C^{\text{unc}}_{K^\theta_{D_U}}) \leq -\Omega_T(\epsilon). \tag{88}$$

Because $C^{\text{unc}}_{K^\theta_{D_U}} = C^\theta_{K^\theta_{D_U}}$, Equation (88) implies that

$$\bar{J}(\theta, C^{\text{unc}}_{K'}) - \bar{J}(\theta, C^\theta_{K^\theta_{D_U}}) \leq -\Omega_T(\epsilon). \tag{89}$$

Note that in Equations (88) and (89), the LHS is not a function of T. We use the notation $-\Omega_T(\epsilon)$ to indicate that the LHS is upper bounded by $-c\epsilon$ for some constant c.

Now we will compare the cost of $C^{\text{unc}}_{K'}$ and $C^\theta_{K'}$ using the following lemma. Note that this lemma is stated very generally so that it can also be used in future results.

Lemma 43 *For $\theta, \hat{\theta}_{\text{L43}} \in \Theta$, suppose $\beta \leq \frac{1}{\log^2(T)}$ satisfies that $\theta \in \hat{\theta}_{\text{L43}} \pm \beta$. Also, suppose K' satisfies $K^\theta_{D_U} - K' \leq \epsilon$ for some $\epsilon > 0$. Furthermore, suppose*

$$\upsilon := (b\epsilon + \beta + |K'|\beta) \leq \min\left(\frac{4B_P}{(\bar{w} + D_U)^2}, \frac{\min(\bar{w}, D_U)/2}{(\bar{w} + D_U)}\right) \tag{90}$$

Define the controller C as follows. For any t, define v_t^{safeU} as the largest u such that for all $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$,

$$a' x_t + b' u \leq D_U,$$

and define v_t^{safeL} *as the smallest* u *such that for all* $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$,

$$D_L \le a' x_t + b' u.$$

Define the controller C *as*

$$C(x_t) = \max\left(\min\left(C_{K'}^{\text{unc}}(x_t), v_t^{\text{safeU}}\right), v_t^{\text{safeL}}\right).$$

Let $|x_0| \le \|D\|_\infty + \bar{w}$. *Then under Assumptions 1–3 and 4,*

$$|\bar{J}(\theta, C, x_0) - \bar{J}(\theta, C_{K'}^{\text{unc}}, x_0)| \le O_T(\upsilon^2). \tag{91}$$

Furthermore, with probability $1 - o_T(1/T^2)$, *for any* $\tau \le T$,

$$|J(\theta, C, \tau, x_0, W') - J(\theta, C_{K'}^{\text{unc}}, \tau, x_0, W')| \le O_T\left(\upsilon \log(1/\upsilon)\left(\upsilon + \frac{\log(T)}{\sqrt{\tau}}\right)\right). \tag{92}$$

The proof of Lemma 43 can be found in Appendix G.9.

We will use Lemma 43 with the ϵ defined in Equation (87), $K' = K_{D_U}^\theta - \epsilon$, $\theta = \theta$, $\hat{\theta}_{\text{L43}} = \theta$, $x_0 = 0$, and $\beta = 0$. Choosing $\hat{\theta}_{\text{L43}} = \theta$ and $\beta = 0$ makes the C in Lemma 43 equivalent to a truncated linear controller. Then, Equation (91) of Lemma 43 gives that

$$|\bar{J}(\theta, C_{K'}^\theta) - \bar{J}(\theta, C_{K'}^{\text{unc}})| \le O_T(\epsilon^2). \tag{93}$$

Putting together Equations (89) and (93), for small enough ϵ we have that

$$\begin{aligned}
&\bar{J}(\theta, C_{K'}^\theta) - \bar{J}(\theta, C_{K_{D_U}^\theta}^\theta) \\
&= \bar{J}(\theta, C_{K'}^\theta) - \bar{J}(\theta, C_{K'}^{\text{unc}}) + \bar{J}(\theta, C_{K'}^{\text{unc}}) - \bar{J}(\theta, C_{K_{D_U}^\theta}^\theta) \\
&\le O_T\left(\epsilon^2\right) - \Omega_T\left(\epsilon\right). && \text{Equations (89), (93)} \\
&< 0. && \text{For small enough } \epsilon
\end{aligned}$$

We have shown that $C_{K'}^\theta$ has lower cost than $C_{K_{D_U}^\theta}^\theta$, and therefore we can conclude that $K_{\text{opt}}(\theta) < K_{D_U}^\theta$, proving the contrapositive and our desired result. ■

G.6. Bounding Linear Scaling away from 0 and 1 (Lemma 39)

Proof By Lemma 42, $F_{\text{opt}}(\theta)$ is the value of $K \in \left(\frac{a-1}{b}, \frac{a}{b}\right]$ that minimizes the function $\frac{q+rK^2}{1-(a-bK)^2}$ (note that we ignore the constant $\sigma_{\mathcal{D}}^2$ as this is a positive constant and does not change the minimization problem). Taking the derivative of this function and equating to 0, we have that $F_{\text{opt}}(\theta)$ is the solution to

$$\frac{2Kr(1-(a-bK)^2) - 2b(a-bK)(q+rK^2)}{(1-(a-bK)^2)^2} = 0.$$

Simplifying, we have

$$abrK^2 + (b^2q + r - a^2r)K - abq = 0$$

Applying the quadratic formula, we get that the positive root is

$$F_{\text{opt}}(\theta) = \frac{a^2r - b^2q - r + \sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}}{2abr}.$$

We also observe that

$$\left(a^2r + b^2q + r\right)^2 - \left(\sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)^2 = 4a^2r^2,$$

which implies that

$$\begin{aligned}&\left(a^2r + b^2q + r\right) - \left(\sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)\\ &= \frac{4a^2r^2}{\left(a^2r + b^2q + r\right) + \left(\sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)}.\end{aligned}$$

Because $\underline{a} \geq a \geq \bar{a}$, $\underline{b} \geq b \geq \bar{b}$, and $r > 0$, this implies that there exists a constant $c^{F1}_{\text{L39}} > 0$ such that

$$\begin{aligned}\frac{a}{b} - F_{\text{opt}}(\theta) &= \frac{a^2r + b^2q + r - \sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}}{2abr}\\ &= \frac{4a^2r^2}{2abr\left(a^2r + b^2q + r + \sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)}\\ &\geq \frac{4\underline{a}^2r^2}{2\bar{a}\bar{b}r\left(\bar{a}^2r + \bar{b}^2q + r + \sqrt{(\bar{b}^2q + r - \underline{a}^2r)^2 + 4\bar{a}^2\bar{b}^2qr}\right)}\\ &:= c^{F1}_{\text{L39}}\\ &> 0.\end{aligned}$$

Similarly, we have that

$$\left(r(a-1)^2 + b^2q\right)^2 - \left(\sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)^2 = -4ar\left((a-1)^2r + b^2q\right).$$

which implies that

$$\begin{aligned}&\left(r(a-1)^2 + b^2q\right) - \left(\sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)\\ &= \frac{-4ar\left((a-1)^2r + b^2q\right)}{\left(r(a-1)^2 + b^2q\right) + \left(\sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)}.\end{aligned}$$

Because $a \geq \bar{a}$ and $r > 0$, this implies that there exists a constant $c^{F2}_{\text{L39}} > 0$ such that

$$\begin{aligned}\frac{a-1}{b} - F_{\text{opt}}(\theta) &= \frac{r(a-1)^2 + b^2q - \sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}}{2abr}\\ &= \frac{-4ar\left((a-1)^2r + b^2q\right)}{2abr\left(r(a-1)^2 + b^2q + \sqrt{(b^2q + r - a^2r)^2 + 4a^2b^2qr}\right)}\\ &\leq -c^{F2}_{\text{L39}}\\ &< 0,\end{aligned}$$

where the constant C^{F2}_{L39} depends on $\bar{a}, \underline{a}, \bar{b}, \underline{b}$. Taking $c^{F}_{\text{L39}} = \min(c^{F1}_{\text{L39}}, c^{F2}_{\text{L39}})$, we have that

$$c^{F}_{\text{L39}} < a - bF_{\text{opt}}(\theta) < 1 - c^{F}_{\text{L39}}. \tag{94}$$

To bound $K_{\text{opt}}(\theta, T)$ away from 0 we need the following lemma:

Lemma 44 *Under Assumptions 1–3, for any $\theta \in \Theta$, if $F_{\text{opt}}(\theta) \geq K^{\theta}_{D_U}$, then $K_{\text{opt}}(\theta) = F_{\text{opt}}(\theta)$.*

Proof If $F_{\text{opt}}(\theta) \geq K^{\theta}_{D_U}$, then $C^{\text{unc}}_{F_{\text{opt}}(\theta)} = C^{\theta}_{F_{\text{opt}}(\theta)}$, i.e. the unconstrained linear controller for $F_{\text{opt}}(\theta)$ is the same as the constrained linear controller for $F_{\text{opt}}(\theta)$. Therefore, $C^{\text{unc}}_{F_{\text{opt}}(\theta)}$ is in the set of constrained controllers. Because the optimal unconstrained controller is linear Anderson and Moore (2007), $C^{\text{unc}}_{F_{\text{opt}}(\theta)}$ is the lowest cost unconstrained controller, and therefore it is also the lowest cost constrained controller. ■

By Lemma 44 and the contrapositive of Lemma 38, either $K_{\text{opt}}(\theta) = F_{\text{opt}}(\theta)$ or $K_{\text{opt}}(\theta) < K^{\theta}_{D_U}$. By Equation (94) and the fact that $a - bK^{\theta}_{D_U} = \frac{D_U}{D_U + \bar{w}}$, we can conclude that

$$a - bK_{\text{opt}}(\theta) \geq \min\left(\frac{D_U}{D_U + \bar{w}}, c^F_{\text{L39}}\right) > 0.$$

Therefore, taking $c_{\text{L39}} = \min\left(\frac{D_U}{D_U+\bar{w}}, c^F_{\text{L39}}\right)$ we have the desired result. ■

G.7. Bounding Conditional Frequency of Non-Linear Controls (Lemma 40)

Proof The structure of this proof is as follows. The bulk of the proof is split into two key lemmas. We then combine these two lemmas to show the desired result. Define

$$\tau := \left\lceil 8\left(\frac{2 + c_{\text{L39}}}{c_{\text{L39}}}\|D\|_\infty + 2\bar{w}\right)/\epsilon^* \right\rceil,$$

where ϵ^* is from Lemma 34. Now, we will define

$$X_j := \{\forall t \in [j : j + \tau], w_t \geq \bar{w} - \epsilon^*/4\}.$$

Note that $\mathbb{P}(X_j) = (\mathbb{P}_{w\sim\mathcal{D}}(w \geq \bar{w} - \epsilon^*/4))^{\tau+1} := p_\epsilon$, and for sufficiently large T, $\tau \leq \lceil \log(T) \rceil$, therefore this X_j has the desired properties.

Lemma 45 *Using the assumptions and notation of Lemma 40, conditional on $E^s_2 \cap \neg E_{\text{E56}} \cap X_j$, there exists an $\ell \in [j : j + \tau]$ such that $u_\ell = u^{\text{safeU}}_\ell$.*

Proof We will first show that conditional on $E^s_2 \cap \neg E_{\text{E56}}$, for any value of x satisfying $D_L - \bar{w} \leq x \leq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}$, and for sufficiently large T, if $w \geq \bar{w} - \epsilon^*/4$, then

$$(a^* - b^* K_{\text{opt}}(\hat{\theta}_s))x + w \geq x + \frac{\epsilon^*}{8}. \tag{95}$$

Under event E^s_2, $\|\theta^* - \hat{\theta}_s\|_\infty \leq \tilde{O}_T(T^{-1/4})$, therefore under event E^s_2 we have the following results:

$$\begin{aligned} a^* - b^* K_{\text{opt}}(\hat{\theta}_s) &\geq \hat{a}_s - \hat{b}_s K_{\text{opt}}(\hat{\theta}_s) - \tilde{O}_T(T^{-1/4}) && \|\theta^* - \hat{\theta}_s\|_\infty \leq \tilde{O}_T(T^{-1/4}) \\ &\geq c_{\text{L39}} - \tilde{O}_T(T^{-1/4}) && \text{Lemma 39} \\ &\geq \frac{c_{\text{L39}}}{2} && \text{suff large } T \end{aligned} \tag{96}$$

and

$$\begin{aligned} a^* - b^* K_{\text{opt}}(\hat{\theta}_s) &\leq \hat{a}_s - \hat{b}_s K_{\text{opt}}(\hat{\theta}_s) + \tilde{O}_T(T^{-1/4}) && \|\theta^* - \hat{\theta}_s\|_\infty \leq \tilde{O}_T(T^{-1/4}) \\ &\leq 1 + \tilde{O}_T(T^{-1/4}). && \text{Lemma 39} \end{aligned} \tag{97}$$

Equation (96) implies that for sufficiently large T,

$$\frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \leq \frac{2D_U}{c_{\text{L39}}} = O_T(1). \tag{98}$$

To prove Equation (95), we will need the following result.

Lemma 46 *Under Assumptions 1–3 and 4, conditional on event $E_2^s \cap \neg E_{\text{E56}}$ and for sufficiently large T,*

$$\frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \leq D_U + \bar{w} - \epsilon^*/2.$$

The proof of Lemma 46 can be found in Appendix G.10.

Conditional on event $E_2^s \cap \neg E_{\text{E56}}$, for sufficiently large T, and for any $D_L - \bar{w} \leq x \leq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}$,

$$\begin{aligned}
&(a^* - b^* K_{\text{opt}}(\hat{\theta}_s))x + \bar{w} - \epsilon^*/4 \\
&= D_U + \left(x - \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}\right)(a^* - b^* K_{\text{opt}}(\hat{\theta}_s)) + \bar{w} - \epsilon^*/2 + \epsilon^*/4 \\
&\geq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} + \left(x - \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}\right)(a^* - b^* K_{\text{opt}}(\hat{\theta}_s)) + \epsilon^*/4 && \text{Lemma 46} \\
&\geq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} + \left(x - \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}\right)(1 + \tilde{O}_T(T^{-1/4})) + \epsilon^*/4 && \text{Eq (97), } x \leq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \\
&= x + \tilde{O}_T\left(T^{-1/4}\left(x - \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}\right)\right) + \epsilon^*/4 \\
&\geq x - \tilde{O}_T\left(T^{-1/4}\left(|D_L| + \bar{w} + \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}\right)\right) + \epsilon^*/4 && D_L - \bar{w} \leq x \leq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \\
&\geq x - \tilde{O}_T\left(T^{-1/4}\right) + \epsilon^*/4 && \text{Eq (98), Assumption 4} \\
&\geq x + \epsilon^*/8. && \text{For sufficiently large } T.
\end{aligned}$$

This in turn implies the statement containing Equation (95).

Recall that u_i is the control at time i of Algorithm 2 and x'_i is the state of Algorithm 2 at time i. Under event $\neg E_{\text{E56}}$, for any $i \in [T_s + 1 : T_{s+1}]$, if $u_{i-1} \neq u_{i-1}^{\text{safeU}}$, then the control at time $i-1$ is either $u_{i-1} = -K_{\text{opt}}(\hat{\theta}_s)x'_{i-1}$ or $u_{i-1} = u_{i-1}^{\text{safeL}} \geq -K_{\text{opt}}(\hat{\theta}_s)x'_{i-1}$. Therefore, under event $\neg E_{\text{E56}}$, if $u_{i-1} \neq u_{i-1}^{\text{safeU}}$ then

$$u_{i-1} \geq -K_{\text{opt}}(\hat{\theta}_s)x'_{i-1}. \tag{99}$$

Combining Equations (95) and (99) gives that for any $i \in [T_s + 1 : 2T_s]$, conditional on the event $\{u_{i-1} \neq u_{i-1}^{\text{safeU}}\} \cap \left\{D_L - \bar{w} \leq x'_{i-1} \leq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)}\right\} \cap E_2^s \cap \neg E_{\text{E56}} \cap X_j$,

$$\begin{aligned}
x'_i &= a^* x'_{i-1} + b^* u_{i-1} + w_{i-1} \\
&\geq a^* x'_{i-1} - b^* K_{\text{opt}}(\hat{\theta}_s)x'_{i-1} + w_{i-1} && \text{Equation (99)} \\
&= (a^* - b^* K_{\text{opt}}(\hat{\theta}_s))x'_{i-1} + w_{i-1} \\
&\geq x'_{i-1} + \frac{\epsilon^*}{8}. && \text{Equation (95)}
\end{aligned} \tag{100}$$

If the control at time $j-1$ is safe (which is guaranteed by construction of the algorithm under event E_2^s), then $x'_j \geq D_L - \bar{w}$. Therefore by Equation (98),

$$\frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} - x'_j \leq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} + |D_L| + \bar{w} \leq \frac{2D_U}{c_{\mathrm{L39}}} + |D_L| + \bar{w} \leq \frac{2 + c_{\mathrm{L39}}}{c_{\mathrm{L39}}} \|D\|_\infty + \bar{w} = O_T(1). \tag{101}$$

By Equation (100), conditional on $E_2^s \cap \neg E_{\mathrm{E56}} \cap X_j$ the state will increase by $\epsilon^*/8$ at each step ℓ if $D_L - \bar{w} \leq x_\ell \leq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)}$ and $u_\ell \neq u_\ell^{\mathrm{safeU}}$. Furthermore, by Equation (101), if the state increases by at least $\frac{2+c_{\mathrm{L39}}}{c_{\mathrm{L39}}} \|D\|_\infty + 2\bar{w}$ from x'_j, then the state will be greater than $\frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)}$. Increasing $\frac{2+c_{\mathrm{L39}}}{c_{\mathrm{L39}}} \|D\|_\infty + 2\bar{w}$ state in increments of at least $\epsilon^*/8$ takes at most $\left\lceil \frac{8(\frac{2+c_{\mathrm{L39}}}{c_{\mathrm{L39}}} \|D\|_\infty + 2\bar{w})}{\epsilon^*} \right\rceil = \tau$ steps. Putting this all together, conditional on $E_2^s \cap \neg E_{\mathrm{E56}} \cap X_j$, either $u_\ell = u_\ell^{\mathrm{safeU}}$ for some $\ell \in [j : j+\tau]$ or $x'_\ell \geq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)}$ for some $\ell \in [j : j+\tau]$. Both of these alternatives imply that $u_\ell = u_\ell^{\mathrm{safeU}}$ for some $\ell \in [j : j+\tau]$, because if $x'_\ell \geq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)}$, then by construction of the algorithm, $u_\ell = u_\ell^{\mathrm{safeU}}$. This is the desired result for this lemma. ■

The next key result is the following lemma.

Lemma 47 *Using the notation and assumptions of the proof of Lemma 40, for sufficiently large T and any $\ell \in [j : j+\tau]$, conditional on $\{u_\ell = u_\ell^{\mathrm{safeU}}\} \cap E_2^s \cap \neg E_{\mathrm{E56}} \cap X_j$, $\ell + 1 \in S''_{T_{s+1}}$.*

Proof Suppose $\ell \in [j : j+\tau]$. Under event E_2^s the control at step $\ell - 1$ is safe, and therefore by the same logic as in Equation (62) in Schiffer and Janson (2024), for sufficiently large T we have that

$$D_U - \epsilon^*/8 \leq D_U - \tilde{O}_T(T^{-1/4}) \leq D_U - 4B_x\epsilon_s \leq a^* x'_\ell + b^* u_\ell^{\mathrm{safeU}}. \tag{102}$$

Therefore, if $u_\ell = u_\ell^{\mathrm{safeU}}$, then

$$a^* x'_\ell + b^* u_\ell \geq D_U - \epsilon^*/8. \tag{103}$$

Therefore, conditional on $\{u_\ell = u_\ell^{\mathrm{safeU}}\} \cap E_2^s \cap \neg E_{\mathrm{E56}} \cap X_j$,

$$\begin{aligned}
x'_{\ell+1} &= a^* x'_\ell + b^* u_\ell + w_\ell \\
&\geq D_U - \epsilon^*/8 + w_\ell && \text{Equation (103)} \\
&\geq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} + 3\epsilon^*/8 + w_\ell - \bar{w} && \text{Lemma 46} \\
&= \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} + w_\ell - (\bar{w} - 3\epsilon^*/8) \\
&\geq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} && \text{Event } X_j
\end{aligned} \tag{104}$$

We also recall again that if $x'_{\ell+1} \geq \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)}$, then $u_{\ell+1} = u_{\ell+1}^{\mathrm{safeU}}$. Therefore, we have shown that conditional on $\{u_\ell = u_\ell^{\mathrm{safeU}}\} \cap E_2^s \cap \neg E_{\mathrm{E56}} \cap X_j$, $u_{\ell+1} = u_{\ell+1}^{\mathrm{safeU}}$. Furthermore, we have for any $G_{\ell+1}$ that

satisfies $\{u_\ell = u_\ell^{\text{safeU}}\} \cap E_2^s \cap \neg E_{\text{E56}}$,

$$\begin{aligned}
&\mathbb{P}\left(u_{\ell+1} = u_{\ell+1}^{\text{safeU}} \,\middle|\, G_{\ell+1}\right) \\
&\geq \mathbb{P}\left(x'_{\ell+1} \geq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \,\middle|\, G_{\ell+1}\right) \\
&= \mathbb{P}\left(a^* x'_\ell + b^* u_\ell + w_\ell \geq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \,\middle|\, G_{\ell+1}\right) \\
&\geq \mathbb{P}\left(D_U - \epsilon^*/8 + w_\ell \geq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} \,\middle|\, G_{\ell+1}\right) && \text{Equation (103)} \\
&= \mathbb{P}\left(w_\ell \geq \frac{D_U}{a^* - b^* K_{\text{opt}}(\hat{\theta}_s)} - D_U + \epsilon/8 \,\middle|\, G_{\ell+1}\right) \\
&\geq \mathbb{P}\left(w_\ell \geq \bar{w} - \epsilon^*/2 + \epsilon^*/8 \mid G_{\ell+1}\right) && \text{Lemma 46} \\
&= \mathbb{P}_{w\sim\mathcal{D}}(w \geq \bar{w} - 3\epsilon^*/8). && (105)
\end{aligned}$$

By Definition of S_t'', Equations (104) and (105) imply the desired result that conditional on $\{u_\ell = u_\ell^{\text{safeU}}\} \cap E_2^s \cap \neg E_{\text{E56}} \cap X_j$, we have that $\ell + 1 \in S''_{T_{s+1}}$. ■

Putting together the two lemmas, we have that conditional on $E_2^s \cap \neg E_{\text{E56}} \cap X_j$, there exists an $\ell \in [j : j + \tau]$ such that $u_\ell = u_\ell^{\text{safeU}}$, and for any $\ell \in [j : j + \tau]$, conditional on $\{u_\ell = u_\ell^{\text{safeU}}\} \cap E_2^s \cap \neg E_{\text{E56}} \cap X_j$, $\ell + 1 \in S''_{T_{s+1}}$. Combining these two lemmas gives that conditional on $E_2^s \cap \neg E_{\text{E56}} \cap X_j$, there exists an $\ell \in [j : j + \tau + 1]$ such that $\ell \in S''_{T_{s+1}}$. For sufficiently large T, $\tau + 1 \leq \lceil \log(T) \rceil$, and therefore this is exactly the desired result. ■

G.8. Total Cost of Linear Controllers (Lemma 42)

Proof Let $x_0, x_1, ...,$ be the series of states when using controller C_K^{unc} under dynamics θ with $x_0 = 0$. Then we have the recursive relationship that $x_0 = 0$ and $x_{i+1} = (a - bK)x_i + w_i$ for all $i \geq 0$. Using this recursive relationship, we have that

$$x_t = \sum_{i=0}^{t-1} w_i (a - bK)^{t-1-i}. \tag{106}$$

If $K = \frac{a-1}{b}$, then $a - bK = 1$. This implies that $x_t^2 \longrightarrow \infty$, and therefore $\bar{J}(\theta, C_K^{\text{unc}}, T) = \infty$.

For the rest of this proof, assume $K \in (\frac{a-1}{b}, \frac{a}{b}]$. Recall that $u_i = -Kx_i$ for all $i \geq 0$. Define $\rho = (a - bK)^2$. Using the above expression for x_t, we have that

$$\begin{aligned}
\bar{J}(\theta, C_K^{\text{unc}}, T) &= \frac{1}{T}\,\mathbb{E}\left[q x_T^2 + \sum_{t=0}^{T-1} q x_t^2 + r u_t^2\right] \\
&= \frac{1}{T}\left(q\,\mathbb{E}[x_T^2] + \sum_{t=1}^{T-1} (q + rK^2)\,\mathbb{E}[x_t^2]\right) && [x_0 = u_0 = 0] \\
&= -\frac{rK^2\,\mathbb{E}[X_T^2]}{T} + \frac{1}{T}\left(\sum_{t=1}^{T} (q + rK^2)\,\mathbb{E}[x_t^2]\right).
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&\frac{1}{T}\left(\sum_{t=1}^{T}(q+rK^2)\,\mathbb{E}[x_t^2]\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}(q+rK^2)\,\mathbb{E}\Big[\Big(\sum_{i=0}^{t-1}w_i(a-bK)^{t-1-i}\Big)^2\Big] && \text{Equation (106)}\\
&=\frac{1}{T}\sum_{t=1}^{T}(q+rK^2)\,\mathbb{E}\Big[\sum_{i=0}^{t-1}\sum_{j=0}^{t-1}w_iw_j(a-bK)^{t-1-i}(a-bK)^{t-1-j}\Big]\\
&=\frac{1}{T}\sum_{t=1}^{T}(q+rK^2)\sum_{i=0}^{t-1}\sigma_{\mathcal{D}}^2(a-bK)^{2(t-1-i)}\\
&=\frac{\sigma_{\mathcal{D}}^2}{T}\sum_{t=1}^{T}(q+rK^2)\sum_{i=0}^{t-1}(a-bK)^{2i}\\
&=\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{T}\sum_{t=1}^{T}\sum_{i=0}^{t-1}\rho^i\\
&=\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{T}\sum_{t=1}^{T}\frac{1-\rho^t}{1-\rho}\\
&=\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{T(1-\rho)}\left(T-\sum_{t=0}^{T-1}\rho^t\right)\\
&=\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{1-\rho}\Big(1-\frac{1-\rho^T}{T(1-\rho)}\Big).
\end{aligned}$$

By the same logic, we have that

$$\frac{rK^2\,\mathbb{E}[X_T^2]}{T}=\frac{rK^2\sigma_{\mathcal{D}}^2\frac{1-\rho^T}{1-\rho}}{T}.$$

Therefore,

$$\begin{aligned}
\bar{J}(\theta,C_K^{\mathrm{unc}})&=\lim_{T\to\infty}\bar{J}(\theta,C_K^{\mathrm{unc}},T)\\
&=\lim_{T\to\infty}-\frac{rK^2\sigma_{\mathcal{D}}^2\frac{1-\rho^T}{1-\rho}}{T}+\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{1-\rho}\Big(1-\frac{1-\rho^T}{T(1-\rho)}\Big)\\
&=\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{1-(a-bK)^2}.
\end{aligned}$$

Now, we note the following derivatives:

$$\frac{d}{dK}\left(\frac{1}{1-(a-bK)^2}\right)=\frac{2b(a-bK)}{(1-(a-bK)^2)^2}$$

and

$$\frac{d}{dK}\left(\frac{K^2}{1-(a-bK)^2}\right)=\frac{2aK(1-(a-bK))}{(1-(a-bK)^2)^2}.$$

For $K \in (\frac{a-1}{b}, \frac{a}{b}]$, if $1-(a-bK) = c > 0$, then $1-(a-bK)^2 > c > 0$, and therefore these derivatives imply that

$$\begin{aligned}\left|\frac{d}{dK}\bar{J}(\theta, C_K^{\text{unc}})\right| &= \left|\frac{d}{dK}\frac{\sigma_{\mathcal{D}}^2(q+rK^2)}{1-(a-bK)^2}\right| \\ &= \left|\sigma_{\mathcal{D}}^2\left(q\frac{2b(a-bK)}{(1-(a-bK)^2)^2} + r\frac{2aK(1-(a-bK))}{(1-(a-bK)^2)^2}\right)\right| \\ &\leq \sigma_{\mathcal{D}}^2\left(q\frac{2b(a-bK)}{c^2} + r\frac{2a|K|(1-(a-bK))}{c^2}\right) \\ &< \infty.\end{aligned}$$

For all $K \in (\frac{a-1}{b}, \frac{a}{b}]$, we also have that

$$\frac{d^2}{dK^2}\left(\frac{1}{1-(a-bK)^2}\right) = b^2\Big(\frac{1}{(1-(a-bK))^3} + \frac{1}{(1+(a-bK))^3}\Big) > 0$$

and

$$\frac{d^2}{dK^2}\left(\frac{K^2}{1-(a-bK)^2}\right) = b^2\Big(\frac{(a-1)^2}{(1-(a-bK))^3} + \frac{(a+1)^2}{(1+(a-bK))^3}\Big) > 0$$

This implies that

$$\frac{d^2}{dK^2}\bar{J}(\theta, C_K^{\text{unc}}) > 0.$$

If $a - bK = 1 - c < 1$, we also have that

$$\frac{d^2}{dK^2}\left(\frac{1}{1-(a-bK)^2}\right) = b^2\Big(\frac{1}{(1-(a-bK))^3} + \frac{1}{(1+(a-bK))^3}\Big) \leq b^2\left(\frac{1}{c^3}+1\right) < \infty$$

and

$$\frac{d^2}{dK^2}\left(\frac{K^2}{1-(a-bK)^2}\right) = b^2\Big(\frac{(a-1)^2}{(1-(a-bK))^3} + \frac{(a+1)^2}{(1+(a-bK))^3}\Big) \leq b^2\left(\frac{(a-1)^2}{c^3} + (a+1)^2\right) < \infty.$$

These two equations imply that for $K \in (\frac{a-1}{b}, \frac{a}{b}]$,

$$\frac{d^2}{dK^2}\bar{J}(\theta, C_K^{\text{unc}}) < \infty.$$

■

G.9. Comparing Cost of Truncated Controller to Cost of Linear (Lemma 43)

Proof We first note the following bounds on K' that we will use throughout this proof that come from the assumptions on ϵ. For any $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$,

$$\begin{aligned}a' - b'K' &= a - bK_{D_U}^{\theta} + (a'-a) + b(K_{D_U}^{\theta} - K') + K'(b-b') \\ &\leq a - bK_{D_U}^{\theta} + b\epsilon + \beta + \beta K' \\ &\leq \frac{D_U}{\bar{w}+D_U} + \upsilon && \text{Def of } K_{D_U}^{\theta} \\ &\leq \frac{D_U + \bar{w}/2}{\bar{w}+D_U} && \text{Equation (90)} \\ &< 1. && (107)\end{aligned}$$

$$\begin{aligned} a' - b'K' &\geq a - bK^{\theta}_{D_U} - b\epsilon - \beta - \beta K' \\ &\geq \frac{D_U}{\bar{w} + D_U} - \upsilon && \text{Def of } K^{\theta}_{D_U} \\ &\geq \frac{D_U/2}{\bar{w} + D_U} && \text{Equation (90)} \\ &> 0. \end{aligned} \tag{108}$$

Let y_t be the state at time t when using controller C and starting at state $y_0 = x_0$ and x_t be the state at time t when using controller $C^{\text{unc}}_{K'}$ and starting at state x_0. Define $d_t := |y_t - x_t|$. Define

$$\theta_m := \arg \max_{\|\theta' - \hat{\theta}_{\text{L43}}\|_\infty \leq \beta} a' - b'K'. \tag{109}$$

Importantly, note that $\theta_m = \arg\min_{\theta' \in \hat{\theta}_{\text{L43}} \pm \beta} \frac{D_U}{a' - b'K'} = \arg\max_{\theta' \in \hat{\theta}_{\text{L43}} \pm \beta} \frac{D_L}{a' - b'K'}$. By construction this means that $C(y_t) = v_t^{\text{safeU}}$ is used if and only if $y_t \geq \frac{D_{\text{U}}}{a_m - b_m K'}$, and similarly v_t^{safeL} is used if and only if $y_t \leq \frac{D_{\text{L}}}{a_m - b_m K'}$.

Lemma 48 *Define $H_t = (y_0, y_1, ..., y_{t-1})$. Using the notation and assumptions in the proof of Lemma 43, for any H_t,*

$$\mathbb{P}\left(C(y_t) = v_t^{\text{safeU}} \;\middle|\; H_t\right) = O_T(\upsilon) \cdot 1_{K' - K^{\theta}_{D_U} \leq \frac{(|K'|+1)\beta}{b}}. \tag{110}$$

Furthermore,

$$\mathbb{P}\left(C(y_t) = v_t^{\text{safeL}} \;\middle|\; H_t\right) = O_T(\upsilon) \cdot 1_{K' - K^{\theta}_{D_U} \leq \frac{(|K'|+1)\beta}{b}}. \tag{111}$$

The proof of Lemma 48 can be found in Appendix G.11. Because the equations in Lemma 48 hold for any H_t, this lemma implies that

$$\mathbb{P}\left(C(y_t) = v_t^{\text{safeU}}\right) = O_T(\upsilon) \cdot 1_{K' - K^{\theta}_{D_U} \leq \frac{(|K'|+1)\beta}{b}} \tag{112}$$

and

$$\mathbb{P}\left(C(y_t) = v_t^{\text{safeL}}\right) = O_T(\upsilon) \cdot 1_{K' - K^{\theta}_{D_U} \leq \frac{(|K'|+1)\beta}{b}}. \tag{113}$$

By Lemma 48, if $K' - K^{\theta}_{D_U} > \frac{(|K'|+1)\beta}{b}$, then for all t,

$$\mathbb{P}\left(C(y_t) = v_t^{\text{safeU}} \text{ or } C(y_t) = v_t^{\text{safeL}}\right) = 0.$$

Therefore in this case, the controllers C and $C^{\text{unc}}_{K'}$ are equivalent, which implies all of the desired results. For the rest of the proof, we will address the case when $K' - K^{\theta}_{D_U} \leq \frac{(|K'|+1)\beta}{b}$. This combined with the definition of ϵ gives that

$$|K' - K^{\theta}_{D_U}| \leq \min\left(\frac{(|K'|+1)\beta}{b}, \epsilon\right) = O_T(\upsilon). \tag{114}$$

Lemma 49 *Using the notation and assumptions in the proof of Lemma 43, if Equation (114) holds then for all $t \geq 0$,*

$$d_{t+1} = \begin{cases} (a - bK')d_t & \text{if } \frac{D_L}{a_m - b_m K'} \leq y_t \leq \frac{D_U}{a_m - b_m K'} \\ (a - bK')d_t + O_T(\upsilon) & \text{otherwise,} \end{cases} \tag{115}$$

and

$$|C^{\text{unc}}_{K'}(x_t) - C(y_t)| = \begin{cases} |K'|d_t & \text{if } \frac{D_L}{a_m - b_m K'} \leq y_t \leq \frac{D_U}{a_m - b_m K'} \\ O_T(\upsilon) & \text{otherwise.} \end{cases} \tag{116}$$

The proof of Lemma 49 can be found in Appendix G.12.

This recursive relationship for d_t in Lemma 49 implies that

$$\begin{aligned}
d_t &= |x_t - y_t| \\
&\leq \sum_{i=1}^{t} (a - bK')^{i-1} O_T(\upsilon) 1_{y_{t-i} \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_{t-i} \leq \frac{D_L}{a_m - b_m K'}} && \text{Lemma 49} \\
&\leq O_T(\upsilon) \sum_{i=0}^{\infty} (a - bK')^i \\
&\leq \frac{O_T(\upsilon)}{1 - (a - bK')} \\
&\leq O_T(\upsilon). && \text{Equation (107)} \qquad (117)
\end{aligned}$$

Note that y_t is by construction safe with respect to dynamics θ_m. Therefore, $|a_m y_t + b_m C(y_t)| \leq \|D\|_\infty$ and $|y_t| \leq \|D\|_\infty + \bar{w}$, which together imply that

$$|C(y_t)| \leq \frac{\|D\|_\infty + a_m |y_t|}{b_m} = O_T(1). \qquad (118)$$

Now we can bound the difference in cost at time $t \geq 0$ as follows:

$$\begin{aligned}
&|qx_t^2 - qy_t^2| + |rC_{K'}^{\text{unc}}(x_t)^2 - rC(y_t)^2| \\
&\leq 2q|y_t|d_t + qd_t^2 + \left(2r|C(y_t)|\,|C_{K'}^{\text{unc}}(x_t) - C(y_t)| + r\,|C_{K'}^{\text{unc}}(x_t) - C(y_t)|^2\right) \\
&\leq 2q|y_t|d_t + qd_t^2 + \left(2rO_T(1)|\,|C_{K'}^{\text{unc}}(x_t) - C(y_t)| + r\,|C_{K'}^{\text{unc}}(x_t) - C(y_t)|^2\right) && \text{Equation (118)} \\
&\leq 2q(\|D\|_\infty + \bar{w})d_t + qd_t^2 + \left(2rO_T(1)\left(|K'|d_t + O_T(\upsilon)1_{y_t \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_t \leq \frac{D_L}{a_m - b_m K'}}\right)\right. \\
&\qquad \left. + r\left(|K'|d_t + O_T(\upsilon)1_{y_t \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_t \leq \frac{D_L}{a_m - b_m K'}}\right)^2\right) && \text{Equation (116)} \\
&= O_T\left(d_t + \upsilon^2 + \upsilon 1_{y_t \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_t \leq \frac{D_L}{a_m - b_m K'}}\right). && \text{Equation (117)} \\
& && (119)
\end{aligned}$$

We will now show that $\mathbb{E}[d_t] \leq O_T(\upsilon^2)$. Importantly, we use that the event

$$\{y_{i-1} \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_{i-1} \leq \frac{D_L}{a_m - b_m K'}\}$$

is equivalent to the event that $C(y_{i-1}) \in \{v_t^{\text{safeU}}, v_t^{\text{safeL}}\}$, which allows us to apply Lemma 48 in the second line.

$$\begin{aligned}
\mathbb{E}[d_t] &\le O_T(\upsilon) \sum_{i=1}^{t} (a - bK')^{t-i}\, \mathbb{E}[1_{y_{i-1} \ge \frac{D_U}{a_m - b_m K'} \text{ or } y_{i-1} \le \frac{D_L}{a_m - b_m K'}}] && \text{Lemma 49} \\
&\le O_T(\upsilon) \sum_{i=1}^{t} (a - bK')^{t-i} O_T(\upsilon) && \text{Lemma 48} \\
&\le O_t(\upsilon^2) \sum_{i=0}^{\infty} (a - bK')^{t-i} \\
&\le \frac{O_T(\upsilon^2)}{1 - (a - bK')} \\
&\le O_T(\upsilon^2). && \text{Equation (107)} \qquad (120)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&|\bar{J}(\theta, C, \tau, x_0) - \bar{J}(\theta, C_{K'}^{\text{unc}}, \tau, x_0)| \\
&\le \mathbb{E}\left[\frac{1}{\tau}\left(q|x_\tau^2 - y_\tau^2| + \sum_{t=0}^{\tau-1} |qx_t^2 - qy_t^2| + |rC_{K'}^{\text{unc}}(x_t)^2 - rC(y_t)^2|\right)\right] \\
&\le \mathbb{E}\left[\frac{1}{\tau}\sum_{t=0}^{\tau} O_T\left(d_t + \upsilon^2 + \upsilon 1_{y_t \ge \frac{D_U}{a_m - b_m K'} \text{ or } y_t \le \frac{D_L}{a_m - b_m K'}}\right)\right] && \text{Equation (119)} \\
&\le \frac{1}{\tau}\sum_{t=0}^{\tau} O_T\left(\mathbb{E}[d_t] + \upsilon^2 + \upsilon\, \mathbb{E}\left[1_{y_t \ge \frac{D_U}{a_m - b_m K'} \text{ or } y_t \le \frac{D_L}{a_m - b_m K'}}\right]\right) \\
&\le \frac{1}{\tau}\sum_{t=0}^{\tau} O_T(\upsilon^2) && \text{Equation (120), Lemma 48} \\
&\le O_T(\upsilon^2).
\end{aligned}$$

Taking a limit as $\tau \to \infty$ of the above equation (where nothing on the right side depends on τ) gives the first desired equation that

$$|\bar{J}(\theta, C, x_0) - \bar{J}(\theta, C_{K'}^{\text{unc}}, x_0)| \le O_T(\upsilon^2).$$

Now we want to bound the difference in cost with high probability instead of in expectation. Let X be the set of times $t \in [0 : \tau]$ such that $C(y_t) \ne -K' y_t$ (i.e. $C(y_t) \in \{v_t^{\text{safeL}}, v_t^{\text{safeU}}\}$). Note that the event $\{t \in X\}$ is the same as the event $\{y_t \ge \frac{D_U}{a_m - b_m K'} \text{ or } y_t \le \frac{D_L}{a_m - b_m K'}\}$.

By Lemma 48, $\mathbb{P}(t \in X \mid H_t) \le c\upsilon$ for some constant $c > 0$ for all t. Therefore, $M_k = \sum_{t=0}^{\tau} (1_{t \in X} - c\upsilon)$ is a supermartingale. By Azuma–Hoeffding's inequality, with probability $1 - o_T(1/T^{10})$,

$$|X| \le O_T(\upsilon\tau) + \log(T)\sqrt{\tau}.$$

Define A as the event that $|X| \le O_T(\upsilon\tau) + \log(T)\sqrt{\tau}$. Define $\kappa = \lceil \log_{a - bK'}(\upsilon) \rceil$. Note that

$$\begin{aligned}
\kappa &= \lceil \log_{a - bK'}(\upsilon) \rceil \\
&\le \left\lceil \frac{\log(\upsilon)}{\log(a - bK')} \right\rceil \\
&= O(\log(\upsilon)) && \text{Lemma 107} \qquad (121)
\end{aligned}$$

Define

$$G = \left\{ t \in [0:\tau] : \exists i \in [t-\kappa : t] \text{ such that } C(y_i) \neq -K' y_i \right\}.$$

Under event A,

$$|G| \leq |X| \cdot (\kappa+1) \leq (O_T(\upsilon\tau) + \log(T)\sqrt{\tau})(\kappa+1). \tag{122}$$

By Lemma 49, if $t \notin G$, then

$$\begin{aligned}
d_t &\leq O_T(\upsilon) \sum_{i=1}^{t} (a-bK')^{t-i} 1_{y_{i-1} \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_{i-1} \leq \frac{D_L}{a_m - b_m K'}} && \text{Lemma 49} \\
&\leq O_T(\upsilon) \sum_{i=1}^{t-\kappa} (a-bK')^{t-i} 1_{y_{i-1} \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_{i-1} \leq \frac{D_L}{a_m - b_m K'}} && t \notin G \\
&\leq O_T(\upsilon)(a-bK')^{\kappa} \sum_{i=1}^{t-\kappa} (a-bK')^{t-i-\kappa} 1_{y_{i-1} \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_{i-1} \leq \frac{D_L}{a_m - b_m K'}} \\
&\leq O_T(\upsilon)(a-bK')^{\kappa} \sum_{i=0}^{\infty} (a-bK')^{i} \\
&\leq O_T(\upsilon^2) \sum_{i=1}^{\infty} (a-bK')^{i} && \text{Definition of } \kappa \\
&= \frac{O_T(\upsilon^2)}{1-(a-bK')} \\
&= O_T(\upsilon^2). && \text{Equation (107)}
\end{aligned} \tag{123}$$

Recall that by Equation (117), for any $t \in G$, $d_t \leq O_T(\upsilon)$, therefore Equation (123) implies that

$$d_t = O_T\left(\upsilon 1_{t \in G} + \upsilon^2\right). \tag{124}$$

Using that $t \in G$ for all t satisfying $y_t \geq \frac{D_U}{a_m - b_m K'}$ or $y_t \leq \frac{D_L}{a_m - b_m K'}$, we have that under event A,

$$\begin{aligned}
&|J(\theta, C, \tau, x_0, W') - J(\theta, C^{\text{unc}}_{K'}, \tau, x_0, W')| \\
&\leq \frac{1}{\tau} \sum_{t=0}^{\tau} |q x_t^2 - q y_t^2| + |r C^{\text{unc}}_{K'}(x_t)^2 - r C(y_t)^2| \\
&= \frac{1}{\tau} \sum_{t=0}^{\tau} O_T\left(d_t + \upsilon^2 + \upsilon 1_{y_t \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_t \leq \frac{D_L}{a_m - b_m K'}} \right) && \text{Equation (119)} \\
&= \frac{1}{\tau} \sum_{t=0}^{\tau} O_T\left(\upsilon \cdot 1_{t \in G} + \upsilon^2 + \upsilon 1_{y_t \geq \frac{D_U}{a_m - b_m K'} \text{ or } y_t \leq \frac{D_L}{a_m - b_m K'}} \right) && \text{Equation (124)} \\
&= O_T(\upsilon^2) + \frac{1}{\tau} \sum_{t=0}^{\tau} O_T\left(\upsilon\right) \cdot 1_{t \in G} \\
&= O_T(\upsilon^2) + O_T\left(\frac{\upsilon \cdot (O_T(\upsilon\tau) + \log(T)\sqrt{\tau})(\kappa+1)}{\tau} \right) && \text{Equation (122)} \\
&= O_T\left(\upsilon \log(1/\upsilon) \left(\upsilon + \frac{\log(T)}{\sqrt{\tau}} \right) \right). && \text{Equation (121)}
\end{aligned}$$

Since this holds under event A and $\mathbb{P}(A) \geq 1 - o_T(1/T^{10})$, this completes the proof. ■

G.10. Estimated Controller Ratio Bound under Large Noise (Lemma 46)

Proof In Algorithm 2, $\hat{\theta}_s$ satisfies

$$\hat{\theta}_s = \operatorname*{arg\,max}_{\|\theta - \hat{\theta}_s^{\mathrm{pre}}\| \le \epsilon_s} a - bK_{\mathrm{opt}}(\theta).$$

Under event E_2^s, we have that $\|\hat{\theta}_s^{\mathrm{pre}} - \theta^*\| \le \epsilon_s$, which implies that

$$\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s) \ge a^* - b^* K_{\mathrm{opt}}(\theta^*).$$

Therefore, we have that (using Lemma 34 in the equality)

$$\frac{D_U}{\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s)} - D_U \le \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\theta^*)} - D_U = \bar{w} - \epsilon^*. \tag{125}$$

Under event E_2^s, we also have that $\|\hat{\theta}_s - \theta^*\|_\infty \le \tilde{O}_T(T^{-1/4})$, therefore

$$\begin{aligned}
&\frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} - D_U \\
&= \frac{D_U}{\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s)} - D_U + \frac{D_U}{a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s)} - \frac{D_U}{\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s)} \\
&\le \bar{w} - \epsilon^* + D_U \frac{(\hat{a}_s - a^*) + (b^* - \hat{b}_s) K_{\mathrm{opt}}(\hat{\theta}_s)}{(a^* - b^* K_{\mathrm{opt}}(\hat{\theta}_s))(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s))} && \text{Eq (125)} \\
&= \bar{w} - \epsilon^* + D_U \frac{(\hat{a}_s - a^*) + (b^* - \hat{b}_s) K_{\mathrm{opt}}(\hat{\theta}_s)}{\left(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s) - (\hat{a}_s - a^*) - (b^* - \hat{b}_s) K_{\mathrm{opt}}(\hat{\theta}_s)\right)(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s))} \\
&\le \bar{w} - \epsilon^* + D_U \frac{\|\theta^* - \hat{\theta}_s\|_\infty \left(1 + |K_{\mathrm{opt}}(\hat{\theta}_s)|\right)}{(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s) - \|\theta^* - \hat{\theta}_s\|_\infty (1 + |K_{\mathrm{opt}}(\hat{\theta}_s)|))(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s))} \\
&\le \bar{w} - \epsilon^* + \frac{D_U \tilde{O}_T(T^{-1/4}) \left(1 + |K_{\mathrm{opt}}(\hat{\theta}_s)|\right)}{(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s) - \tilde{O}_T(T^{-1/4})(1 + |K_{\mathrm{opt}}(\hat{\theta}_s)|))(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s))} \\
&\le \bar{w} - \epsilon^*/2. && \text{Eq (127)}
\end{aligned} \tag{126}$$

To see the last inequality, note that Lemma 39 gives that $1 > \hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s) \ge c_{\mathrm{L39}}$. This implies that $|K_{\mathrm{opt}}(\hat{\theta}_s)| = O_T(1)$, and therefore for sufficiently large T we have that

$$\begin{aligned}
&\frac{D_U \tilde{O}_T(T^{-1/4}) \left(1 + K_{\mathrm{opt}}(\hat{\theta}_s)\right)}{(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s) - \tilde{O}_T(T^{-1/4})(1 + K_{\mathrm{opt}}(\hat{\theta}_s)))(\hat{a}_s - \hat{b}_s K_{\mathrm{opt}}(\hat{\theta}_s))} \\
&\le \frac{\tilde{O}_T(T^{-1/4}) D_U \left(1 + O_T(1)\right)}{(c_{\mathrm{L39}} - \tilde{O}_T(T^{-1/4})(1 + O_T(1)))c_{\mathrm{L39}}} \\
&\le \epsilon^*/2.
\end{aligned} \tag{127}$$

Finally, rearranging Equation (126) gives exactly the desired result. ■

G.11. Conditional Probability of Using u_t^{safeU} (Lemma 48)

Lemma 50 *Using the same notation and assumptions of Lemma 48, for all $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$, the controls used by controller C are safe for dynamics θ' for all $t \in [0 : T-1]$.*

The proof of Lemma 50 can be found in Appendix G.13

By definition, $C(y_t) = v_t^{\text{safeU}}$ if and only if there exists a $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$ such that $y_t \geq \frac{D_U}{a'-b'K'}$. Equivalently, $C(y_t) = v_t^{\text{safeU}}$ if and only if $y_t \geq \frac{D_U}{a_m - b_m K'}$. We also note that

$$\begin{aligned}
(a - bK^\theta_{D_U}) - (a_m - b_m K') &= (a - a_m) + b(K' - K^\theta_{D_U}) + K'(b_m - b) \\
&\geq -\beta + b(K' - K^\theta_{D_U}) - |K'|\beta \\
&= b(K' - K^\theta_{D_U}) - (1 + |K'|)\beta \\
&\geq \Big(b(K' - K^\theta_{D_U}) - (|K'| + 1)\beta\Big) 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} \\
&\geq -(b\epsilon + (|K'| + 1)\beta) 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} \\
&= -\upsilon 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}}. \qquad (128)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\left(C(y_t) = v_t^{\text{safeU}} \mid H_t\right) \\
&= \mathbb{P}\left(y_t \geq \frac{D_U}{a_m - b_m K'} \mid H_t\right) \\
&= \mathbb{P}\left(a y_{t-1} + bC(y_{t-1}) + w_{t-1} \geq \frac{D_U}{a_m - b_m K'} \mid H_t\right) \\
&\leq \mathbb{P}\left(|w_{t-1}| \geq \frac{D_U}{a_m - b_m K'} - D_U \mid H_t\right) && \text{Lemma 50} \\
&= \mathbb{P}\left(|w_{t-1}| \geq \frac{D_U}{a_m - b_m K'} + \bar{w} - \frac{D_U}{a - bK^\theta_{D_U}}\right) && \text{Definition of } K^\theta_{D_U} \\
&= \mathbb{P}\left(|w_{t-1}| \geq \bar{w} + \frac{D_U(a - bK^\theta_{D_U}) - D_U(a_m - b_m K')}{(a - bK^\theta_{D_U})(a_m - b_m K')}\right) \\
&\leq \mathbb{P}\left(|w_{t-1}| \geq \bar{w} - \frac{D_U \upsilon}{(a - bK^\theta_{D_U})(a_m - b_m K')}\right) 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} && \text{Equation (128)} \\
&\leq \mathbb{P}\left(|w_{t-1}| \geq \bar{w} - \frac{D_U \upsilon}{(a - bK^\theta_{D_U})(a - bK^\theta_{D_U} - \upsilon)}\right) 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} && \text{Equation (128)} \\
&\leq \mathbb{P}\left(|w_{t-1}| \geq \bar{w} - \frac{D_U \upsilon}{(\frac{D_U}{\bar{w}+D_U})(\frac{D_U}{\bar{w}+D_U} - \frac{D_U}{2(\bar{w}+D_U)})}\right) 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} && \text{Def 16, } \upsilon \leq \frac{(D_U/2)}{(\bar{w}+D_U)} \\
&= \mathbb{P}\left(|w_{t-1}| \geq \bar{w} - \frac{2\upsilon(\bar{w} + D_U)^2}{D_U}\right) 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} \\
&\leq \frac{4B_P \upsilon(\bar{w} + D_U)^2}{D_U} 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\beta}{b}} && \mathcal{D} \text{ pdf bounded by } B_P \\
&= O_T(\upsilon) \cdot 1_{K' - K^\theta_{D_U} \leq \frac{(|K'|+1)\epsilon}{b}}. && (129)
\end{aligned}$$

Therefore, the safety truncation v_t^{safeU} is only applied with probability at most $O_T(\upsilon)$ at every time step. By definition, $C(y_t) = v_t^{\text{safeL}}$ if and only if there exists a $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$ such that $y_t \le \frac{D_L}{a'-b'K'}$. This only happens if and only if $y_t \le \frac{D_L}{a_m - b_m K'}$. We also have by Equations (107) and (108) that because $D_L < 0$,

$$\left| \frac{D_L}{a_m - b_m K'} - D_L \right| = \frac{|D_L|}{a_m - b_m K'} - |D_L|. \tag{130}$$

Also by Equations (107) and (108), we have because $D_U \le |D_L|$ that

$$\frac{|D_L|}{a_m - b_m K'} - |D_L| \ge \frac{D_U}{a_m - b_m K'} - D_U. \tag{131}$$

Therefore,

$$\begin{aligned}
&\mathbb{P}\left(C(y_t) = v_t^{\text{safeL}} \,\middle|\, H_t \right) \\
&= \mathbb{P}\left(y_t \le \frac{D_L}{a_m - b_m K'} \,\middle|\, H_t \right) \\
&= \mathbb{P}\left(a y_{t-1} + b C(y_{t-1}) + w_{t-1} \le \frac{D_L}{a_m - b_m K'} \,\middle|\, H_t \right) \\
&\le \mathbb{P}\left(w_{t-1} \le \frac{D_L}{a_m - b_m K'} - D_L \,\middle|\, H_t \right) && \text{Lemma 50} \\
&\le \mathbb{P}\left(|w_{t-1}| \ge \frac{|D_L|}{a_m - b_m K'} - |D_L| \,\middle|\, H_t \right) && \text{Equation (130)} \\
&\le \mathbb{P}\left(|w_{t-1}| \ge \frac{D_U}{a_m - b_m K'} - D_U \,\middle|\, H_t \right) && \text{Equation (131)} \\
&\le O_T(\upsilon) \cdot 1_{K' - K^{\theta}_{D_U} \le \frac{(|K'|+1)\epsilon}{b}} \cdot . && \text{Equation (129)}
\end{aligned} \tag{132}$$

This is exactly the second result we need and therefore we are done.

G.12. Bounding Difference in Position when Truncating (Lemma 49)

If $\frac{D_L}{a_m - b_m K'} \le y_t \le \frac{D_U}{a_m - b_m K'}$, then $C(y_t) = -K' y_t$, and therefore

$$|C(y_t) - C^{\text{unc}}_{K'}(x_t)| = |K'| d_t \tag{133}$$

and

$$d_{t+1} = |a y_t + b C(y_t) + w_t - (a x_t + b C^{\text{unc}}_{K'}(x_t) + w_t)| = (a - bK') d_t. \tag{134}$$

This proves the first case of both equations in Lemma 49. Now we will prove the second case of both equations.

Under Equation (114), we have that for any $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$

$$\begin{aligned}
\left| (a - bK^{\theta}_{D_U}) - (a' - b'K') \right| &\le |a - a'| + b|K' - K^{\theta}_{D_U}| + |K'||b' - b| \\
&\le (\beta + b O_T(\upsilon) + |K'|\beta) \\
&= O_T(\upsilon).
\end{aligned} \tag{135}$$

If $y_t > \frac{D_U}{a_m - b_m K'}$, then for some $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$, $C(y_t) = \frac{D_U - a' y_t}{b'}$. Therefore,

$$
\begin{aligned}
&|C(y_t) - C^{\text{unc}}_{K'}(y_t)| \\
&= \left|C(y_t) + K' y_t\right| \\
&= \left|\frac{D_U - a' y_t}{b'} + K' y_t\right| \\
&= \frac{1}{b'}\left|D_U - \left(a' - b' K'\right) y_t\right| \\
&= \frac{a' - b' K'}{b'}\left|\frac{D_U}{a' - b' K'} - y_t\right| && \text{Equations (107), (108)} \\
&\leq \frac{a' - b' K'}{b'}\left|\frac{D_U}{a' - b' K'} - (D_U + \bar{w})\right| && \frac{D_U}{a' - b' K'} \leq y_t \leq D_U + \bar{w} \text{ by Lemma 50} \\
&= \frac{a' - b' K'}{b'}\left|\frac{D_U}{a' - b' K'} - \frac{D_U}{a - b K^{\theta}_{D_U}}\right| \\
&= \frac{D_U}{b'}\left|\frac{(a - b K^{\theta}_{D_U}) - (a' - b' K')}{a - b K^{\theta}_{D_U}}\right| \\
&\leq \frac{D_U}{b'}\left(\frac{O_T(\upsilon)}{a - b K^{\theta}_{D_U}}\right) && \text{Equation (135), Equation (108)} \\
&= \frac{(D_U + \bar{w}) O_T(\upsilon)}{b'} && a - b K^{\theta}_{D_U} = \frac{D_U}{D_U + \bar{w}} \\
&= O_T(\upsilon). && (136)
\end{aligned}
$$

Because the controls used by C are safe with respect to θ by Lemma 50, if $D_L - \frac{D_L}{a_m - b_m K'} > \bar{w}$, then $\mathbb{P}\left(y_t \leq \frac{D_L}{a_m - b_m K'}\right) = 0$. Therefore, if $y_t \leq \frac{D_L}{a_m - b_m K'}$ then it also must be the case that $D_L - \frac{D_L}{a_m - b_m K'} \leq \bar{w}$. By Equations (107) and (108), we have that $a_m - b_m K' \leq \frac{D_U}{D_U + \bar{w}} + O_T(\upsilon)$ and $a_m - b_m K' \geq \frac{D_U}{D_U + \bar{w}} - O_T(\upsilon)$. Therefore, if $y_t \leq \frac{D_L}{a_m - b_m K'}$, then $D_L - \frac{D_L}{a_m - b_m K'} \leq \bar{w}$, which implies that

$$
\begin{aligned}
D_L &\geq \frac{\bar{w}}{1 - \frac{1}{a_m - b_m K'}} \\
&= \bar{w}\frac{a_m - b_m K'}{a_m - b_m K' - 1} \\
&\geq \bar{w}\frac{\frac{D_U}{D_U + \bar{w}} - O_T(\upsilon)}{\frac{D_U}{D_U + \bar{w}} + O_T(\upsilon) - 1} \\
&= \bar{w}\left(\frac{\frac{D_U}{D_U + \bar{w}}}{\frac{-\bar{w}}{D_U + \bar{w}}} - O_T(\upsilon)\right) \\
&= -D_U - O_T(\upsilon).
\end{aligned}
$$

This combined with the fact that $D_U \leq |D_L|$ by Assumption 4, we have that if $y_t \leq \frac{D_L}{a_m - b_m K'}$, then

$$
||D_L| - D_U| \leq O_T(\upsilon). \tag{137}
$$

Therefore, if $y_t < \frac{D_L}{a_m - b_m K'}$, then for some $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$, $C(y_t) = \frac{D_L - a' y_t}{b'}$. Therefore,

$$
\begin{aligned}
&|C(y_t) - C^{\text{unc}}_{K'}(y_t)| \\
&= |C(y_t) + K' y_t| \\
&= \left| \frac{D_L - a' y_t}{b'} + K' y_t \right| \\
&= \frac{1}{b'} |D_L - (a' - b'K')\, y_t| \\
&\le \frac{1}{b'} |D_L - (a' - b'K')\, (D_L - \bar{w})| && D_L - \bar{w} \le y_t \le \frac{D_L}{a' - b'K'}, \text{Eq (108)} \\
&= \frac{1}{b'} ||D_L| - (a' - b'K')\, (|D_L| + \bar{w})| \\
&\le \frac{1}{b'} |D_U - (a' - b'K')\, (D_U + \bar{w})| + |D_U - |D_L|| \\
&\quad + |(a' - b'K')(D_U - |D_L|)| \\
&\le \frac{1}{b'} |D_U - (a' - b'K')\, (D_U + \bar{w})| + |D_U - |D_L|| + |D_U - |D_L|| && \text{Equation (108), (107)} \\
&\le \frac{1}{b'} |D_U - (a' - b'K')\, (D_U + \bar{w})| + O_T(\upsilon) && \text{Equation (137)} \\
&\le \frac{(a' - b'K')}{b'} \left| \frac{D_U}{(a' - b'K')} - (D_U + \bar{w}) \right| + O_T(\upsilon) \\
&= O_T(\upsilon). && \text{As in Equation (136)} \qquad (138)
\end{aligned}
$$

Combining Equations (136) and (138) gives that if $y_t > \frac{D_U}{a_m - b_m K'}$ or $y_t < \frac{D_L}{a_m - b_m K'}$,

$$
\begin{aligned}
|C^{\text{unc}}_{K'}(x_t) - C(y_t)| &= \left| -K' x_t + K' y_t - K' y_t - C(y_t) \right| \\
&= \left| K' x_t - K' y_t \right| + \left| K' y_t + C(y_t) \right| \\
&\le K' d_t + O_T(\upsilon) \\
&\le O_T(\upsilon). && \text{Equation (117)} \qquad (139)
\end{aligned}
$$

Now we can use this to bound the value of d_{t+1} as follows:

$$
\begin{aligned}
d_{t+1} &= |(a - bK')x_t - (a y_t - b C(y_{t_1})| \\
&= |(a - bK')x_t - (a - bK')y_t + bK' y_t - bC(y_t)| \\
&\le |(a - bK')x_t - (a - bK')y_t| + |bK' y_t - bC(y_t)| \\
&\le (a - bK')d_t + b O_T(\upsilon) && \text{Equations (136) and (138)} \\
&\le (a - bK')d_t + O_T(\upsilon). && (140)
\end{aligned}
$$

Equations (139) and (140) give the second half of both desired piecewise equations.

G.13. Safety of Truncated Controller (Lemma 50)

Proof We will proceed by induction. For the base case, we have that $y_0 = x_0$ satisfies $|y_0| \le \|D\|_\infty + \bar{w}$. Define $z := \frac{D_U - a y_0 - 2\beta(\|D\|_\infty + \bar{w} + \log(T))}{b}$. For sufficiently large T, because $\beta \le 1/\log^2(T)$ and $\|D\|_\infty = O_T(1)$, we have that

$$
|z| \le \frac{D_U + a(\|D\|_\infty + \bar{w}) + 2\beta(\|D\|_\infty + \bar{w} + \log(T))}{b} \le \frac{D_U + a(\|D\|_\infty + \bar{w}) + \frac{2(\|D\|_\infty + \bar{w} + \log(T))}{\log^2(T)}}{b} \le \log(T).
$$

Because $\theta \in \hat{\theta}_{\text{L43}} \pm \beta$,

$$\begin{aligned}\max_{\theta' \in \hat{\theta}_{\text{L43}} \pm \beta} a'y_0 + b'z &\leq ay_0 + bz + 2\beta|y_0| + 2\beta|z| \\ &\leq ay_0 + bz + 2\beta(\|D\|_\infty + \bar{w} + \log(T)) \\ &= ay_0 + D_U - ay_0 - 2\beta(\|D\|_\infty + \bar{w} + \log(T)) + 2\beta(\|D\|_\infty + \bar{w} + \log(T)) \\ &= D_U.\end{aligned}$$

Therefore,

$$v_t^{\text{safeU}} \geq z = \frac{D_U - ay_0 - 2\beta(D_U + \bar{w} + \log(T))}{b}.$$

By similar logic, we have that

$$v_t^{\text{safeL}} \leq \frac{D_L - ay_0 + 2\beta(\|D\|_\infty + \bar{w} + \log(T))}{b}.$$

For sufficiently large T, $4\beta(\|D\|_\infty + \bar{w} + \log(T)) \leq \frac{1}{\log(T)}$. Therefore, because $D_U \geq D_L + \frac{1}{\log(T)}$, we have that

$$v_t^{\text{safeL}} \leq v_t^{\text{safeU}}.$$

Finally, this implies by construction of the controller C that the control $C(y_0)$ will be safe for all $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$. This completes the base case.

For the inductive step, we note that if $C(y_{t-1})$ is safe for all $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$, then it is safe for θ. This implies that $D_L \leq ay_{t-1} + bC(y_{t-1}) \leq D_U$, which implies that $|y_t| \leq \|D\|_\infty + \bar{w}$. We can therefore use the exact same logic as in the base case to get that $C(y_t)$ will be safe for all $\theta' \in \hat{\theta}_{\text{L43}} \pm \beta$. This completes the proof by induction. ■

Appendix H. Bounding Sources of Regret for Small Noise Case

H.1. Regret of Using Estimated Optimal Unconstrained Controller (Proposition 29)

Proof By Lemma 39, $a^* - b^* F_{\text{opt}}(\theta^*) < 1 - c_{\text{L39}}$, which implies by Lemma 42 that $\bar{J}(\theta^*, C_F^{\text{unc}})$ is twice differentiable at the point $F = F_{\text{opt}}(\theta^*)$ with first and second derivatives that are both finite and independent of T. We also have by Lemma 37 that $|F_{\text{opt}}(\hat{\theta}_{\text{wu}}) - F_{\text{opt}}(\theta^*)| \leq \tilde{O}_T(T^{-1/4})$ conditional on event E_2^0. Therefore, conditional on event E_2^0 and for sufficiently large T, we can do a second order Taylor expansion of $\bar{J}(\theta^*, C_F^{\text{unc}})$ around $F = F_{\text{opt}}(\theta^*)$ to get that

$$\left| T \cdot \bar{J}\left(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}\right) - T \cdot \bar{J}\left(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\theta^*)}\right) \right| = \tilde{O}_T(\sqrt{T}). \tag{141}$$

Because the lowest-cost unconstrained linear controller $C^{\text{unc}}_{F_{\text{opt}}(\theta^*)}$ has the lowest cost among all unconstrained controllers Anderson and Moore (2007),

$$T \cdot \bar{J}(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\theta^*)}) - T \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*, T)}) \leq 0. \tag{142}$$

Combining Equations (141) and (142) and multiplying by $(T - T_0)/T$, we have

$$(T - T_0) \cdot \bar{J}(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}) - (T - T_0) \cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\text{opt}}(\theta^*, T)}) = \tilde{O}_T(\sqrt{T}). \tag{143}$$

Now we just need to convert this to a result about finite time cost rather than infinite cost which requires the following lemma.

Lemma 51 *Under Assumptions 1–3 and 4, for any $\theta \in \Theta$ and K satisfying $1-(a-bK) = \epsilon > 0$ for some $\epsilon = \Omega_T(1)$,*

$$|\bar{J}(\theta, C_K^{\mathrm{unc}}, T) - \bar{J}(\theta, C_K^{\mathrm{unc}})| = O_T\left(\frac{1}{T}\right).$$

The proof of Lemma 51 can be found in Appendix H.5.

For sufficiently large T, conditional on event E_2^0,

$$\begin{aligned} 1-(a^*-b^*F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})) &\geq 1-\left(a^*-b^*F_{\mathrm{opt}}(\theta^*)-\tilde{O}_T(T^{-1/4})\right) && \text{Lemma 37} \\ &> c_{\mathrm{L39}}/2. && \text{Lemma 39} \end{aligned}$$

Therefore, we can apply Lemmas 36 and 51 to Equation (143) to get the desired result that conditional on event E_2^0,

$$(T-T_0)\cdot \bar{J}(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0) - (T-T_0)\cdot \bar{J}(\theta^*, C^{\theta^*}_{K_{\mathrm{opt}}(\theta^*,T)}, T-T_0) = \tilde{O}_T(\sqrt{T}).$$

■

H.2. Regret from Randomness (Proposition 30)

Proof We will apply the standard McDiarmid's inequality to the function

$$f(\{w_t\}_{t=T_0}^{T-1}) = (T-T_0)J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, W').$$

To do this, we need a bounded difference inequality. We will show the following.

Lemma 52 *For $i \in [T_0 : T-1]$, let $\{w'_t\}_{t=T_0}^{T-1}$ be such that $w'_t = w_t$ for $t \neq i$ and $w'_i \sim \mathcal{D}$ is independent of $\{w_t\}_{t=T_0}^{T-1}$. If $|F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}}) - F_{\mathrm{opt}}(\theta^*)| \leq \tilde{O}_T(T^{-1/4})$, then for sufficiently large T,*

$$|(T-T_0)\cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1}) - (T-T_0)\cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, \{w'_t\}_{t=T_0}^{T-1})| \leq c.$$

for some $c = \tilde{O}_T(1)$.

The proof of Lemma 52 can be found in Appendix H.6.

Under event E_2^0, by Lemma 37 we have that $|F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}}) - F_{\mathrm{opt}}(\theta^*)| \leq \tilde{O}_T(T^{-1/4})$. Furthermore, conditional on E_2^0 and $\hat{\theta}_{\mathrm{wu}}$ the random variables $\{w_t\}_{t=T_0}^{T-1}$ are still i.i.d. because the noise random variables are independent of the history. Therefore, conditional on event E_2^0, we can use Lemma 52 with the standard McDiarmid's inequality McDiarmid et al. (1989) and get

$$\begin{aligned} &\mathbb{P}\left(|(T-T_0)\cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1}) - \mathbb{E}[(T-T_0)\cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1})]| \geq \epsilon \;\middle|\; \hat{\theta}_{\mathrm{wu}}\right) \\ &\leq 2\exp\left(-2\frac{\epsilon^2}{c^2(T-T_0)}\right). \end{aligned}$$

Because

$$\mathbb{E}[(T-T_0)\cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1})] = (T-T_0)\cdot \bar{J}(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0),$$

taking $\epsilon = \sqrt{T}c\log(T)$ gives conditional on E_2^0,

$$\begin{aligned} &\mathbb{P}\left(|(T-T_0)\cdot J(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1}) - T-T_0)\cdot \bar{J}(\theta^*, C^{\mathrm{unc}}_{F_{\mathrm{opt}}(\hat{\theta}_{\mathrm{wu}})}, T-T_0)| \geq \sqrt{T}c\log(T) \;\middle|\; \hat{\theta}_{\mathrm{wu}}\right) \\ &= o_T(1/T). \end{aligned} \tag{144}$$

Define

$$E_{\text{P30}} := \left\{ |(T-T_0) \cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1}) - T - T_0) \cdot \bar{J}(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0)| < \sqrt{T} c \log(T) \right\}.$$

By the law of total expectation, Equation (144) implies that

$$\mathbb{P}(E_{\text{P30}} \mid E_2^0) = \mathbb{E}[\mathbb{P}(E_{\text{P30}} \mid \hat{\theta}_{\text{wu}}, E_2^0) \mid E_2^0] \geq 1 - o_T(1/T).$$

Because $\mathbb{P}(E_2^0) \geq \mathbb{P}(E) = 1 - o_T(1/T)$, we therefore have that

$$\mathbb{P}(E_{\text{P30}}) \geq \mathbb{P}(E_{\text{P30}} \mid E_2^0)\mathbb{P}(E_2^0) = 1 - o_T(1/T)$$

as desired.

■

H.3. Regret from Starting Position After Warm-Up (Proposition 31)

Proof

Lemma 53 *Under Assumptions 1–3 and 4, for any $\theta \in \Theta$ and any $K \in [\frac{a-1}{b}, \frac{a}{b}]$, when using controller C_K^{unc} under dynamics θ where $1 - (a - bK) = \epsilon = \Omega_T(1)$ and starting at state $x_0 = x$, then for all t, the state x_t at time t satisfies*

$$|x_t| \leq |x| + \frac{\bar{w}}{\epsilon}.$$

Furthermore, for any x, y, W' and $\tau \leq T$,

$$\begin{aligned} |\tau J(\theta, C_K^{\text{unc}}, \tau, x, W') - \tau J(\theta, C_K^{\text{unc}}, \tau, y, W')| &\leq \frac{(q + rK^2)(x-y)^2 + 2(q+rK^2)\left(|x| + \frac{\bar{w}}{\epsilon}\right)|x-y|}{\epsilon} \\ &= O_T\left((x-y)^2 + |x(x-y)|\right). \end{aligned}$$

The proof of Lemma 53 can be found in Appendix H.7.

By Lemma 37, under event $E \subseteq E_2^0$, we have $|F_{\text{opt}}(\hat{\theta}_{\text{wu}}) - F_{\text{opt}}(\theta^*)| \leq O_T(T^{-1/4})$. Therefore, by Lemma 39, under event E and for large enough T,

$$1 - (a^* - b^* F_{\text{opt}}(\hat{\theta}_{\text{wu}})) \geq c_{\text{L39}} - b^* |F_{\text{opt}}(\hat{\theta}_{\text{wu}}) - F_{\text{opt}}(\theta^*)| \geq c_{\text{L39}}/2.$$

Conditional on event E, C^{alg} is safe for dynamics θ^*, and therefore by Lemma 57, the state of C^{alg} at time T_0 satisfies $|x'_{T_0}| \leq B_x = \tilde{O}_T(1)$ conditional on E. Therefore, by Lemma 53, conditional on E,

$$(T-T_0) \cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, x'_{T_0}, \{w_t\}_{t=T_0}^{T-1}) - (T-T_0) \cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T-T_0, 0, \{w_t\}_{t=T_0}^{T-1}) = \tilde{O}_T(1).$$

■

H.4. Regret from Enforcing Safety (Proposition 32)

Proof Under event E_2^0, we have that for sufficiently large T,

$$\hat{a} - \hat{b}F_{\text{opt}}(\hat{\theta}_{\text{wu}}) \geq a^* - b^*F_{\text{opt}}(\theta^*) - \tilde{O}_T(T^{-1/4}) > 0 \tag{145}$$

by Lemma 39 and Lemma 37. Conditional on event $E_2^0 \cap E_{\text{E56}}$ and for sufficiently large T we have the following result:

$$\begin{aligned}
&\tilde{O}_T(T^{-1/4}) \\
&\geq D_U + \bar{w} - \frac{D_U}{\hat{a} - \hat{b}F_{\text{opt}}(\hat{\theta}_{\text{wu}})} && \text{Equation (56)} \\
&= \frac{D_U}{a^* - b^*K_{D_U}^{\theta^*}} - \frac{D_U}{\hat{a} - \hat{b}F_{\text{opt}}(\hat{\theta}_{\text{wu}})} && \text{Definition of } K_{D_U}^{\theta^*} \\
&\geq \frac{D_U}{a^* - b^*K_{D_U}^{\theta^*}} - \frac{D_U}{a^* - b^*F_{\text{opt}}(\theta^*) - \tilde{O}_T(T^{-1/4})} && \text{Equation (145)} \\
&= \frac{-D_U\tilde{O}_T(T^{-1/4}) + b^*D_U(K_{D_U}^{\theta^*} - F_{\text{opt}}(\theta^*))}{(a^* - b^*F_{\text{opt}}(\theta^*) - \tilde{O}_T(T^{-1/4}))(a^* - b^*K_{D_U}^{\theta^*})} \\
&\geq \frac{-D_U\tilde{O}_T(T^{-1/4}) + b^*D_U(K_{D_U}^{\theta^*} - F_{\text{opt}}(\theta^*))}{(a^* - b^*F_{\text{opt}}(\theta^*))(a^* - b^*K_{D_U}^{\theta^*})} \\
&= (K_{D_U}^{\theta^*} - F_{\text{opt}}(\theta^*))\frac{b^*D_U}{(a^* - b^*F_{\text{opt}}(\theta^*))(a^* - b^*K_{D_U}^{\theta^*})} \\
&\quad - \frac{D_U\tilde{O}_T(T^{-1/4})}{(a^* - b^*F_{\text{opt}}(\theta^*))(a^* - b^*K_{D_U}^{\theta^*})}.
\end{aligned} \tag{146}$$

Because $\theta^*, D_U, F_{\text{opt}}(\theta^*), K_{D_U}^{\theta^*}$ are all independent of T, we can rearrange Equation (146) to get

$$K_{D_U}^{\theta^*} - F_{\text{opt}}(\theta^*) \leq \tilde{O}_T(T^{-1/4}).$$

Combining this with Lemma 37 which states that $|F_{\text{opt}}(\hat{\theta}_{\text{wu}}) - F_{\text{opt}}(\theta^*)| = \tilde{O}_T(T^{-1/4})$ we have that

$$K_{D_U}^{\theta^*} - F_{\text{opt}}(\hat{\theta}_{\text{wu}}) \leq \tilde{O}_T(T^{-1/4}). \tag{147}$$

Conditional on event $E_2^0 \cap E_{\text{safe}}^{\text{wu}}$, $\|\hat{\theta}_{\text{wu}} - \theta^*\|_\infty \leq \epsilon_0 = \tilde{O}_T(T^{-1/4})$ and $|x'_{T_0}| \leq \|D\|_\infty + \bar{w}$. Conditional on $E_2^0 \cap E_{\text{safe}}^{\text{wu}}$, we can apply Lemma 43 with $\theta = \theta^*$, $\hat{\theta}_{\text{L43}} = \hat{\theta}_{\text{wu}}$, $K' = F_{\text{opt}}(\hat{\theta}_{\text{wu}})$, ϵ as the right hand side of Equation (147), $\beta = \epsilon_0$, $\tau = T - T_0$, and $x_0 = x'_{T_0}$. With this choice of parameters, the controller C in Lemma 43 is exactly equivalent to $C^{\text{alg}'}$ under event E_{E56}. Conditional on E_2^0, ϵ and β satisfy the necessary inequality for Lemma 43 as both are $\tilde{O}_T(T^{-1/4})$.

The event $E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}$ depends only on noise random variables before time T_0, which means we can apply Lemma 43 conditional on these events. Equation (92) of Lemma 43 gives that for sufficiently

large T, conditional on $E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}$, and with conditional probability $1 - o_T(1/T)$,

$$\begin{aligned}
&\left|(T-T_0)\cdot J(\theta^*, C^{\text{alg}'}, T-T_0, x'_{T_0}, W') - (T-T_0)\cdot J(\theta^*, C_{K'}^{\text{unc}}, T-T_0, x'_{T_0}, W')\right| \\
&\le (T-T_0)O_T\Bigg(\left(b\epsilon+\epsilon_0+|F_{\text{opt}}(\hat{\theta}_{\text{wu}})|\epsilon_0\right)\log\left(\frac{1}{b\epsilon+\epsilon_0+|F_{\text{opt}}(\hat{\theta}_{\text{wu}})|\epsilon_0}\right) \\
&\qquad\times\left(\left(b\epsilon+\epsilon_0+|F_{\text{opt}}(\hat{\theta}_{\text{wu}})|\epsilon_0\right)+\frac{\log(T)}{\sqrt{T-T_0}}\right)\Bigg) \\
&\le (T-T_0)O_T\left(\tilde{O}_T(T^{-1/4})\log(1/\tilde{\Omega}_T(T^{-1/4}))\left(\tilde{O}_T(T^{-1/4})+\frac{\log(T)}{\sqrt{T-T_0}}\right)\right) && [\epsilon_0=\tilde{\Omega}_T(1)] \\
&= \tilde{O}_T(\sqrt{T}). && (148)
\end{aligned}$$

Taking E_{P32} to be the event that Equation (148) holds gives the desired result that $\mathbb{P}(E_{\text{P32}} \mid E_2^0 \cap E_{\text{safe}}^{\text{wu}} \cap E_{\text{E56}}) = 1 - o_T(1/T)$. ■

H.5. Regret from Finite to Infinite Optimal Linear Controller (Lemma 51)

Proof By Lemma 53, when starting at $x_0 = 0$ and using controller C_K^{unc} we have that

$$|x_T| \le \frac{\bar{w}}{\epsilon}. \tag{149}$$

Therefore, we can conclude that (for $W' = \{w_t\}_{t=0}^{T-1}$):

$$\begin{aligned}
&\left|\bar{J}(\theta, C_K^{\text{unc}}, 2T) - \bar{J}(\theta, C_K^{\text{unc}}, T)\right| \\
&= \left|\frac{T\cdot\bar{J}(\theta, C_K^{\text{unc}}, T) + T\cdot\mathbb{E}\left[\bar{J}(\theta, C_K^{\text{unc}}, T, x_T)\right]}{2T} - \bar{J}(\theta, C_K^{\text{unc}}, T)\right| \\
&= \left|\frac{\mathbb{E}\left[\bar{J}(\theta, C_K^{\text{unc}}, T, x_T)\right]}{2} - \frac{1}{2}\bar{J}(\theta, C_K^{\text{unc}}, T)\right| \\
&= \frac{1}{2T}\left|\mathbb{E}\left[T\bar{J}(\theta, C_K^{\text{unc}}, T, x_T)\right] - T\bar{J}(\theta, C_K^{\text{unc}}, T)\right| \\
&= \frac{1}{2T}\left|\mathbb{E}\left[\mathbb{E}\left[TJ(\theta, C_K^{\text{unc}}, T, x_T, W') - TJ(\theta, C_K^{\text{unc}}, T, 0, W') \;\middle|\; x_T\right]\right]\right| \\
&= \frac{1}{2T}\left|\mathbb{E}\left[O_T\left(x_T^2\right)\right]\right| && \text{Lemma 53} \\
&= O_T\left(\frac{1}{T}\mathbb{E}[x_T^2]\right) \\
&= O_T\left(\frac{1}{T}\mathbb{E}\left[\frac{\bar{w}^2}{\epsilon^2}\right]\right) && \text{Equation (149)} \\
&= O_T\left(\frac{1}{T}\right).
\end{aligned}$$

Furthermore, we have that

$$\begin{aligned}|\bar{J}(\theta, C_K^{\text{unc}}, T) - \bar{J}(\theta, C_K^{\text{unc}})| &= \left|\sum_{i=0}^{\infty} \bar{J}(\theta, C_K^{\text{unc}}, 2^i T) - \bar{J}(\theta, C_K^{\text{unc}}, 2^{i+1}T)\right| \\ &\leq \sum_{i=0}^{\infty} \left|\bar{J}(\theta, C_K^{\text{unc}}, 2^i T) - \bar{J}(\theta, C_K^{\text{unc}}, 2^{i+1}T)\right| \\ &= \sum_{i=0}^{\infty} O_T\left(\frac{1}{T2^i}\right) \\ &= O_T\left(\frac{1}{T}\right).\end{aligned}$$

■

H.6. McDiarmid's Bounded Difference For Total Cost (Lemma 52)

Proof Define $x_{T_0}, .., x_T$ as the states with noise $\{w_t\}_{t=T_0}^{T-1}$ when using controller $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ starting at $x_{T_0} = 0$ and define $y_{T_0}, ..., y_T$ as the states with noise $\{w'_t\}_{t=T_0}^{T-1}$ when using controller $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ starting at $y_{T_0} = 0$. By construction, the cost up until time i is the same for both trajectories. At time $i+1$, we have that

$$|y_{i+1} - x_{i+1}| = |w_i - w'_i| \leq 2\bar{w} = O_T(1). \tag{150}$$

The remaining difference in cost is simply the difference in cost of two length $T' = T - i - 1$ trajectories using controller $C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}$ starting at states y_{i+1} and x_{i+1} respectively. By the assumption of this lemma on $F_{\text{opt}}(\hat{\theta}_{\text{wu}})$ and Lemma 39, we have that for sufficiently large T,

$$1 - (a^* - b^* F_{\text{opt}}(\hat{\theta}_{\text{wu}})) \geq 1 - (a^* - b^* F_{\text{opt}}(\theta^*)) - \tilde{O}_T(T^{-1/4}) \geq c_{\text{L39}}/2.$$

Therefore we can combine Lemma 53 and Equation (150) to get that the difference in the cost from time $i+1$ onward is upper bounded by

$$|T' \cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T', x_{i+1}, \{w_t\}_{t=i+1}^{T-1}) - T' \cdot J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T', y_{i+1}, \{w_t\}_{t=i+1}^{T-1})| = O_T(1). \tag{151}$$

Therefore, we have that (see below for justification)

$$\begin{aligned}&|(T - T_0)J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T - T_0, 0, \{w_t\}_{t=T_0}^{T-1}) - (T - T_0)J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T - T_0, 0, \{w'_t\}_{t=T_0}^{T-1})| \\ &= |(i - T_0)J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, i - T_0, \{w_t\}_{t=T_0}^{i-1}) - (i - T_0)J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, i - T_0, \{w'_t\}_{t=T_0}^{i-1})| + \\ &\qquad |T' J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T', x_{i+1}, \{w_t\}_{t=i+1}^{T-1}) - T' J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T', y_{i+1}, \{w_t\}_{t=i+1}^{T-1})| \\ &= |J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T - i - 1, x_{i+1}, \{w_t\}_{t=i+1}^{T-1}) - J(\theta^*, C^{\text{unc}}_{F_{\text{opt}}(\hat{\theta}_{\text{wu}})}, T - i - 1, y_{i+1}, \{w_t\}_{t=i+1}^{T-1})| \\ &= O_T(1).\end{aligned}$$

Note that in the first equality we also cancelled out the controls at time i which are the same for both trajectories. In the second equality, we used the fact that $\{w_t\}_{t=T_0}^{i-1} = \{w'_t\}_{t=T_0}^{i-1}$, and in the final line we used Equation (151). ■

H.7. Bounding Long-term Position Deviation (Lemma 53)

Proof

By construction, when using C_K^{unc} we have the recursive relationship that $x_t = (a - bK)x_{t-1} + w_{t-1}$. Because we assume that $a - bK = 1 - \epsilon < 1$, we have that

$$|x_t| \leq |x| + \sum_{i=0}^{\infty}(a - bK)^i \bar{w} = |x| + \frac{\bar{w}}{1 - (a - bK)} = |x| + \frac{\bar{w}}{\epsilon} = \beta,$$

where we define $\beta = |x| + \frac{\bar{w}}{\epsilon}$. This proves the first part of the lemma. Furthermore, this implies that the magnitude of the control is never greater than

$$|u_t| = |K||x_t| \leq |K|\beta.$$

Using controller C_K^{unc}, let $x_0, x_1, ..., x_T$ be the sequence of states starting at $x_0 = x$ and let $y_0, y_1, ..., y_T$ be the series of states starting at $y_0 = y$. Define $d_t = |x_t - y_t|$. Note that $d_0 = |x - y|$. Furthermore, for all t,

$$d_t = (a - bK)d_{t-1}.$$

and

$$|C_K^{\text{unc}}(x_t) - C_K^{\text{unc}}(y_t)| = Kd_t.$$

Therefore, we have the following bound.

$$\begin{aligned}
&\left|TJ(\theta, C_K^{\text{unc}}, T, x, W') - TJ(\theta, C_K^{\text{unc}}, T, y, W')\right| \\
&= \left|(qx_T^2 - qy_T^2) + \sum_{t=0}^{T-1} qx_t^2 - qy_t^2 + r(Kx_t)^2 - r(Ky_t)^2\right| \\
&\leq \sum_{t=0}^{T} 2q|x_t|d_t + qd_t^2 + 2r|Kx_t||Kd_t| + rK^2d_t^2 \\
&\leq (2q + 2rK^2)\beta \sum_{t=0}^{T} d_t + (q + rK^2)\sum_{t=0}^{T} d_t^2 && |x_t| \leq \beta \\
&\leq (2q + 2rK^2)\beta \sum_{t=0}^{\infty}(a - bK)^t d_0 + (q + rK^2)\sum_{t=0}^{\infty}(a - bK)^{2t} d_0^2 \\
&= 2(q + rK^2)\beta\frac{|x - y|}{1 - (a - bK)} + \frac{(q + rK^2)(x - y)^2}{1 - (a - bK)^2} \\
&\leq 2(q + rK^2)\beta\frac{|x - y|}{1 - (a - bK)} + \frac{(q + rK^2)(x - y)^2}{1 - (a - bK)} && a - bK < 1 \\
&\leq \frac{2(q + rK^2)\left(|x| + \frac{\bar{w}}{\epsilon}\right)|x - y| + (q + rK^2)(x - y)^2}{\epsilon}. && (152)
\end{aligned}$$

This is exactly the desired result of the second equation of Lemma 53. ■

Appendix I. General Technical Lemmas

The following four lemmas are used throughout the appendix and follow directly from results in Schiffer and Janson (2024).

Lemma 54 (Lemma 13 in Schiffer and Janson (2024)) *Suppose w_t for $t < T$ are sub-Gaussian and F is an event such that $\mathbb{P}(F) = 1 - o_T(1/T^{11})$. Then*

$$\mathbb{E}[\max_{i \le t} w_i^2 \mid \neg F]\mathbb{P}(\neg F) = o_T\left(\frac{1}{T^{10}}\right).$$

Lemma 55 (Lemma 11 in Schiffer and Janson (2024)) *Let x, y be two random variables independent of noises $W' = \{w_i'\}_{i=0}^{t-1}$ such that for some $L = \tilde{O}_T(1)$, both $\mathbb{P}(|x| \ge L)\,\mathbb{E}[x^2 \mid |x| \ge L] = o_T\left(\frac{1}{T^{10}}\right)$ and $\mathbb{P}(|y| \ge L)\,\mathbb{E}[y^2 \mid |y| \ge L] = o_T\left(\frac{1}{T^{10}}\right)$ and $\mathbb{P}(|x| \le 4\log^2(T)) = 1 - o_T(1/T^{11})$ and $\mathbb{P}(|y| \le 4\log^2(T)) = 1 - o_T(1/T^{11})$. Then under Assumptions 1–3, if $\|\theta - \theta^*\|_\infty = \epsilon \le \epsilon_{\mathrm{L3}}$, then for any $K \in (\frac{a-1}{b}, \frac{a}{b})$ and $t \le T$,*

$$\left|\mathbb{E}\left[t \cdot J(\theta^*, C_K^\theta, t, x, W') - t \cdot J(\theta^*, C_K^\theta, t, y, W')\right]\right| = \tilde{O}_T\left(\mathbb{E}[|x-y|] + \epsilon + \frac{1}{T^2}\right). \tag{153}$$

Proof This follows directly by Lemma 11 in Schiffer and Janson (2024) and Lemmas 2 and 3. ■

Lemma 56 (Lemma 12 in Schiffer and Janson (2024)) *Let $x_0, x_1, ...x_T$ be the sequences of states when starting at state $x_0 = x$ and using controller C_t at time t. Suppose that the control $C_t(x_t)$ is safe for dynamics θ_t and $\|\theta_t - \theta^*\| \le \frac{1}{\log(T)}$ for all $t < T$. For sufficiently large T,*

$$\forall t \le T,\ |x_t| = O_T(|x| + \|D\|_\infty + \max_{i \le t-1} |w_i|).$$

$$\forall t < T,\ |C_t(x_t)| = O_T(|x| + \|D\|_\infty + \max_{i \le t-1} |w_i|).$$

Lemma 57 (Lemma 4 in Schiffer and Janson (2024)) *Let $|x_0| \le 4\log^2(T)$. Suppose for all $t < T$, the control used by controller C_t at time t is safe for fixed dynamics θ_t and for all $t \le T$,*

$$\|\theta^* - \theta_t\|_\infty \le \frac{1}{\log(T)}. \tag{154}$$

Then under Assumptions 1–3, for sufficiently large T and conditioned on event E_1, using this controller C_t with dynamics θ^* for T steps starting at x_0 will give states $(x_0, ..., x_T)$ and controls $(u_0, ..., u_{T-1})$ satisfying the following equations.

$$|x_t| \overset{a.s.}{\le} 4\log^2(T) < \log^3(T) := B_x \tag{155}$$

$$|u_t| \overset{a.s.}{\le} O_T(\log^2(T)) < \log^3(T) := B_x. \tag{156}$$

Furthermore, if x_0 and the controller C_t are deterministic, then the states $(x_0, ..., x_T)$ and controls $(u_0, ..., u_{T-1})$ satisfy

$$\mathbb{E}[|x_t|] \le 4\log^2(T) < \log^3(T) := B_x \tag{157}$$

$$\mathbb{E}[|u_t|] \le O_T(\log^2(T)) < \log^3(T) := B_x. \tag{158}$$

$$\mathbb{P}(E_1) \ge 1 - \sum_{t=0}^{T-1} 2\exp\left(-\log^4(T)/\alpha\right) = 1 - o_T\left(\frac{1}{T^{\log(T)}}\right). \tag{159}$$

Appendix J. Higher Dimension Extended Discussions

To prove Theorem 2, we introduced several new tools for technical analysis of non-linear controllers. We believe that many of our results and techniques will generalize to higher dimensions, but due to the already very complex nature of the proofs in one-dimension (and already very long appendix), this is beyond the scope of the current paper. That being said, in this section we give a high-level discussion on how we believe that our results from Theorem 2 can extend to higher dimensions with a series of unproven conjectures.

For this section only, we will assume that the states are in $\mathbb{R}^n$ and the controls are in $\mathbb{R}^m$, and the unknown dynamics $\Theta^* = (A^*, B^*)$ are matrices of the appropriate dimensions. We then let the cost at time t be $x_t^\top Q x_t + u_t^\top R u_t$ for known $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$. We can then generalize the safety of Definition 1 to require that for all t, $D\left(A^* x_t + B^* u_t\right) \leq d$ for known matrix D and vector d (and element-wise comparisons). As in the single dimension case, this definition is strictly more general than restricting the realized position when noise is bounded. The first important decision we must make for higher dimensions is how to generalize the class of truncated linear controllers. One natural such way is the following:

$$C_K^\Theta(x) = \begin{cases} -Kx & \text{if } D(A - BK)x \leq d \\ \underset{u : D(Ax + Bu) \leq d}{\arg\min} \; u^\top R u & \text{otherwise} \end{cases} \tag{160}$$

With this generalization truncated linear controllers, we expect that versions of Lemmas 2 and 3 will directly hold. Intuitively, the Lemma 2 generalization implies that the cost of a truncated linear controller is continuous in the dynamics. This should still be true in higher dimensions, as the controls themselves are still continuous in the dynamics Θ and the parameter K, which implies that $\|C^{\Theta^*}_{K_{\text{opt}}(\Theta^*,t)}(x) - C^{\Theta}_{K_{\text{opt}}(\Theta,t)}(x)\|_2$ will be small for any x as long as $\|\Theta - \Theta^*\|_2$ is small. To see that Lemma 3 will generalize, recall that this lemma states that the T step cost of using the truncated linear controller is continuous in the starting position. As long as the eigenvalues of $A - BK$ are sufficiently small (as should be true for the optimal controller), the truncated linear controller will always be shrinking the magnitude of the state. Therefore, the difference between starting at state x versus y will shrink over time, leading to a small difference in the total cost. To see formal statements of the conjectured generalizations of these lemmas, see Appendix J.1.

We next describe a natural way that Algorithm 2 can generalize to higher dimensions. Note that the warm-up phase and regularized least squares estimate of the unknown dynamics in Algorithm 2 can directly generalize to higher dimensions, as the uncertainty bound in Abbasi-Yadkori and Szepesvári (2011) holds in higher dimensions. There are two key modifications that Algorithm 2 needs for higher dimensions. The first is to generalize how we define C_s^{alg}, i.e. how to split the cases of "large noise" versus "small noise". One natural generalization is the following, where we split depending on whether the best unconstrained linear controller for the estimated dynamics keeps the state sufficiently inside the safety-constrained region.

$$C_s^{\text{alg}} \leftarrow \begin{cases} C^{\text{unc}}_{F_{\text{opt}}(\hat{\Theta}_{\text{wu}})} & \text{if } \max_{x \in \mathbb{R}^n : |x_i| \leq c + \bar{w}} \min_{z' : Dz' \leq d} \|(\hat{A}_{\text{wu}} - \hat{B}_{\text{wu}} F_{\text{opt}}(\hat{\Theta}_{\text{wu}}))x - z'\|_2 \leq O_T(T^{-1/4}) \\ C^{\hat{\Theta}_s}_{K_{\text{opt}}(\hat{\Theta}_s)} & \text{otherwise} \end{cases}$$

The second important change in higher dimensions is determining how to generalize the safe controls u_t^{safeU} and u_t^{safeL}. One natural way to do this is to consider the lowest cost control that takes the position sufficiently within the safety region.

$$u_t^{\text{safe}} \leftarrow \begin{cases} C_s^{\text{alg}}(x_t) & \text{if } D(\hat{A}_s x_t + \hat{B}_s C_s^{\text{alg}}(x_t)) \leq d - \epsilon_s \\ \underset{u \in \mathbb{R}^m}{\arg\min} \left\{ u^\top R u : D(\hat{A}_s x_t + \hat{B}_s u) \leq d - \epsilon_s \right\} & \text{otherwise} \end{cases}$$

To see a more complete sketch of this algorithm, see Appendix J.1. We leave formal studying of whether this algorithm gives provably low regret to future works, however we discuss one simple higher dimensional

case in which we believe our results can directly generalize. Suppose we are in the setting where $D \in \mathbb{R}^{2n \times n}$ and the $2i-1$ row of D is $\mathbf{e}_i$ and the $2i$ row of D is $-\mathbf{e}_i$ for $i \in [1:n]$. Further, suppose that $d = (c, c, ..., c) \in \mathbb{R}^{2n}$. In other words, the safe region is a box centered at the origin with edge length $2|c|$. We will also assume for simplicity that B^* is invertible and that the noise distribution is symmetric around the origin. As in the one-dimensional setting, we expect that in this case there should be a perfect dichotomy between when the noise is "large enough" that the controls will be sufficiently non-linear and when the noise is "small enough" that the optimal unconstrained linear controller is close to being safe. Intuitively, this is because the symmetry of the constraints and noise implies that we effectively can split the analysis of C_s^{alg} into two cases again. The first case is when the noise is sufficiently small that the optimal linear controller is close to being safe, and therefore our algorithm would use the optimal certainty equivalence linear controller with small additional truncations as needed. Because the noise is sufficiently small in this case, the small truncations cause only a small amount of regret, and we know that the optimal certainty equivalence linear controller also has low regret. The second case is when the noise is sufficiently large that the truncation (use of u_t^{safe}) will occur sufficiently many times, in which case the dynamics will be sufficiently non-linear for the uncertainty to decrease at a faster rate leading to low regret. We leave formal studying of this setting and the general higher dimensional setting to future work. See Appendix J.1 for more formal conjectures and algorithm.

J.1. Higher Dimension Conjectures

Below we state conjectures that are the higher-dimension equivalents of Lemmas 2 and 3.

Conjecture 3 *There exists $\epsilon_{\text{C3}} = \tilde{\Omega}_T(1)$ such that for any $\|\Theta - \Theta^*\|_\infty \le \epsilon_{\text{C3}}$ and $t \le T$,*

$$|J^*(\Theta^*, C^{\Theta}_{K_{\text{opt}}(\Theta,t)}, t) - J^*(\Theta^*, C^{\Theta^*}_{K_{\text{opt}}(\Theta^*,t)}, t)| \le \tilde{O}_T\left(\|\Theta - \Theta^*\|_\infty + \frac{1}{T^2}\right).$$

Conjecture 4 *There exist $\epsilon_{\text{C4}}, \delta_{\text{C4}} = \tilde{\Omega}_T(1)$ such that for any Θ satisfying $\|\Theta - \Theta^*\|_\infty \le \epsilon_{\text{C4}}$ the following holds. For $t < T$, let $W' = \{w_i\}_{i=0}^{t-1}$. Then for any $K \in \mathbb{R}^{m \times n}$ such that the eigenvalues of $A - BK$ have magnitude strictly less than 1, there exists a set $\mathcal{Y}_{\text{C4}} \in \mathbb{R}^t$ that depends only on C_K^{Θ} such that the following holds. Define $E_{\text{C4}}\left(C_K^{\Theta}, W'\right)$ as the event that $W' \in \mathcal{Y}_{\text{L4}}$. Then $\mathbb{P}(E_{\text{C4}}\left(C_K^{\Theta}, W'\right)) \ge 1 - o_T(1/T^{10})$ and for any $\|x\|_2, \|y\|_2 \le 4\log^2(T)$ such that $\|x - y\|_2 \le \delta_{\text{C4}}$, conditional on event $E_{\text{C4}}\left(C_K^{\Theta}, W'\right)$,*

$$\left|t \cdot J(\Theta^*, C_K^{\Theta}, t, x, W') - t \cdot J(\Theta^*, C_K^{\Theta}, t, y, W')\right| \le \tilde{O}_T(\|x - y\|_2 + \|\Theta - \Theta^*\|_\infty). \tag{161}$$

Below we present a sketch of the algorithm described in the body that is a higher dimension generalization of Algorithm 2 which we conjecture has low regret.

Algorithm 3 Algorithm Sketch for Higher Dimensions

Input: $D, \mathcal{D}, \Theta, C^{\text{init}}, T, \lambda$

for $t \leftarrow 0$ *to* $\sqrt{T} - 1$ **do**

  $\phi_t \sim \text{Rademacher}(0.5)$ Use control $u_t = C^{\text{init}}(x_t) + \frac{\phi_t}{\log(T)}$

$\hat{\Theta}_{\text{wu}} \leftarrow (Z_{\sqrt{T}}^\top Z_{\sqrt{T}} + \lambda I)^{-1} Z_{\sqrt{T}}^\top X_{\sqrt{T}}$

for $s \leftarrow 0$ *to* $\log_2(\sqrt{T}) - 1$ **do**

  $T_s \leftarrow 2^s \sqrt{T}$

  $\epsilon_s \leftarrow$ Error based on Abbasi-Yadkori and Szepesvári (2011)

  $\hat{\Theta}_s^{\text{pre}} \leftarrow$ Regularized least squares estimate of Θ^*

  $\hat{\Theta}_s \leftarrow \arg\max_{\|\Theta - \hat{\Theta}_s^{\text{pre}}\| \le \epsilon_s} \|A - BK_{\text{opt}}(\Theta)\|_2$

  $$C_s^{\text{alg}} \leftarrow \begin{cases} C^{\text{unc}}_{F_{\text{opt}}(\hat{\Theta}_{\text{wu}})} & \text{if } \max_{x \in \mathbb{R}^n : |x_i| \le c + \bar{w}} \min_{z' : Dz' \le d} \|(\hat{A}_{\text{wu}} - \hat{B}_{\text{wu}} F_{\text{opt}}(\hat{\Theta}_{\text{wu}}))x - z'\|_2 \le O_T(T^{-1/4}) \\ C^{\hat{\Theta}_s}_{K_{\text{opt}}(\hat{\Theta}_s)} & \text{otherwise} \end{cases}$$

  for $t \leftarrow T_s$ *to* $2T_s - 1$ **do**

    $$\text{Use } u_t^{\text{safe}} \leftarrow \begin{cases} C_s^{\text{alg}}(x_t) & \text{if } D(\hat{A}_s x_t + \hat{B}_s C_s^{\text{alg}}(x_t)) \le d - \epsilon_s \\ \arg\min_{u \in \mathbb{R}^m} \left\{ u^\top R u : D(\hat{A}_s x_t + \hat{B}_s u) \le d - \epsilon_s \right\} & \text{otherwise} \end{cases}$$