Minimal Expected Regret in Linear Quadratic Control

Yassir Jedra
KTH, Royal Institute of Technology
Stockholm, Sweden, jedra@kth.se

Alexandre Proutiere
KTH, Royal Institute of Technology
Stockholm, Sweden, alepro@kth.se

Abstract

We consider the problem of online learning in Linear Quadratic Control systems whose state transition and state-action transition matrices A and B may be initially unknown. We devise an online learning algorithm and provide guarantees on its expected regret. This regret at time T is upper bounded (i) by $\widetilde{O}((d_u + d_x)\sqrt{d_x T})$ when A and B are unknown, (ii) by $\widetilde{O}(d_x^2 \log(T))$ if only A is unknown, and (iii) by $\widetilde{O}(d_x(d_u + d_x)\log(T))$ if only B is unknown and under some mild non-degeneracy condition (d_x and d_u denote the dimensions of the state and of the control input, respectively). These regret scalings are minimal in T, d_x and d_u as they match existing lower bounds in scenario (i) when $d_x \leq d_u$ (Simchowitz and Foster, 2020), and in scenario (ii) (Lai, 1986). We conjecture that our upper bounds are also optimal in scenario (iii) (there is no known lower bound in this setting).

Existing online algorithms proceed in epochs of (typically exponentially) growing durations. The control policy is fixed within each epoch, which considerably simplifies the analysis of the estimation error on A and B and hence of the regret. Our algorithm departs from this design choice: it is a simple variant of certainty-equivalence regulators, where the estimates of A and B and the resulting control policy can be updated as frequently as we wish, possibly at every step. Quantifying the impact of such a constantly-varying control policy on the performance of these estimates and on the regret constitutes one of the technical challenges tackled in this paper.

Proceedings of the 25[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151.

1 INTRODUCTION

The Linear Quadratic Regulator (LQR) problem arguably constitutes the most iconic, studied, and applied problem in control theory. In this problem, the system dynamics are approximated by those of a linear system, which, in discrete time, are $x_{t+1} = Ax_t + Bu_t + \eta_t$. x_t and u_t represent the state and action vectors at time t, respectively, and typically, the noise sequence $(\eta_t)_{t\geq 0}$ is i.i.d. Gaussian. The decision maker experiences an instantaneous cost quadratic in both the state and the control input $x_t^\top Q x_t + u_t^\top R u_t$, where Q and R are positive semidefinite matrices. Her objective is to devise a control policy minimizing her accumulated long term expected cost. In the perfect information setting (as investigated in this paper), the state is observed without noise, and is plugged in as an input in the control policy. When the state transition and state-action transition matrices A and B are known, the optimal control policy is a simple feedback control $u_t = K_\star x_t$ where $K_\star = -\left(R + B^\top P_\star B\right)^{-1} B^\top P_\star A$ and $P_\star$ solves the discrete algebraic Riccati equation.

This paper considers the online learning version of the LQR problem, where the matrices A and B may be initially unknown. In such a scenario, the decision maker must control the system while learning these matrices. Early efforts in the control community (Åström and Wittenmark, 1973; Kumar, 1985) were devoted to establish the convergence and asymptotic properties of adaptive control algorithms. The regret analysis of these algorithms, initiated by Lai (1986); Lai and Wei (1987) (in very specific cases) and by Abbasi-Yadkori and Szepesvári (2011), has attracted a lot of attention recently, see e.g., Mania et al. (2019); Cohen et al. (2019); Shirani Faradonbeh et al. (2020); Simchowitz and Foster (2020); Abeille and Lazaric (2020); Lale et al. (2020); Cassel et al. (2020). Refer to §2 for details, and to Appendix B for an even longer discussion. Despite these efforts, the picture remains blurry. We are unable from the aforementioned literature (see Ziemann and Sandberg (2021)) to determine the conditions on (A, B, R, Q) under which one can achieve a regret scaling as $O(\log(T))$ or $O(\sqrt{T})$.

The objective of this work is to devise online algorithms for the LQR problem whose expected regret exhibits the best scaling in both the time horizon T and the dimensions d_x and d_u of the state and control input, in the three envisioned scenarios: (I) when (A, B) are unknown, (II) when A only is unknown, (III) when B only is unknown. In addition to guaranteeing minimal regret, we wish our algorithms to enjoy a simple and natural design and to require as little inputs as possible. Existing algorithms proceed in *epochs* of exponentially growing durations. The control policy is fixed within each epoch. This doubling trick considerably simplifies the analysis of the estimation error on A and B and hence of the regret, but seems rather impractical. It is also worth noting that most algorithms take as input a level of confidence δ, the time horizon T, a stabilizer $K_\circ$, but also sometimes known upper bounds on the norms of A and B (for details, refer to §2 and to Appendix B).

Contributions. We propose CEC($\mathcal{T}$) a family of algorithms based on the certainty equivalence principle and with the following desirable properties. The control policy is not fixed within epochs, and may be updated continuously as the estimates of A and B and hence of the optimal control improve. These updates are performed at the rounds in $\mathcal{T} \subset \mathbb{N}$, and our regret guarantees hold for a wide variety of set $\mathcal{T}$ ranging from $\mathcal{T} = \mathbb{N}$ (updates every round) to $\mathcal{T} = \{2^k : k \in \mathbb{N}\}$ (doubling trick). CEC($\mathcal{T}$) is *anytime*, and does not leverage any a priori information about the system (except for the knowledge of a stabilizing controller). We provide upper bounds of the expected regret of our algorithms in the three envisioned scenarios:

In Scenario I where A and B are unknown, the expected regret of CEC($\mathcal{T}$) is[1] $\widetilde{O}((d_u+d_x)\sqrt{d_x T})$, which matches the existing lower bound derived by Simchowitz and Foster (2020) when $d_x \leq d_u$.

In Scenario II where A only is unknown, the expected regret of CEC($\mathcal{T}$) is $\widetilde{O}(d_x^2 \log(T))$, matching the lower bound derived by Lai (1986).

In Scenario III where B only is unknown, the expected regret of CEC($\mathcal{T}$) is $\widetilde{O}(d_x(d_u + d_x)\log(T))$ under the assumption that $K_\star$ is full-rank; there is no lower bound in this setting, but we conjecture that our upper bound exhibits the optimal scalings in T, d_x, and d_u.

It is worth comparing the design and performance guarantees of CEC($\mathcal{T}$) to those of existing algorithms. For Scenario I, Simchowitz and Foster (2020) present an algorithm achieving the same regret upper bound as ours but with probability $1-\delta$ where δ is an input of the algorithm (see the next section for a detailed discussion on the difference between regret in expectation and with high probability), and using a doubling trick. For Scenarios II and III, the best regret guarantees were $O(\mathrm{poly}(d_x, d_u)\log^2(T))$ (Cassel et al., 2020) with unspecified polynomial dependence in (d_x, d_u) (with degree at least 8 as far as we can infer from the analysis of Cassel et al. (2020)). These guarantees were achieved by an algorithm with several inputs, including the time horizon, a stabilizer, upper bounds on $\|A\|$, $\|B\|$, on the minimal ergodic cost, and that achieved under the stabilizer. Refer to Appendix B for a detailed discussion.

Further note that our algorithm, CEC($\mathcal{T}$), is anytime and does not apply any kind of doubling trick: it is a simple variant of certainty-equivalence regulators, where the estimates of A and B and the resulting control policy can be updated as frequently as we wish, possibly at every step. Quantifying the impact of such a constantly-varying control policy on the performance of these estimates and on the regret has eluded researchers and constitutes the main technical challenges tackled in this paper. We address this challenge by developing a novel decomposition of the cumulative covariates matrix (sometimes referred to as Gram matrix), and by deriving concentration results on its spectrum.

2 RELATED WORK

The online LQR problem has received a lot of attention in the control and learning communities. The research efforts towards the development of algorithms with regret guarantees have recently intensified. We may categorize these algorithms into two classes.

In the first class, we find algorithms based on the so-called self-tuning regulators (Åström and Wittenmark, 1973), as those developed in second half of the 20th century in the control community, see Kumar (1985); Matni et al. (2019). Self-tuning regulators work as follows. At any given step, they estimate the unknown matrices A and B, and apply a control policy corresponding to the optimal control obtained replacing A and B by their estimators. To ensure an appropriate level of excitation of the system, and the ability to learn A and B, the control inputs are typically perturbed using white noise. The algorithms developed by Lai (1986); Lai and Wei (1987); Rantzer (2018); Shirani Faradonbeh et al. (2020); Mania et al. (2019); Simchowitz and Foster (2020); Cassel et al. (2020) obey these principles. In the second class, we find algorithms applying the Optimism in Front of Uncertainty (OFU) principle, extensively used to devise regret optimal algorithms in stochastic bandit problems Lai and Robbins (1985); Lattimore and Szepesvári (2020). These algorithms maintain confidence ellipsoids where the

[1] The notation $\widetilde{O}$ hides logarithmic factors.

system parameters lie with high probaility, and select optimistically a system in this ellipsoid to compute the control policy, see Abbasi-Yadkori and Szepesvári (2011); Faradonbeh et al. (2017); Cohen et al. (2019); Abeille and Lazaric (2020); Lale et al. (2020). A description of these algorithms and of their regret guarantees can be found in Appendix B.

All these algorithms use a doubling trick which considerably simplifies their analysis, and most of them are not anytime (as they use the time-horizon as input). For Scenario I where A and B are unknown, all are designed in the fixed confidence setting, and enjoy regret guarantees with a fixed confidence level δ. The best regret upper bound so far is $\widetilde{O}(d_u\sqrt{d_x T\log(1/\delta)})$ with probability $1-\delta$ (Simchowitz and Foster, 2020). For Scenarios II (B known) and III (A known), Cassel et al. (2020) present an algorithm with an expected regret scaling as $O(\mathrm{poly}(d_x,d_u)\log^2(T))$. CEC($\mathcal{T}$) offers much better guarantees with a simplified design, and is anytime.

We conclude this section with a brief discussion on the differences between regret guarantees in expectation or with high probability. Most existing online algorithms for the LQR problem have regret guarantees holding with a fixed level of confidence $1-\delta$ where δ is an input of the algorithms. Their regret analysis consists in identifying a "good" event under which the algorithm behaves well and holding with probability at least $1-\delta$. Devising algorithms with expected regret upper bounds is more involved since one needs to also analyze the behavior of the algorithm under the complementary event (the "bad" event). In turn, analyzing the expected regret requires a deeper understanding of the problem. There is however a method to transform an algorithm with regret guarantees with probability $1-\delta$ to an algorithm with expected regret guarantees: it consists in tuning δ as a function of the time horizon T, and in controlling the regret under the bad event. For example, consider the algorithm of Simchowitz and Foster (2020); in Scenario I where A and B are unknown, the regret of this algorithm is upper bounded by $C\sqrt{T}\log(1/\delta)$ with probability $1-\delta$. Now choosing $\delta = 1/T^2$ and by applying the stabilizing controller when the state norm exceeds some threshold, it can be shown that the expected regret of the modified algorithm scales at most as $C\sqrt{T}\log(T)$. Note that this method induces a multiplicative regret cost proportional to $\log(T)$ and leads to an algorithm that requires the time horizon as input. We believe that because of the additional $\log(T)$ multiplicative cost, this method would lead to sub-optimal expected regret guarantees in Scenarios II and III (there is, anyway, no algorithms in these settings with high probability regret upper bounds).

3 PRELIMINARIES AND ASSUMPTIONS

The LQR problem. We consider a linear dynamical system $x_{t+1} = Ax_t + Bu_t + \eta_t$ as described in the introduction, and initial state $x_0 = 0$. $(\eta_t)_{t\geq 0}$ is a sequence of i.i.d. zero-mean, isotropic[2], σ^2-sub-gaussian random vectors. The objective of the decision maker is to identify a control policy $(u_t)_{t\geq 0}$ minimizing the following ergodic cost $\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}[\sum_{t=1}^T x_t^\top Q x_t + u_t^\top R u_t]$, where Q and R are positive semidefinite matrices. Under the assumption that (A,B) is stabilizable, the Discrete Algebraic Riccati Equation (DARE) $P = A^\top PA - A^\top PB(R+B^\top P^\top B)^{-1}B^\top PA + Q$ admits a unique positive definite solution $P_\star$ (Kučera, 1972), and that the optimal control $(u_t)_{t\geq 0}$ that minimizes the above objective is defined as

$$\forall t \geq 0, \quad u_t = K_\star x_t, \tag{1}$$

with $K_\star = -\left(R + B^\top P_\star B\right)^{-1} B^\top P_\star A$. The minimum ergodic cost achieved under the feedback controller $K_\star$ is denoted by $\mathcal{J}^\star_{(A,B)} = \limsup_{T\to\infty}\frac{1}{T}\mathbb{E}[\sum_{t=1}^T x_t^\top (Q + K_\star^\top R K_\star) x_t]$.

Regret in the online LQR problem. We investigate scenarios where A and / or B are initially unknown. An adaptive control algorithm π is defined as a sequence of measurable functions u_t from the past observations to a control input: for any $t \geq 0$, u_t is $\mathcal{F}_t$-measurable where $\mathcal{F}_t = \sigma(x_0, u_0, \ldots, x_{t-1}, u_{t-1}, x_t)$. The performance of the algorithm π is assessed through its regret defined as:

$$R_T(\pi) = \sum_{t=1}^T (x_t^\top Q x_t + u_t^\top R u_t) - T\mathcal{J}^\star_{(A,B)},$$

where (x_t, u_t) are the state and control input at time t under π. The above regret definition is used in most related papers, and somehow assumes that an Oracle algorithm (aware of A and B initially) would pay a cost of $\mathcal{J}^\star_{(A,B)}$ in each round. We discuss an alternative definition in Appendix C, and justify this definition when the expected regret is the quantity of interest.

Assumptions. Throughout the paper, we make the following assumptions. (i) $(\eta_t)_{t\geq 0}$ is a sequence of i.i.d. zero-mean, isotropic, σ^2-sub-gaussian random vectors. (ii) We assume w.l.o.g.[3] that $Q \succ I_{d_x}$ and $R = I_{d_u}$. (iii)

[2] We say that a random vector η is isotropic if $\mathbb{E}[\eta\eta^\top] = I_d$. If η is zero-mean, isotropic and σ^2-subgaussian, then we also have $1 \leq 4\sigma^2$. The isotropy assumption is without loss of generality because if $\mathbb{E}[\eta\eta^\top] = \Sigma \succ 0$ then we can rescale the dynamics by $\Sigma^{-1/2}$.

[3] This is achieved by a change of basis in the state and input spaces, and by rescaling the dynamics. See Simchowitz and Foster (2020).

We assume as in most existing papers that the system (A, B) is stabilizable, and that the learner has access to a stabilizing controller $K_\circ$ (i.e, $\rho(A + BK_\circ) < 1$).

4 THE CEC($\mathcal{T}$) ALGORITHM

The pseudo-code of our algorithm, CEC($\mathcal{T}$), is presented in Algorithm 1. It essentially based on the Certainty Equivalence principle: the control policy applied at time t is the optimal control policy obtained by replacing (A, B) by their Least Squares Estimators (LSEs). However CEC($\mathcal{T}$) includes three additional components described in more details below. First, the control inputs are perturbed to ensure a sufficient excitation of the system (so that the LSE is consistent). Then, CEC($\mathcal{T}$) exploits the stabilizing controller $K_\circ$ to avoid pathological cases where the system state could become unstable. The use of $K_\circ$ is driven by an hysteresis switching mechanism. Finally, the LSE of (A, B) and the corresponding optimal policy can be updated at will, as frequently as we wish. Hence, CEC($\mathcal{T}$) allows for lazy updates, which can be interesting in case of low computational budget.

Certainty Equivalence and lazy updates. CEC($\mathcal{T}$) takes as input an infinite set $\mathcal{T} \subset \mathbb{N}$ corresponding to the times when the control policy is updated. At such times, we compute the LSE (A_t, B_t) of (A, B) (see Appendix F for a pseudo-code). For example in Scenario I, we have: for $t \geq 2$,

$$[A_t \quad B_t] = \left(\sum_{s=0}^{t-2} x_{s+1} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right) \left(\sum_{s=0}^{t-2} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right)^{-1}. \tag{2}$$

Refer to Appendix F for the expressions of the LSE in the other scenarios. You will note that, when B only is unknown (Scenario III), then at time t, we only use the sample path $(x_0, u_0, \ldots, x_{t-2}, u_{t-2}, x_{t-1})$ to compute the LSE of B. This ensures that η_t and K_{t+1} are independent, which will turn to be crucial in our analysis. From the LSE (A_t, B_t), we compute the updated control policy by solving Ricatti equations:

$$P_t = A_t^\top P_t A_t - A_t^\top P_t B_t (R + B_t^\top P_t B_t)^{-1} B_t^\top P_t A_t + Q, \tag{3}$$

$$K_t = -(R + B_t^\top P_t B_t)^{-1} B_t^\top P_t A_t. \tag{4}$$

Regarding the set $\mathcal{T}$ of update times, we just assume that for some constant $C > 1$,

$$\mathcal{T} = (t_k)_{k\geq 1} \quad \text{with} \quad \forall k \geq 1, \quad t_k < t_{k+1} \leq C t_k. \tag{5}$$

These conditions are very general and are compatible with $\mathcal{T} = \mathbb{N}$ (update every round) and $\mathcal{T} = \{2^k, k \in \mathbb{N}\}$ (doubling trick).

Hysteresis switching and stability. In CEC($\mathcal{T}$), the calls of the stabilizer $K_\circ$ is driven by an hysteresis switching mechanism defined by two sequences of stopping times $(\tau_k, \upsilon_k)_{k\geq 1}$ where

$$\tau_k = \inf\left\{t > \upsilon_k : \sum_{s=0}^{t} \|x_s\|^2 > \sigma^2 d_x \, g(t)\right\},$$

$$\upsilon_k = \inf\left\{t > \tau_{k-1} : \sum_{s=0}^{t} \|x_s\|^2 < \sigma^2 d_x \, f(t)\right\},$$

with $\tau_0 = 0$ and $g(t) \geq f(t)$ for all $t \geq 1$. By construction, the sequences are interlaced: for all $k \geq 1$, $\tau_{k-1} < \upsilon_k < \tau_k$. The use of $K_\circ$ is done as follows. For $\tau_k < t < \upsilon_{k+1}$, we use $K_\circ$ until the growth rate of $\sum_{s=0}^t \|x_s\|^2$ decreases from that of $g(t)$ to that of $f(t)$. For $\upsilon_{k+1} \leq t < \tau_{k+1}$, we use the adaptive controller K_t. We choose $f(t) = t^{1+\gamma/2}$, $g(t) = t^{1+\gamma}$ and $h(t) = t^\gamma$, where $\gamma > 0$ (for the analysis, we need to have $g(t) > f(t) > t$). With these choices, we will establish that the expected number of times when $K_\circ$ is used is finite. This means that after some time, CEC($\mathcal{T}$) only uses the certainty equivalence controller.

Input perturbations. A sufficient excitation of the system is achieved by sometimes adding noise to the control inputs – mainly in Scenario I. In CEC($\mathcal{T}$), $(\nu_t)_{t\geq 0}$ and $(\zeta_t)_{t\geq 0}$ are sequences of independent random vectors where for all $t \geq 1$, $\nu_t \sim \mathcal{N}(0, \sigma_t^2 I_{d_u})$ and $\zeta_t \sim \mathcal{N}(0, I_{d_u})$. We choose $\sigma_t^2 = \sqrt{d_x}\sigma^2/\sqrt{t}$.

5 REGRET GUARANTEES

In this section, we present our main results. We provide finite-time expected regret upper bounds for the CEC($\mathcal{T}$) algorithm in the three envisioned scenarios, and give a sketch of the way they are derived. In the statement of the results, for simplicity, we use the following notations: the relationship $\lesssim$ corresponds to $\leq$ up to a universal multiplicative constant. For any matrix M with $\rho(M) < 1$, we define $\mathcal{G}_M = \sum_{s=0}^\infty \|M^s\|$, and for any matrix K such that $\rho(A + BK) < 1$, we denote $P_\star(K) = \sum_{t=0}^\infty ((A+BK)^t)^\top (Q + K^\top R K)(A + BK)^t$. We further introduce $\mathcal{G}_\circ = \mathcal{G}_{A+BK_\circ}$, $C_B = \max(\|B\|, 1)$, $C_\circ = \max(\|A\|, \|B\|, \|BK_\circ\|, \|K_\circ\|, 1)$, and $C_K = \max(\|K_\circ\|, \|K_\star\|, 1)$.

5.1 Expected Regret Upper Bounds

Theorem 1. *(Scenario I – A and B unknown) Let the set of update times $\mathcal{T}$ satisfy (5). The regret of $\pi = \text{CEC}(\mathcal{T})$ with input $\mathcal{T}$ satisfies in Scenario I: for all $T \geq 1$,*

$$\mathbb{E}[R_T(\pi)] \leq C_1 \sqrt{d_x}(d_x + d_u)\sqrt{T}\log(T) + C_2,$$

with $C_1 \lesssim \sigma^2 C_K^2 C_B^2 \|P_\star\|^{9.5} \log(e\sigma d_x d_u \mathcal{G}_\circ C_\circ \|P_\star\| C_B)^2$ and $C_2 \lesssim \text{poly}(\sigma, d_x, d_u, \mathcal{G}_\circ, C_\circ, \|P_\star\|, C_B)$.

Algorithm 1: Certainty Equivalence Control ($\mathcal{T}$) (CEC($\mathcal{T}$))

input : Cost matrices Q and R, a stabilizing controller $K_\circ$, variance proxy σ of the noise, set of rounds $\mathcal{T}$ for the controller updates.

$\ell_{-1} \leftarrow 0$, $K_{-1} \leftarrow 0$;
for $t \geq 0$ **do**

$$\ell_t \leftarrow \begin{cases} 0 & \text{if} \quad \sum_{s=0}^{t} \|x_s\|^2 > \sigma^2 d_x g(t), \\ 1 & \text{if} \quad \sum_{s=0}^{t} \|x_s\|^2 < \sigma^2 d_x f(t), \\ \ell_{t-1} & \text{otherwise.} \end{cases}$$

if $(t \in \mathcal{T})$ compute (A_t, B_t) (applying the LSE algorithm, see Appendix F);
compute (P_t, K_t) (solving Riccati equations using (A_t, B_t));
else $K_t \leftarrow K_{t-1}$;
Scenario I – (A, B) unknown:

$$u_t \leftarrow \begin{cases} K_t x_t + \nu_t & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \text{ and } \lambda_{\min}\left(\sum_{s=0}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right) \geq t^{1/4} \\ K_\circ x_t + \nu_t & \text{otherwise.} \end{cases}$$

Scenario II – A only unknown:

$$u_t \leftarrow \begin{cases} K_t x_t & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \\ K_\circ x_t & \text{otherwise.} \end{cases}$$

Scenario III – B only unknown:

$$u_t \leftarrow \begin{cases} K_t x_t & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \text{ and } \lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq \sqrt{t} \\ K_\circ x_t + \zeta_t & \text{otherwise.} \end{cases}$$

end

The above theorem states that the expected regret is $\widetilde{O}((d_x + d_u)\sqrt{d_x T})$. It is worth noting that the regret upper bounds match the lower bound derived by Simchowitz and Foster (2020) when $d_x \leq d_u$. When $\mathcal{T} = \{2^k, k \in \mathbb{N}\}$, we can improve our upper bound and show:

$$\mathbb{E}[R_T(\pi)] \lesssim \sigma^2 C_B^2 \|P_\star\|^{5.25} \sqrt{d_x} d_u \sqrt{T} \log(T) + \sigma^2 \|P\|^{9.5} d_x^2 \log(T) \text{poly}(\sigma, d_x, \mathcal{G}_\circ, C_\circ, \|P_\star\|).$$

Simchowitz and Foster (2020) prove a similar regret upper bound, but in the fixed confidence setting, i.e., with probability $1 - \delta$, for an algorithm taking δ and T as inputs. Our algorithm, CEC($\mathcal{T}$), is anytime and enjoys regret guarantees in expectation. Finally, we note that the second term in the regret upper bound of Theorem 1, $\text{poly}(\sigma, d_x, d_u, \mathcal{G}_\circ, C_\circ, \|P_\star\|)$, corresponds to the regret generated in rounds where the stabilizer is used.

The next two theorems provide regret upper bounds for CEC($\mathcal{T}$) in the remaining scenarios.

Theorem 2. *(Scenario II – A only unknown) Let the set of update times $\mathcal{T}$ satisfy (5). The regret of $\pi = \text{CEC}(\mathcal{T})$ with input $\mathcal{T}$ satisfies in Scenario II: for all $T \geq 1$,*

$$\mathbb{E}[R_T(\pi)] \leq C_1 d_x^2 \log(T) + C_2,$$

with constants $C_1 \lesssim \sigma^2 \|P_\star\|^{9.5} \log(e\sigma \mathcal{G}_\circ C_\circ \|P_\star\| d_x)^2$ and $C_2 \lesssim \text{poly}(\sigma, d_x, \mathcal{G}_\circ, C_\circ, \|P_\star\|)$.

Theorem 3. *(Scenario III – B only unknown) Let the set of update times $\mathcal{T}$ satisfy (5). Assume that $K_\star K_\star^\top \succ 0$. The regret of $\pi = \text{CEC}(\mathcal{T})$ with input $\mathcal{T}$ satisfies in Scenario III: for all $T \geq 1$,*

$$\mathbb{E}[R_T(\pi)] \leq C_1 d_x (d_x + d_u) \log(T) + C_2,$$

with $C_1 \lesssim \sigma^2 \|P_\star\|^{9.5} \mu_\star^{-2} \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| \mu_\star^{-1} C_B d_x d_u)^2$, and $C_2 \lesssim \text{poly}\left(\sigma, d_x, \mathcal{G}_\circ, C_\circ, \|P_\star\|, \mu_\star^{-1}\right)$, where we denote $\mu_\star^2 = \min(\lambda_{\min}(K_\star K_\star^\top .1))$.

The results presented in the two above theorems significantly improve those derived by Cassel et al. (2020). There, the authors devise an algorithm whose inputs include upper bounds on $\|A\|$ and $\|B\|$, and on the minimal ergodic cost. The expected regret of this algorithm is upper bounded by $O(\text{poly}(d_x, d_u) \log^2(T))$. In contrast, CEC($\mathcal{T}$) has an expected regret $O(d_x^2 \log(T))$ when B is known and $O(d_x(d_u + d_x) \log(T))$ when A is known. Note that these scalings are natural and similar to the optimal regret scalings one would typically get in stochastic bandit problems. In fact, Lai in Lai (1986) (see Section 3) establishes that the expected regret cannot be smaller than $d_x^2 \log(T)$ in Scenario II. The author is not able to present an algorithm with regret matching this lower bound (refer to Appendix B for details).

5.2 Sketch Of The Regret Analysis

We provide below a brief description of the strategy used to establish Theorems 1, 2, and 3. In this subsection, and only for simplicity, the notations $\gtrsim$ and $\lesssim$ will sometimes hide the problem dependent constants that appear in the analysis. Let π denote $\mathrm{CEC}(\mathcal{T})$.

Step 1. Regret decomposition and integration. Our strategy is to establish that for all $\delta \in (0,1)$, the following

$$\mathbb{P}\left(R_T(\pi) \gtrsim c_1\psi(T)\log(e/\delta) + c_2\mathrm{poly}(\log(e/\delta))\right) \geq c_3\mathrm{poly}(\log(e/\delta))\delta \tag{6}$$

holds with probability at most $c_3\mathrm{poly}(\log(e/\delta))\delta$, where c_1, c_2, c_3 are positive problem-dependent constants and $\psi(T)$ is the targeted regret rate (e.g., $\psi(T) = \sqrt{T}$ in Scenario I). Integrating over δ, we obtain the desired upper bound in expectation $\mathbb{E}[R_T(\pi)] \lesssim c_1\log(c_3)\psi(T) + c_2\mathrm{poly}(\log(c_3))$. In order to show (6), we define for each $\delta \in (0,1)$ a "nice" event $\mathcal{E}_\delta$ such that

$$\mathcal{E}_\delta \subseteq \{R_T(\pi) \lesssim c_1\psi(T)\log(e/\delta) + c_2\mathrm{poly}(\log(e/\delta))\}, \tag{7}$$

$$\mathbb{P}(\mathcal{E}_\delta) \geq 1 - c_3\mathrm{poly}(\log(e/\delta))\delta, \tag{8}$$

which in turn will imply (6). Next we explain the construction of the event $\mathcal{E}_\delta$ satisfying (7). The likelihood of $\mathcal{E}_\delta$ and the proof of (8) are discussed in Step 2.

To construct $\mathcal{E}_\delta$ (refer to Appendix C for details), we consider the regret generated before and after a certain time τ that will be properly defined later. The first conditions we impose on $\mathcal{E}_\delta$ and τ allow us to write the regret generated after τ as

$$R_T(\pi) - R_\tau(\pi) \lesssim \sum_{t=\tau+1}^{T} \|x_t\|^2_{P_\star(K_t)-P_\star(K_{t-1})} + \mathrm{tr}(P_\star(K_t) - P_\star) + \sigma_t^2\|P_\star(K_t)\|.$$

These conditions are that ***(i)***-***(ii)*** hold at times between $\tau \leq t \leq T$,

- ***(i)*** the algorithm only uses the certainty equivalence controller K_t;
- ***(ii)*** the controller K_t is sufficiently close to $K_\star$, so that that $\rho(A + BK_t) < 1$.

We can further prove that:

$$R_T(\pi) - R_\tau(\pi) \lesssim c_1 \sum_{t=\tau+1}^{T} \varepsilon_t^2 + \sigma_t^2 \lesssim c_1(\psi(T) - \psi(\tau)),$$

if for all $\tau \leq t < T$, we ensure that:

- ***(iii)*** $\|P_\star(K_t) - P_\star\| \lesssim \varepsilon_t^2\log(1/\delta)$ where $(\varepsilon_t)_{t\geq 1}$ is non-increasing and satisifies $\sum_{s=1}^{T}\varepsilon_s^2 \lesssim \psi(T)$,
- ***(iv)*** $\|P_\star(K_t)\| = O(1)$,
- ***(v)*** $\|x_t\|^2 = O(1)$,

A proper choice of τ and condition ***(i)*** will allow us to upper bound the regret up to τ. Indeed, if ***(i)*** holds for $t \geq \tau$, we are using the certainty equivalence controller after τ. In view of the design of our algorithm, we must have $\sum_{s=0}^{\tau}\|x_\tau\|^2 \leq \sigma^2 d_x f(s)$, and that $\max_{0\leq s\leq\tau}\|K_s\|^2 \leq h(\tau)$, which in turn implies that $R_\tau(\pi) \lesssim c_2'\mathrm{poly}(\tau)$ for some problem dependent constant $c_2' > 0$. Therefore, $\{\forall t \geq \tau :$ ***(i)*** **-** ***(v)*** hold$\} \subseteq \{R_T(\pi) \lesssim c_1\psi(T) + \mathrm{poly}(\tau)\}$. Now, if we choose $\tau \geq c_2''\mathrm{poly}(\log(e/\delta))$ for some appropriately chosen constant $c_2'' > 0$,

$$\mathcal{E}_\delta = \{\forall t \geq c_2''\mathrm{poly}(\log(e/\delta)) : \textbf{\textit{(i)}} \textbf{ - } \textbf{\textit{(v)}} \text{ hold}\}, \tag{9}$$

then we have the desired set inclusion $\mathcal{E}_\delta \subseteq \{R_T(\pi) \lesssim c_1\psi(T)\log(e/\delta) + c_2\mathrm{poly}(\log(e/\delta))\}$. These arguments are made precise in Appendix C for each of the scenarios I, II, and III. The regret decomposition is stated in Lemma 1, as for the integration of the high probability bound, we refer the reader to Lemma 2.

Step 2. Nice event likelihood. It remains to establish (8). The proof relies on several important ingredients. Note that the conditions ***(i)***, ***(ii)*** and ***(iii)*** that the event $\mathcal{E}_\delta$ must satisfy concern the fact that K_t is played after τ and our ability to control $\|B(K_t - K_\star)\|$ and $\|P_\star(K_t) - P_\star\|$. We show that these three conditions will be satisfied if the error of our LSE (A_t, B_t) of (A, B) is small enough. To this aim, we use the perturbation bounds derived in Proposition 16. These bounds allow us to bound $\|B(K_t - K_\star)\|$ and $\|P_\star(K_t) - P_\star\|$ as a function of $\max(\|A_t - A\|^2, \|B_t - B\|^2)$. Observe that if $\|B(K_t - K_\star)\|$ is small, then playing K_t will stabilize the system, and we will keep using K_t, which leads to the condition ***(i)*** of $\mathcal{E}_\delta$. The perturbation bounds directly control $\|B(K_t - K_\star)\|$ and $\|P_\star(K_t) - P_\star\|$, and yield the conditions ***(ii)*** and ***(iii)*** of $\mathcal{E}_\delta$. In summary, we can establish (8) provided that we are able to control $\max(\|A_t - A\|^2, \|B_t - B\|^2)$. More precisely, we just need to prove the following probabilistic statement:

$$\forall t \gtrsim \log(e/\delta), \quad \max(\|A_t - A\|^2, \|B_t - B\|^2) \lesssim \frac{\log(t)}{t^{1/2}}\log(e/\delta), \tag{10}$$

holds with probability at least $1 - \delta$.

The results regarding the event $\mathcal{E}_\delta$ are established in Appendix D. Appendix E is devoted to proving that the algorithm eventually commits to the certainty equivalence controller K_t. The analysis of LSE is presented in Appendix F. Results about the perturbations bounds for Riccati equations are stated in Appendix I.

Step 3. Performance of the LSE under varying control. In this last and most interesting step, we prove (10). It is well established, see e.g. Mania et al. (2019) that the error of the LSE (A_t, B_t) heavily depends on the spectral properties of what we refer to as the cumulative covariates matrix. This matrix is defined as $\sum_{s=0}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top$ for Scenario I, $\sum_{s=0}^{t-1} x_s x_s^\top$ for Scenario II, and $\sum_{s=0}^{t-1} u_s^\top u_s^\top$ for Scenario III. We show that for example in Scenario I, the critical condition to obtain (10) is that:

$$\forall t \gtrsim \log(e/\delta), \quad \lambda_{\min}\left(\sum_{s=1}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right) \gtrsim t^{1/2} \quad (11)$$

holds with probability at least $1-\delta$. Establishing (11) is one of the main technical contribution of this paper, and is detailed in the next section.

6 SPECTRUM OF THE CUMULATIVE COVARIATES MATRIX

As explained in the previous section, a critical step in the analysis of the regret of CEC($\mathcal{T}$) is to characterize the performance of the LSE even if the underlying controller evolves over time. To this aim, we need to control the spectrum of the cumulative covariates matrix. We present a new decomposition of this matrix, and show how the decomposition leads to concentration results on its smallest eigenvalue. We believe that our method is of independent general interest. We apply it to the analysis of the cumulative covariates matrix in Scenario I. Refer to Appendix G for details, and for the treatment of the two other scenarios. In this section, we denote by $\widetilde{K}_t$ the controller used by CEC($\mathcal{T}$) at time t, i.e., either $K_\circ$ or K_t. Again, we hide problem dependent constants in the notations $\lesssim$ and $\gtrsim$.

6.1 A Generic Recipe: Decomposition And Concentration

We sketch here a method to study the smallest eigenvalue of a random random matrix of the form $\sum_{s=1}^t y_s y_s^\top$ where $y_s = z_s + M_s \xi_s$ and where $(z_s, M_s, \xi_s)_{s\geq 1}$ is a stochastic process such that ξ_s is independent of $(z_1, \dots, z_s)$ and $(M_1, \dots, M_s)$ for all $s \geq 1$. This model covers the cumulative covariates matrices obtained in the various scenarios. Indeed, for Scenario I, we have $y_s = \begin{bmatrix} x_s \\ u_s \end{bmatrix} = z_s + M_s \xi_s$, with

$$z_s = \begin{bmatrix} Ax_{s-1} + Bu_{s-1} \\ \widetilde{K}_s(Ax_{s-1} + Bu_{s-1}) \end{bmatrix}, \quad M_s = \begin{bmatrix} I_{d_x} & O \\ \widetilde{K}_s & I_{d_u} \end{bmatrix},$$
$$\xi_s = \begin{bmatrix} \eta_{s-1} \\ \nu_s \end{bmatrix}.$$

For Scenario II, we have $y_s = x_{s+1} = z_s + M_s \xi_s$, with :

$$z_s = (A + B\widetilde{K}_s)x_s, \quad M_s = I_{d_x}, \quad \xi_s = \eta_s.$$

For Scenario III, we have $y_s = u_s = z_s + M_s \xi_s$, with:

$$z_s = \widetilde{K}_s(Ax_{s-1} + Bu_{s-1}), \quad M_s = \widetilde{K}_s,$$
$$\xi_s = \widetilde{K}_s \eta_{s-1} + 1_{\{\widetilde{K}_s = K_\circ\}} \zeta_s.$$

Next, we claim that for some $\alpha > 0$, the smallest eigenvalue of $\sum_{s=1}^t y_s y_s^\top$ grows at least as t^α as t grows large when **(C1)** $\lambda_{\min}(\sum_{s=1}^t M_s M_s^\top)$ is growing at least as t^α and **(C2)** $\lambda_{\max}\left(\sum_{s=1}^t z_s z_s^\top\right)$ grows at most polynomially in t. As a consequence of this claim, in all scenarios, to complete Step 3 of the regret analysis, we just need to verify that the conditions **(C1)** and **(C2)** hold. This verification is explained in Scenario I in the next subsection. For complete statements and proofs, refer to Appendix G.

The first step towards our claim is the easiest but perhaps the most insightful: it consists in applying Lemma 10 in Appendix G to show that[4] for all $V \succ 0$,

$$\sum_{s=1}^t y_s y_s^\top \succeq \sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - V$$
$$\underbrace{- \left(\sum_{s=1}^t z_s (M_s\xi_s)^\top\right)^\top (V_t + V)^{-1} \left(\sum_{s=1}^t z_s (M_s\xi_s)^\top\right)}_{(\star)}.$$

where $V_t = \sum_{s=1}^t z_s z_s^\top$. The second step consists in observing that the second term $(\star)$ in the above inequality is a self-normalized matrix valued process (see for example Abbasi-yadkori et al. (2011), or Proposition 9). Concentration results for such processes lead to

$$\left\| (V_t + V)^{-1/2} \left(\sum_{s=1}^t z_s (M_s\xi_s)^\top\right) \right\|^2 \lesssim \log\left(\lambda_{\max}(V_t)\right)$$

with high probability, provided the matrices $(M_s)_{s\geq 1}$ are bounded. In the last step, we derive a concentration inequality for the matrix $\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top$ (see Proposition 8). It concentrates around $\sum_{s=1}^t M_s M_s^\top$.

In summary, we have proved that under condition **(C2)**, $\lambda_{\min}(\sum_{s=1}^t y_s y_s^\top)$ scales at least as $\lambda_{\min}(\sum_{s=1}^t M_s M_s^\top)$, which combined with **(C1)** provides the desired claim.

[4] Here we mean lower bound in the sense of the Lwner partial order over symmetric matrices.

6.2 The Recipe At Work In Scenario I

We first establish a weak growth rate for $\lambda_{\min}(\sum_{s=1}^{t} y_s y_s^\top)$ of order $t^{1/4}$ where we have a priori no information on the boundedness of the matrix sequence $(M_s)_{s\geq 1}$ (see statement (12) Theorem 4). A consequence of this first result is that LSE is consistent and therefore (A_t, B_t) will eventually be sufficiently close to (A, B) (See Appendix F). Using the perturbation bounds of Proposition 16 (See Appendix I), we can then guarantee that eventually the sequence of $(M_s)_{s\geq 1}$ will become uniformly bounded over time w.h.p.. Using this result, we show that the growth rate of $\lambda_{\min}(\sum_{s=1}^{t} y_s y_s^\top)$ may be refined to an order of $t^{1/2}$ (see the statement (13) of Theorem 4).

Theorem 4 (*Informal*). *Under CEC($\mathcal{T}$), for all $\delta \in (0,1)$ we have for all $t \gtrsim \log(e/\delta)$,*

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \gtrsim t^{1/4}\right) \geq 1-\delta. \quad (12)$$

Furthermore, provided we can guarantee that $\forall t \gtrsim \log(e/\delta)$, $\mathbb{P}(\|\widetilde{K}_t - K_\star\| \leq C_K) \geq 1-\delta$, then for all $\delta \in (0,1)$, we have for all $t \gtrsim \log(e/\delta)$,

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \gtrsim t^{1/2}\right) \geq 1-\delta. \quad (13)$$

The proof of Theorem 4 relies on showing that the conditions **(C1)** and **(C2)** hold. We can start by establishing **(C2)** since its proof is common to both statements (12) and (13). We can observe that

$$\lambda_{\max}\left(\sum_{s=1}^{t} z_s z_s^\top\right) \lesssim h(t)\left(\sum_{s=1}^{t}\|x_s\|^2 + \sum_{s=1}^{t}\|\eta_{s-1}\|^2\right),$$

where we use that $Ax_{s-1} + Bu_{s-1} = x_s - \eta_{s-1}$, and $\|\widetilde{K}_s\|^2 \leq h(s)$ for all $s \geq 1$. We can deduce, via a matrix concentration argument (See Proposition 8 in Appendix G), that $\sum_{s=1}^{t}\|\eta_s\|^2 \lesssim t$ w.h.p.. We can also establish that $\sum_{s=1}^{t}\|x_s\|^2$ does not grow more than a polynomial of order $g(t)h(t)$ w.h.p. under CEC($\mathcal{T}$) (see Proposition 15 in Appendix H). Thus, condition **(C2)** is satisfied.

Establishing **(C1)** is slightly more involved especially for the statement (13), but essentially we can prove that $\lambda_{\min}(\sum_{s=1}^{t}(M_s\xi_s)(M_s\xi_s)^\top) \gtrsim t^{1/4}$ w.h.p., and provided we can guarantee that for all $t \geq \log(e/\delta)$, $\mathbb{P}(\|\widetilde{K}_t - K_\star\| \leq C_K) \geq 1-\delta$, then $\lambda_{\min}(\sum_{s=1}^{t}(M_s\xi_s)(M_s\xi_s)^\top) \gtrsim t^{1/2}$ w.h.p. for $t \gtrsim \log(e/\delta)$. We refer the reader to Appendix G for a detailed proof of these claims. At a high level these results follow because of the special structure of the sequence of matrices $(M_s)_{s\geq 1}$ and the independence between the sequences $(\eta_s)_{s\geq 0}$ and $(\nu_s)_{s\geq 0}$.

The precise statements of Theorem 4 and its proof are deferred to Appendix G.

7 EXPERIMENTS

To illustrate the performance of CEC($\mathcal{T}$), we run the algorithm on the following simple example, proposed by Abeille and Lazaric (2020) (refer to Appendix A for details):

$$A = \begin{pmatrix}1.01 & 0.01\\0.01 & 0.5\end{pmatrix}, \qquad B = Q = R = I_2.$$

The initially known stabilizer is $K_\circ$ and the unknown optimal controller $K_\star$ are given by:

$$K_\circ \approx \begin{pmatrix}-0.27 & -0.01\\-0.01 & -0.13\end{pmatrix}, \quad K_\star \approx \begin{pmatrix}-0.63 & -0.007\\-0.007 & -0.27\end{pmatrix}.$$

Abeille and Lazaric (2020) compare the regrets of their algorithm LagLQ and of CECCE (Simchowitz and Foster, 2020) in Scenario I when (A, B) are unknown. In Figure 3 of Abeille and Lazaric (2020), the regret of these two algorithms are plotted *after* the initialization phase required by both algorithms and where a stabilizing controller (with added noise) is used. This initial stabilizing controller is not described by Abeille and Lazaric (2020). But since the initialization phase lasts for 2.10^4 rounds, and if $K_\circ$ as defined above was used, it would generate an expected regret approximately equal to $2.10^4 \times (\mathrm{tr}(P_\star(K_\circ)) - \mathrm{tr}P_\star(K_\star)) \approx 17.10^3$. For LagLQ and CECCE, the initialization phase is required so that the algorithm starts with a set of stabilizing controllers close to $K_\star$ to ensure that the state remains bounded when using a controller in this set. The use of hysteresis switching in CEC($\mathcal{T}$) allows us to remove the initialization phase and the associated regret, which on this example confers to CEC($\mathcal{T}$) much lower regrets. In Figure 1, we plot the expected regret of CEC($\mathcal{T}$) averaged over 100 runs for the three scenarios (we do not compare with LagLQ and CECCE, as their regrets would be very large at time 0 – after the initialization phase, but we do expect that LagLQ will asymptotically peform better because of its improved exploration subroutine that translates to better problem dependent constants in the regret scaling).

These curves clearly illustrate the regret scalings in the various scenarios: $\sqrt{T}$ when (A, B) is unknown, and $\log(T)$ when either A or B is known. We present additional experiments in Appendix A, and in particular, quantify the gain in not using a doubling trick, but rather updating the estimated optimal controller more frequently.

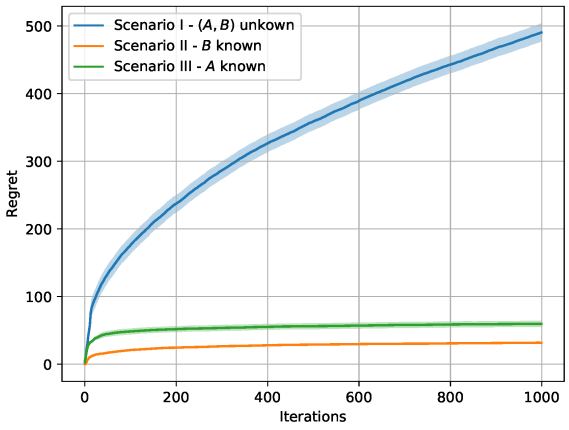


Figure 1: Regret vs. time of CEC($\mathcal{T}$) averaged over 100 runs (shaded areas correspond to the standard error of the mean (SEM)).

8 CONCLUSION

In this paper, we have designed CEC($\mathcal{T}$) a simple certainty equivalence-based algorithm for the online LQR problem, and derived upper bounds on its expected regret in the three envisioned scenarios (I: (A, B) unknown, II: A only unknown, III: B only unknown). The upper bounds have an optimal scaling in the time horizon T in all scenarios and in the dimensions of the state and control input vectors at least in Scenarios I and II (we believe that the dependence in the dimensions is also optimal in Scenario III). Importantly, CEC($\mathcal{T}$) allows for the estimates of (A, B) and the corresponding certainty-equivalent control to be continuously updated as new data is observed. Studying the performance of such a continuously evolving control is the main technical challenge solved in the paper.

Many interesting questions remain open. CEC($\mathcal{T}$) exploits a stabilizer when needed, and we proved that the expected regret generated in rounds where the stabilizer is used is finite. Does it mean that we can get rid of the stabilizer? Another interesting research direction is to investigate whether our approach and results extend to LQG systems where the decision maker receives noisy measurements of the state.

Acknowledgements

This research was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

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.

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *NIPS*, volume 11, pages 2312–2320, 2011.

Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

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

Marc Abeille and Alessandro Lazaric. Efficient optimistic exploration in linear-quadratic regulators via lagrangian relaxation. In Hal Daum III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 23–31. PMLR, 13–18 Jul 2020.

Lilian Besson and Emilie Kaufmann. What Doubling Tricks Can and Can't Do for Multi-Armed Bandits. working paper or preprint, Feb 2018.

Asaf Cassel, Alon Cohen, and Tomer Koren. Logarithmic regret for learning linear quadratic regulators efficiently. In *International Conference on Machine Learning*, pages 1328–1337. PMLR, 2020.

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

Sarah Dean, Horia Mania, Nikolai Matni, Benjamin Recht, and Stephen Tu. On the sample complexity of the linear quadratic regulator. *Foundations of Computational Mathematics*, pages 1–47, 2019.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite time analysis of optimal adaptive policies for linear-quadratic systems. *CoRR*, abs/1711.07230, 2017.

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

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite-time adaptive stabilization of linear systems. *IEEE Trans. Autom.*

Control., 64(8):3498–3505, 2019. doi: 10.1109/TAC.2018.2883241.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1467–1476, Stockholmsmssan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

Aurelien Garivier, Tor Lattimore, and Emilie Kaufmann. On explore-then-commit strategies. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

Daniel Hsu, Sham Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab.*, 17:6 pp., 2012. doi: 10.1214/ECP.v17-2079.

Michail M Konstantinov, P Hr Petkov, and Nicolai D Christov. Perturbation analysis of the discrete riccati equation. *Kybernetika*, 29(1):18–29, 1993.

Vladimír Kučera. The discrete riccati equation of optimal control. *Kybernetika*, 8(5):430–447, 1972.

R. P. Kumar. A survey of some results in stochastic adaptive control. *Siam Journal on Control and Optimization*, 1985.

Tze Leung Lai. Iterated least squares in multiperiod control. *Advances in Applied Mathematics*, 3(1): 50–73, 1982. ISSN 0196-8858.

Tze Leung Lai. Asymptotically efficient adaptive control in stochastic regression models. *Advances in Applied Mathematics*, 7(1):23–45, 1986. ISSN 0196-8858.

Tze Leung Lai and Herbert Robbins. Adaptive Design and Stochastic Approximation. *The Annals of Statistics*, 7(6):1196 – 1221, 1979.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–2, 1985.

Tze Leung Lai and Ching Zong Wei. Asymptotically efficient self-tuning regulators. *SIAM Journal on Control and Optimization*, 25(2):466–481, 1987. doi: 10.1137/0325026.

Sahin Lale, Kamyar Azizzadenesheli, Babak Hassibi, and Anima Anandkumar. Explore more and improve regret in linear quadratic regulators. *arXiv preprint arXiv:2007.12291*, 2020.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Horia Mania, Stephen Tu, and Benjamin Recht. Certainty equivalence is efficient for linear quadratic control. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Nikolai Matni, Aalexandre Proutiere, Anders Rantzer, and Stephen Tu. From self-tuning regulators to reinforcement learning and back again. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 3724–3740, 2019. doi: 10.1109/CDC40024.2019.9029916.

Yi Ouyang, Mukul Gagrani, and Rahul Jain. Control of unknown linear systems with thompson sampling. In *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1198–1205. IEEE, 2017.

Anders Rantzer. Concentration bounds for single parameter adaptive control. In *Proceedings of American Control Conference*, volume 2018-June, pages 1862–1866, United States, jun 2018. Institute of Electrical and Electronics Engineers Inc. ISBN 978-153865428-6. doi: 10.23919/ACC.2018.8431891. American Control Conference 2018, ACC 2018 ; Conference date: 27-06-2018 Through 29-06-2018.

Karl Johan Åström and Björn Wittenmark. On self tuning regulators. *Automatica*, 9(2):185?199, mar 1973. ISSN 0005-1098.

Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Input perturbations for adaptive control and learning. *Automatica*, 117:108950, 2020. ISSN 0005-1098.

Max Simchowitz and Dylan Foster. Naive exploration is optimal for online LQR. In Hal Daum III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 8937–8948. PMLR, 13–18 Jul 2020.

William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933. ISSN 00063444.

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

Ingvar Ziemann and Henrik Sandberg. On uninformative optimal policies in adaptive lqr with unknown b-matrix. In *Learning for Dynamics and Control*, pages 213–226. PMLR, 2021.

Supplementary Material: Minimal Expected Regret in Linear Quadratic Control

Contents

Notations and Assumptions

- $f(x) \gtrsim g(x)$ means there exists an universal constant $c > 0$ such that $f(x) \geq cg(x)$.
- $f(x) \lesssim g(x)$ means there exists an universal constant $c > 0$ such that $f(x) \leq cg(x)$.
- $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue.
- $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue.
- $\|\cdot\|$ denotes operator norm for matrices or ℓ_2-norm for vectors.
- $\|\cdot\|_F$ denotes Frobeinus norm.
- $\|x\|_M = \sqrt{x^\top M x}$ for any vectors x.
- $\|a_{1:t}\|_\infty = \max_{1\leq s\leq t} |a_s|$ where $(a_s)_{s\geq 1}$ is a scalar valued sequence.
- $\|a_{1:t}\|_2 = \sqrt{\sum_{s=1}^t |a_s|^2}$ where $(a_s)_{s\geq 1}$ is a scalar valued sequence.
- d_x and d_u denote respectively the dimension of the state/input space.
- $d = d_x + d_u$.
- γ is a postive constant used to define $f(t) = t^{1+\gamma/2}, g(t) = t^{1+\gamma}$ and $h(t) = t^\gamma$.
- $\gamma_\star = \max\{1, \gamma\}$.
- $C_\circ = \max(\|A\|, \|B\|, \|BK_\circ\|, \|K_\circ\|, 1)$.
- $\mathcal{G}_M = \sum_{s=0}^\infty \|M^s\|$.
- $\mathcal{G}_M(\varepsilon) = \sup\left\{\sum_{s=0}^\infty \|\prod_{k=0}^s (M + \Delta_k)\| : \Delta :\ (\Delta_t)_{t\geq 1}, \sup_{t\geq 0} \|\Delta_t\| \leq \varepsilon\right\}$.
- $P(A, B)$ solution to the DARE corresponding to the LQR problem (A, B, Q, R).
- $K(A, B)$ optimal gain matrix corresponding to the LQR problem (A, B, Q, R).
- $\mathcal{L}(M, N)$ is the solution to the Discrete Lyapunov equation $X = M^\top X M + N$.
- $P(A, B, K) = \mathcal{L}(A + BK, Q + K^\top R K)$.

Shorthands for the true parameters (A, B).

- $C_B = \max(\|B\|, 1)$.
- $C_\circ = \max(\|K_\circ\|, 1)$.
- $\mathcal{G}_\circ = \mathcal{G}_{A+BK_\circ}$.
- $\mathcal{G}_\star = \mathcal{G}_{A+BK_\star}$.
- $\mathcal{G}_\star(\varepsilon) = \mathcal{G}_{A+BK_\star}(\varepsilon)$.
- $P_\star = P(A, B)$.
- $K_\star = K(A, B)$.
- $P_\star(K) = P(A, B, K)$.
- $\mu_\star = \min(\sqrt{\lambda_{\min}(K_\star K_\star^\top)}, 1)$.

Assumptions.

- We assume without loss of generality that $Q \succ I_{d_x}$ and $R = I_{d_u}$. These can be enforced by a change of basis of the state and input spaces, and rescaling the dynamics (see e.g., Simchowitz and Foster (2020)).
- The noise sequence $(\eta_t)_{t\geq 1}$ is assumed to be i.i.d. zero-mean, isotropic and σ^2-sub-gaussian random vectors. Isotropy here is assumed for simplicity and is without loss of generality. Observe that isotropy implies $4\sigma^2 \geq 1$.
- In all the envisioned scenarios, we assume access to a stabilizing controller. That is we know $K_\circ \in \mathbb{R}^{d_u \times d_x}$ such that $\rho(A + BK_\circ) < 1$.
- In scenario II – (A known), we assume that $\mu_\star > 0$.

A EXPERIMENTS

A.1 The LQR System

The LQR system used in all our experiments was suggested by Abeille and Lazaric (2020), and is described by

$$A = \begin{pmatrix} 1.01 & 0.01 \\ 0.01 & 0.5 \end{pmatrix}, \qquad B = Q = R = I_2.$$

Note that when the control inputs are set to zero, the resulting system is marginally stable $\lambda_{\max}(A) \approx 1.01 > 1$. The initially known stabilizer $K_\circ$ and unkown optimal controller $K_\star$ are given by

$$K_\circ \approx \begin{pmatrix} -0.27 & -0.01 \\ -0.01 & -0.13 \end{pmatrix}, \qquad K_\star \approx \begin{pmatrix} -0.63 & -0.007 \\ -0.007 & -0.27 \end{pmatrix}.$$

We further have the ergodic cost under $K_\circ$, and optimal cost

$$\mathrm{tr}(P_\star(K_\circ)) \approx 3.63, \qquad \mathrm{tr}(P_\star) \approx 2.78.$$

The noise sequence $(\eta_t)_{t\geq 1}$ is a sequence of i.i.d. random vectors distributed as $\mathcal{N}(0, I_2)$. Finally, the experiments we run here mainly concern Scenario I when A and B are unknown.

A.2 Sensitivity Of CEC($\mathcal{T}$) To The Doubling Trick With Or Without Data Forgetting

The experiments presented here have the following objectives.

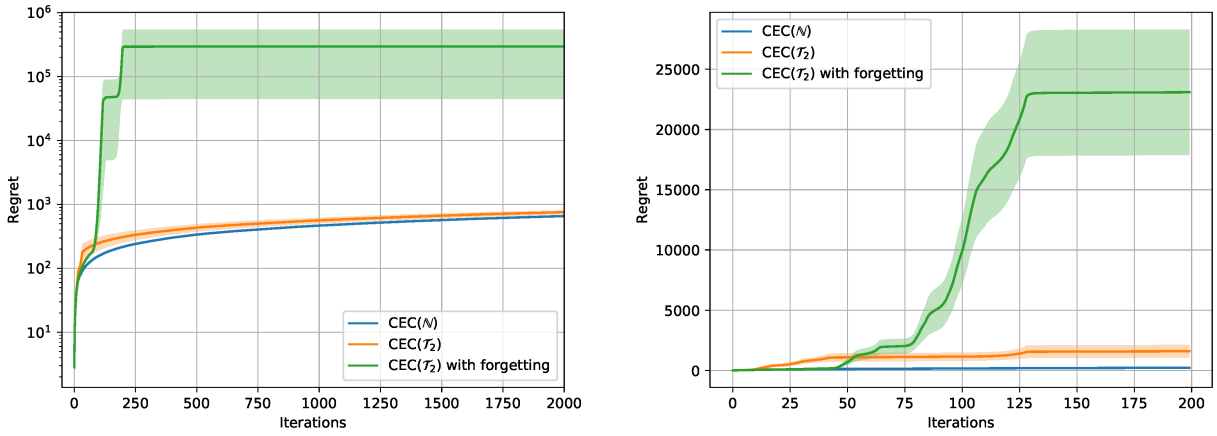
1. Impact of sparse updates. We first wish to assess the impact of rarefying the control policy updates in CEC($\mathcal{T}$), i.e., to quantify the impact of different choices of $\mathcal{T}$. To this aim, we compare the regrets of CEC($\mathbb{N}$) (updates every step) and of CEC($\mathcal{T}_2$) where $\mathcal{T}_2 = \{2^t : t \in \mathbb{N}^\star\}$.

2. Impact of data forgetting. In most existing algorithms, the control policy is updated in the steps of $\mathcal{T}_2$, i.e., at the beginning of phases of increasing durations (phase k lasts for 2^k steps). In addition, to further simplify the analysis of the LSE, the control policy applied in phase k only depends on the data gathered during the previous phase $k-1$. This is the case for example in the algorithm Certainty Equivalence Control with Continual Exploration (CECCE) proposed by Simchowitz and Foster (2020). In turn, this allows to analyze the LSE under a *fixed* control policy. We refer to this simplification as the *data forgetting* trick; note that half of the data gathered is discarded using this trick. Although applying such principle does not affect the asymptotic scaling of the regret (Simchowitz and Foster, 2020), it must have an impact on the practical performance of the algorithm. We wish to quantify this impact. To this aim, we consider a variant of CEC($\mathcal{T}_2$) where the algorithm applies the data forgetting trick. More precisely, for all $k \in \mathbb{N}^\star$, at times $2^k, \dots, 2^{k+1} - 1$, the algorithm either uses $K_\circ$ or a fixed estimate K_k of K, where K_k is constructed using only samples gathered at times $2^{k-1}, \dots, 2^k - 1$. We refer to this algorithm as CEC($\mathcal{T}_2$) with forgetting.

All the algorithms CEC($\mathbb{N}$), CEC($\mathcal{T}_2$), and CEC($\mathcal{T}_2$) with forgetting are using $f(t) = (t+1)^2$, $g(t) = (t+1)^3$ (hysteresis switching), and $h(t) = t + 1$.

The results are presented in Figure 2. We observe that CEC($\mathbb{N}$), CEC($\mathcal{T}_2$) both have a regret scaling roughly as $\sqrt{T}$ as predicted by Theorem 1. Although we do not provide a theoretical guarantee on CEC($\mathcal{T}_2$) with forgetting, we believe that the regret of this algorithm is also of the order $\sqrt{T}$. However CEC($\mathcal{T}_2$) with forgetting has a much higher regret than CEC($\mathbb{N}$) (note that the y-axis in Fig. 2 (a) are in log scale). In addition, the regret of CEC($\mathcal{T}_2$) with forgetting suffers from a very high variance, visible even after averaging over $N = 3000$ runs. This high variance can be explained by the fact that, during each epoch, the estimated controller is not updated. If it happens that such a controller is bad, then the algorithm has to wait until $\sum_{s=0}^{t} \|x_t\|^2 > d_x g(t) = d_x(t+1)^3$ to revert back to the stabilizing controller. CEC($\mathbb{N}$) is the least affected by the choice of g and exhibits the best performance.

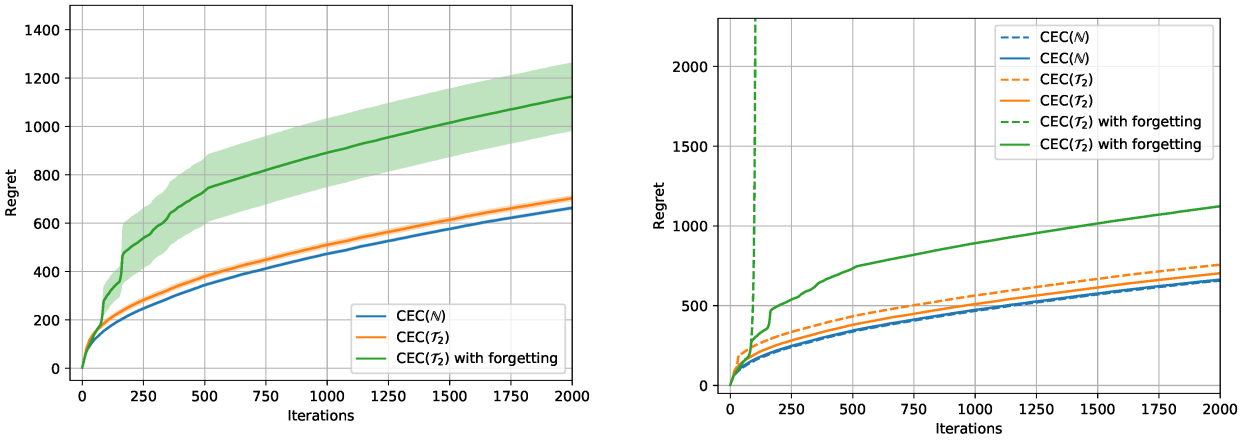
The choice of the functions f and g that control the hysteresis switching has been so far rather arbitrary. We run the experiment again with $f(t) = \log(t)(t+1)$ and $g(t) = \log(t)^2(t+1)$. The results are presented in Figure 3. As we can see, the initial regret of CEC($\mathcal{T}_2$) and CEC($\mathcal{T}_2$) with forgetting drastically improve. The variance issue still remains with forgetting, however it is now less prominent. Note that the performance of CEC($\mathbb{N}$) remained unchanged, which suggests that algorithms based on continuous updates are more robust to the choice of f, and g.



(a) Averaged over $N = 100$ runs.

(b) Averaged over $N = 3000$ runs.

Figure 2: Regret vs. time averaged over N runs of CEC($\mathbb{N}$), CEC($\mathcal{T}_2$), and CEC($\mathcal{T}_2$) with forgetting (shaded areas correspond to the standard error of the mean (SEM)).



(a) Hysteresis switching with $f(t) \approx t\log(t)$ and $g(t) \approx t\log(t)^2$

(b) Shaded areas are not plotted

Figure 3: Regret vs. time averaged over 100 runs of CEC($\mathbb{N}$), CEC($\mathcal{T}_2$), and CEC($\mathcal{T}_2$) with forgetting (when plotted, shaded areas correspond to the standard error of the mean (SEM)). The solid (resp. dashed) curves correspond to using hysteresis switching with $f(t) = (t+1)^2$ and $g(t) = (t+1)^3$ (resp. $f(t) = \log(t+1)(t+1)$ and $g(t) = \log(t+1)^2(t+1)$).

A.3 CEC($\mathcal{T}$) Efficient Use Of The Initial Stabilizer

In this experiment, we wish to compare the performance of our algorithm to that of CECCE (Certainty Equivalence Control with Continual Exploration) proposed by Simchowitz and Foster (2020). CECCE requires as an input a confidence level δ and a stabilizing controller. We choose $\delta = 0.05$ and use $K_\circ$ as an initial stabilizer. The comparison to other algorithms is left for future work. These algorithms are not proven to be optimal, often require knowledge of upper bounds on unknown quantities, are not anytime, or require knowledge of a stabilizing controller that is sufficiently close to the optimal controller (see Appendix B for more details). Refer to Abeille and Lazaric (2020) for a comparison of LagLQ (the algorithm proposed in Abeille and Lazaric (2020)) to CECCE. LagLQ takes as input the knowledge of a stabilizing controller sufficiently close to the optimal controller, which in turn, requires an initialization phase to identify such controller (such initialization phase is not accounted for

in Abeille and Lazaric (2020)).

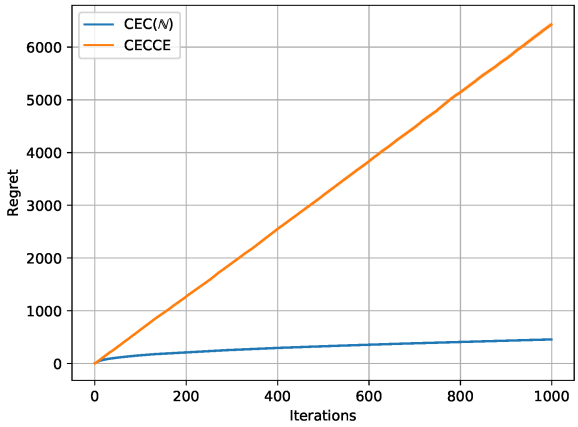


Figure 4: Regret vs. time of CEC($\mathcal{T}$) averaged over 100 runs (shaded areas correspond to the standard error of the mean (SEM)).

We present the results in Figure 4. We observe that CECCE suffers a linear regret over the considered horizon $T = 1000$, which indicates that the algorithm is still in the initialization phase. In fact, it was also highlighted by Abeille and Lazaric (2020) that CECCE requires an initial phase of length $T_0 \approx 10^8$. This suggests that the design of the initialisation phase of CECCE is too conservative, and hence that the algorithm suffers a huge initial regret cost. This conservativeness can be traced back the perturbation bounds on DARE provided by Simchowitz and Foster (2020), and upon which CECCE is based. On the other hand, the results suggest that CEC($\mathbb{N}$) has a strikingly shorter initialization phase, and wisely uses the stabilizing controller. This property can be explained by the hysteresis switching mechanism upon which CEC($\mathcal{T}$) relies which gives much more flexibility in choosing under what worst case scenario the algorithm should revert back to the stabilizing controller. In summary, CEC($\mathbb{N}$) performs better than CECCE, at least non-asymptotically.

B RELATED WORK

In this section, we describe existing learning algorithms for the online LQR problem. These algorithms may be roughly categorized into two classes. In the first class, we find algorithms based on slightly perturbing the so-called self-tuning regulators (Åström and Wittenmark, 1973), as those developed in second half of the 20th century in the control community, see Kumar (1985) for a survey and Matni et al. (2019) for a more recent discussion. The second class of algorithms applies the Optimism in Front of Uncertainty (OFU) principle, extensively used to devise regret optimal algorithms in stochastic bandit problems (Lai and Robbins, 1985; Lattimore and Szepesvári, 2020). Before describing these two classes of algorithm in more detail, we start by discussing how algorithms may differ in terms of regret guarantees, design principles, and the assumptions made towards their analysis.

One may also consider a third class of algorithms one that we may refer to as Thompson Sampling based algorithms (Thompson, 1933). These algorithms assume that the true parameters of the system are samples from some prior distribution, as such they follow a bayesian approach in their design and analysis. We will not discuss these algorithm since they often enjoy guarantees these are different than ours and the ones we compare with (e.g., Ouyang et al. (2017) analyse the bayesian regret, Abeille and Lazaric (2017) have sub-optimal regret rates in T, and Faradonbeh et al. (2018) provide almost sure guarantees).

B.1 Types Of Guarantees, Algorithm Design, And Assumptions

Regret guarantees. We may assess the performance of an algorithm by establishing various kinds of regret guarantees. Most often, the regret guarantees are in the *fixed confidence setting* only, in the following sense. The regret R_T^π of an algorithm π up to time T satisfies a probabilistic guarantee of the form: $\mathbb{P}\left(R_T^\pi \leq \psi(T)\left(\log\left(1/\delta\right)\right)^{1/\gamma}\right) \geq 1-\delta$, for some $\gamma \leq 2$ and some increasing function ψ. Typically such a guarantee is shown for a fixed confidence level (a fixed δ), as the algorithm π is most often actually parametrized by δ. As a consequence, this probabilistic guarantee cannot be integrated over δ to obtain an upper bound of the expected regret. However, as discussed in Section 2, by carefully choosing δ as a function of the time horizon T, one may modify the algorithms with fixed confidence guarantees and obtain guarantees in expectations, at the expense of an extra $\log(T)$ multiplicative factor in the regret upper bound. As far as we are aware, strictly speaking, upper bounds on the *expected* regret have been investigated by Rantzer (2018); Cassel et al. (2020) in Scenario II where A or B is known only. Finally, it is worth mentioning that early work on adaptive control have focussed on deriving *asymptotic* regret guarantees. For example, Lai (Lai, 1986; Lai and Wei, 1987) devised, for systems where control inputs induce no cost $R = 0$, algorithms whose regret satisfies $\limsup_{T\to\infty} R_T^\pi/\log(T) \leq C$ almost surely. This type of guarantee does not imply either guarantees w.h.p. as described above or guarantees in expectation.

Algorithm design. Over the last few years, we have witnessed a significant research effort towards the design of learning algorithms with regret guarantees. Most choices in this design have been made to simplify the algorithm analysis rather than to improve their performance in practice. We discuss these choices below.
(i) Doubling trick. This trick is usually applied in online optimization problems (including bandits) (Besson and Kaufmann, 2018; Lattimore and Szepesvári, 2020) to come up with algorithms with *anytime* regret guarantees. The doubling trick generally comes with a cost in terms of regret (Besson and Kaufmann, 2018). In linear quadratic control, the doubling trick is used in all recent papers to simplify the analysis, but also to reduce the computational complexity of the algorithms. It consists in splitting time into successive *phases* whose durations grow exponentially. The control policy is computed at the beginning of each phase, and is applied throughout the epoch. The analysis may then leverage the fact that the control policy is fixed within each phase. Even if recent algorithms include a doubling trick, most of them still take the time horizon T as an input, and hence are not anytime.
(ii) Explore-Then-Commit and the known time horizon and confidence level. As already mentioned, most of the recent algorithms target regret guarantees with a fixed confidence level, parametrized by δ. To this aim, they adopt an Explore-Then-Commit (ETC) strategy, namely they rely on statistical tests to decide to switch from an exploration phase to an exploitation phase. These tests require of course the knowledge of δ. Applying an ETC strategy with confidence level $(1-\delta)$ imposes some constraints on the time horizon T (it has to be greater than some decreasing function of δ so that the statistical tests end). Observe that to further simplify the design and analysis of algorithms, the time horizon is often assumed to be known in advance. It is finally interesting to note that ETC strategies are known to be sub-optimal, even in the simplest of the stochastic bandit problems (Garivier et al., 2016).

Assumptions. The set of assumptions made to design and analyze algorithms varies in the literature, which makes it hard to report and compare existing results precisely. Most existing work assume that we have access to a stabilizer. Recent attempts to remove this assumption include Lale et al. (2020). There, for example, the authors assume that the algorithm knows that the system (A, B) belongs to a set of systems (A', B') such that $\|A' + B'K_{(A',B')}\| \leq \Upsilon < 1$ and $\|[A', B']\|_F \leq S$, which in particular implies that $\|P_{(A',B')}\| \leq L$, for some constants Υ, S, and L. We will provide a description as precise as possible of the set of assumptions made in each paper reported below.

We propose an algorithm that does not take as input the time horizon T, or a certain level of confidence δ. The control policy used in the algorithm can be updated every step, but also as frequently as we wish (we may decide reduce the computational complexity of the algorithm). Our analysis provides regret guarantees in expectation in all scenarios.

B.2 Existing Algorithms

Next we describe selected recent learning algorithms. The first set of algorithms consist in slightly perturbing the control policies obtained when applying the certainty equivalence principle. The second set consists of algorithms applying the OFU principle.

Perturbed Self-Tuning Regulators

Self-tuning regulators work as follows. At any given step, they estimate the unknown matrices A and B, and apply a control policy corresponding to the optimal control obtained replacing A and B by their estimators. Unfortunately as proved by Lai (1982), self-tuning regulators may fail at converging – the certainty equivalence principle does not always hold. This is due to the fact that under these regulators, the system may not be as excited as needed to obtain precise estimators. To circumvent this difficulty, the natural idea is to introduce some noise in the control inputs, leading to what we refer to as perturbed self-tuning regulators. We list below papers applying this idea, and analyzing the resulting regret.

As far as we know, the first regret analysis of perturbed self-tuning regulators is due to Lai and co-authors in the 80's, see e.g., Lai (1986); Lai and Wei (1987). The focus is on a scenario where A and B are unknown, but where the control inputs do not contribute to the costs ($R = 0$). Lai first establishes, using the techniques developed in Lai and Robbins (1979), that even if B is known, asymptotically the regret cannot be smaller than $d_x^2 \log(T)$ when the noise process $(\eta_t)_{t\geq 0}$ is i.i.d. with distribution $\mathcal{N}(0, I_{d_x})$. More precisely, it is shown that for the best learning algorithm π, $\liminf_{T\to\infty} \frac{R_T^\pi}{\log(T)} \geq d_x^2$ almost surely. This lower bound is established using a Bayesian method, and it is easy to see that it also holds in expectation. Lai (1986) devises a perturbed self-tuning regulator, and analyzes its regret, but only under the assumption (1.7) which specifies that the state remains bounded and that the minimum eigenvalue of the cumulative covariates matrix grows linearly with time asymptotically a.s.. These assumptions are strong, and hard to remove. In Lai and Wei (1987), these assumptions are removed but for very specific systems. In both papers, the regret guarantees take the form $\limsup_{T\to\infty} R_T^\pi / \log(T) \leq C$ almost surely, which as already mentioned, does not imply guarantees in expectation.

Shirani Faradonbeh et al. (2020) study Scenario I. They propose a perturbed self-tuning regulator, referred to as Perturbed Greedy Regulator, where the variance of the noise added to the inputs decreases over time. The regulator uses a doubling trick. The authors establish that with probability $1 - \delta$, a regret is bounded by $\widetilde{\mathcal{O}}(\sqrt{T}\log(1/\delta)^4)$ for $T \geq g(\delta)$. The dependence of the upper bound in the system and its dimensions is not explicit. It is worth noting that the authors assume that the algorithm has access to a stabilizer, which according to their companion paper (Faradonbeh et al., 2019) can be learnt in finite time. The authors further assume that the system remains stable during the execution of their algorithm.

Mania et al. (2019) do not explicitly propose a perturbed self-tuning regulator with regret guarantees. However, they show that if the estimation error is sufficiently small then the resulting algorithm achieves a $\widetilde{O}(\sqrt{T})$ regret with explicit dependence on the problem dimensions d_x, d_u. Their main contribution is a perturbation bound on the solution to the Discrete Algebraic Riccati equations, an important piece of the regret analysis. They propose an alternative proof to that of Konstantinov et al. (1993) and compute explicitly the problem dependent constants.

Simchowitz and Foster (2020) propose, for Scenario I, a perturbed self-tuning regulator, that takes as input δ

and T, as well as a stabilizer. Again to simplify the analysis, a doubling trick is used. The algorithm achieves a regret of $\widetilde{O}(d_u\sqrt{d_xT\log(1/\delta)})$ with probability $1-\delta$. The authors further derive what they refer to as a *local minimax* lower bound on the expected regret. This lower bound is obtained by varying the potential system matrices (A,B) around those of the true system, and is in a sense close to a problem-specific lower bound. The lower bound is scaling as $\Omega(d_u\sqrt{d_xT\log(1/\delta)})$.

Cassel et al. (2020) present perturbed self-tuning regulators for Scenarios II and III (when A or B is known). The regulators have numerous inputs, including a stabilizer K_0 (actually a strongly stable control) and upper bounds on $\|A\|$, $\|B\|$, on the minimal ergodic cost, and that achieved under K_0. When A is known, the proposed regulator is shown to have an expected regret of order $O(\mathrm{poly}(d_x,d_u)\log^2(T))$ (where the degree of the polynomial scale in the dimensions is not precised). The regulator presented for the case when B is unknown achieves similar expected regret guarantees provided that the optimal controller $K_\star$ satisfies $K_\star K_\star^\top \succeq \mu > 0$.

OFU-based algorithms

A typical OFU-based algorithm proceeds as follows. It maintains a confidence set $\mathcal{C}$ (an ellipsoid) where the system parameters (A,B) lie with high probability. At a given step, the algorithm selects $(A',B')\in\mathcal{C}$ that minimizes the optimal cost $J_{(A',B')}$ (with sometimes an additional margin). The controller $K_{(A',B')}$ is then applied. Since updating the controller requires solving a complex optimization problem, referred to as the optimistic LQR below, this update should be done rarely (using a doubling trick).

Abbasi-Yadkori and Szepesvári (2011) present OFU-LQ, an algorithm taking as inputs a confidence level δ and T, as well as a bounded set $\mathcal{S}$ where (A,B) lies and an upper bound on $\left\|\begin{bmatrix}A & B\end{bmatrix}\right\|_F$. OFU-LQ leverages a *random* doubling trick: the controller is updated each time the determinant of the covariates matrix is doubled. This determinant roughly grows as $t^{d_x+d_u}$, and hence the successive phases have durations multiplied by $2^{1/(d_x+d_u)}>1$. OFU-LQ has a regret of order $\widetilde{\mathcal{O}}\sqrt{T\log(1/\delta)}$ with probability $1-\delta$. Here, the notation $\widetilde{\mathcal{O}}(\cdot)$ hides polynomial factors in $\log(T)$ and multiplicative constants exponentially growing in d_x+d_u. Abbasi-Yadkori and Szepesvári (2011) does not indicate how to solve the optimistic LQR, and cannot be implemented directly. The same conclusion holds for the algorithm proposed by Faradonbeh et al. (2017) (there, the authors were able to relax some assumptions on the noise and stability, but an efficient and practical implementation of the algorithm is not investigated).

A first practical implementation of OFU-based algorithms was presented by Cohen et al. (2019). The algorithm, OSLO, uses an SDP formulation to solve the optimistic LQR. It uses a somewhat random doubling trick similar to Abbasi-Yadkori and Szepesvári (2011) and requires the knowledge of δ, T, upper bounds on the norms of A,B and $P_{(A,B)}$, as well as a stabilizing controller. OSLO achieves a regret of order $\widetilde{O}((d_x+d_u)^3\sqrt{T\log(1/\delta)^4})$.

Abeille and Lazaric (2020) present an efficient algorithm to implement OFU-based algorithms proposed by Abbasi-Yadkori and Szepesvári (2011). The idea is to perform a relaxation of the optimistic LQR. The analysis requires the knowledge of δ and the horizon T. And a similar random doubling trick to that of Abbasi-Yadkori and Szepesvári (2011) is used. They obtain a regret of order $\widetilde{O}((d_u+d_x)\sqrt{d_xT}\log(1/\delta))$ with probability $1-\delta$. Note that the learner is assumed to know an initial state that is sufficiently good so that stability is maintained throughout the learning process.

Lale et al. (2020) provide another improvement on the OFU based algorithm of Abbasi-Yadkori and Szepesvári (2011). Their goal is to improve the dependence of the regret upper bound on the dimension and to remove the assumption of having access to a stabilizing controller. Doing so, the authors need to introduce other assumptions about the set of systems to which the algorithm applies: A, B and $A+BK_{(A,B)}$ have bounded norms. The regret of the proposed algorithm is with probability $1-\delta$ of order $\widetilde{O}(\mathrm{poly}(d_x+d_u)\sqrt{T\log(1/\delta)^\gamma})$ with $\gamma\geq 1$ but unspecified, and the degree $\mathrm{poly}(d)$ is also unspecified.

Summary of existing results and comparison with $\mathrm{CEC}(\mathcal{T})$

In summary, all the aforementioned algorithms use a doubling trick so that the algorithm becomes amenable to theoretical analysis (using the independence between epochs etc). Most of the algorithms are designed in the fixed confidence setting. They are variants of ETC strategies and their construction rely heavily on knowledge of the confidence level δ. Most of them also take as input the time horizon T, i.e., they are not anytime.

The assumptions made towards the regret analysis of these algorithms are not unified. Therefore it is very

hard to obtain fair comparisons between these algorithms. Except for Cassel et al. (2020), the algorithms have regret guarantees in the fixed confidence setting, i.e., with probability $1-\delta$. The best regret dependence in δ is $\sqrt{\log(1/\delta)}$, but it is not achieved by all algorithms ($\log(1/\delta)$ (Abeille and Lazaric, 2020) and $\log(1/\delta)^2$ (Shirani Faradonbeh et al., 2020; Cohen et al., 2019)).

The table below summarizes the regret guarantees achieved by the various algorithms.

Scenario I - A and B unknown			
paper	regret upper bound w.p. $1-\delta$	expected regret upper bound	required inputs
Shirani Faradonbeh et al. (2020)	$\widetilde{O}(g(d_x+d_u)\sqrt{T\log(1/\delta)^4})$	-	δ (unclear)
Simchowitz and Foster (2020)	$\widetilde{O}(d_u\sqrt{d_xT\log(1/\delta)})$	-	δ and T
Mania et al. (2019)	$\widetilde{O}(g(d_x+d_u)\sqrt{T}f(\log(1/\delta)))$	-	unkown
Lale et al. (2020)	$\widetilde{O}(\mathrm{poly}(d_x+d_u)\sqrt{T\log(1/\delta)})$	-	δ
Abbasi-Yadkori and Szepesvári (2011)	$\widetilde{O}(\sqrt{c^{d_x+d_u}T\log(1/\delta)})$	-	δ and T
Abeille and Lazaric (2020)	$\widetilde{O}((d_x+d_u)\sqrt{d_xT\log(1/\delta)^2})$	-	δ and T
Cohen et al. (2019)	$\widetilde{O}((d_x+d_u)^3\sqrt{T\log(1/\delta)^4})$	-	δ and T
this paper	-	$\widetilde{O}((d_u+d_x)\sqrt{d_xT})$	-

Scenarios II and III - A or B known			
paper	assumptions	expected regret upper bound	required inputs
Cassel et al. (2020)	A known, $\|B\|\leq M$	$O(\mathrm{poly}(d_x,d_u)\log^2(T))$	T, M (among others)
this paper	A known	$O(d_x(d_x+d_u)\log(T))$	-
Cassel et al. (2020)	B known, $\|A\|\leq M$	$O(\mathrm{poly}(d_x,d_u)\log^2(T))$	T, M (among others)
this paper	B known	$O(d_x^2\log(T))$	-

Table 1: Regret guarantees of existing algorithms. The notation $\widetilde{O}(\cdot)$ hides polynomial factors in $\log(T)$ and additive low order terms in T with potentially worse dependencies in $\log(1/\delta)$, and the constants depends on problem parameters.

C REGRET DEFINITIONS AND ANALYSIS

This appendix includes in C.1 a discussion about the definition of the regret of an adaptive control algorithm. In C.2, we provide a useful decomposition of the expected regret that will serve as the starting point of our analysis in the three scenarios. We present in C.3 the so-called integration lemma, that will helps us to derive expected regret upper bounds based on high probability bounds. The three last subsections give the proofs of our main theorems: the regret upper bound of CEC($\mathcal{T}$) in Scenario I (Theorem 1) is proved in C.4. The proof of Theorem 3 for Scenario III (A known) is given in C.5. That of Theorem 2 for Scenario II (B known) is finally presented in C.6.

C.1 Regret Definitions

Let us denote by Π the set of all possible adaptive control policies. For a poolicy $\pi \in \Pi$, $(x_1^\pi, u_1^\pi, \dots, x_t^\pi, u_t^\pi)$ is the sequence of states and control inputs generated under π. Remember that $x_0^\pi = 0 = u_0^\pi$.

We define the ergodic cost of a policy $\pi \in \Pi$ as

$$\mathcal{J}(\pi) = \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} (x_t^\pi)^\top Q(x_t^\pi) + (u_t^\pi)^\top R(u_t^\pi)\right] \qquad \textit{(ergodic cost / objective)}$$

It can be shown under suitable assumptions on (A, B, Q, R) that there exists a policy $\pi_\star \in \arg\min_{\pi\in\Pi} \mathcal{J}(\pi)$. Let $\mathcal{J}_\star = \mathcal{J}(\pi^\star)$. The optimal policy $\pi_\star$ can be found explicitly: for all $t \geq 1$, $\pi^\star$ defines the feedback control $u_t^{\pi^\star} = K_\star x_t^{\pi^\star}$. The matrix $K_\star$ can be computed by solving the Riccati equations. We have the useful identity that $P_\star = (A + BK_\star)^\top P_\star (A + BK_\star) + Q + K_\star^\top R K_\star$ where $P_\star \succ 0$ is the solution to the Ricatti equations. Furthermore, we have $\mathcal{J}_\star = \mathrm{tr}(P_\star)$.

Now, we define the regret of a policy $\pi \in \Pi$ as

$$\sum_{t=1}^{T} (x_t^\pi)^\top Q(x_t^\pi) + (u_t^\pi)^\top R(u_t^\pi) - \mathbb{E}\left[\sum_{t=1}^{T} (x_t^{\pi^\star})^\top Q(x_t^{\pi^\star}) + (u_t^{\pi^\star})^\top R(u_t^{\pi^\star})\right]. \tag{14}$$

This definition is natural as we compare the cost under π to that under $\pi^\star$, cumulated over T steps. During these T steps to compute the costs, we follow the trajectory of the system. This contrasts with the definition of regret often used in the literature:

$$R_T(\pi) = \sum_{t=1}^{T} (x_t^\pi)^\top Q(x_t^\pi) + (u_t^\pi)^\top R(u_t^\pi) - T\mathcal{J}_\star. \tag{15}$$

In fact when considering the expected regret, the above two definitions coincide up to a constant. Indeed, note that for all $t \geq 1$, $\|x_t^{\pi^\star}\|_{P_\star}^2 = \|x_{t+1}^{\pi^\star} - \eta_t\|_{P_\star}^2 + \|x_t^{\pi^\star}\|_Q^2 + \|u_t^{\pi^\star}\|_R^2$. Taking expectation (note $\mathbb{E}[\eta_t^\top X] = 0$ provided η_t is independent of X) gives

$$\mathbb{E}\left[\|x_t^{\pi^\star}\|_{P_\star}^2\right] = \mathbb{E}\left[\|x_{t+1}^{\pi_\star}\|_{P_\star}^2 - \|\eta_t\|_{P_\star}^2 + \|x_t^{\pi^\star}\|_Q^2 + \|u_t^{\pi^\star}\|_R^2\right].$$

Summing over time after rearranging gives:

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} (x_t^{\pi^\star})^\top Q(x_t^{\pi^\star}) + (u_t^{\pi^\star})^\top R(u_t^{\pi^\star})\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \|x_t^{\pi^\star}\|_{P_\star}^2 - \|x_{t+1}^{\pi_\star}\|_{P_\star}^2 + \|\eta_t\|_{P_\star}^2\right] \\ &= \mathbb{E}\left[\|x_1\|_{P_\star}^2 - \|x_{T+1}^{\pi_\star}\|_{P_\star}^2\right] + TJ_\star.\end{aligned}$$

Note that $\mathbb{E}[\|x_{T+1}^{\pi_\star}\|_{P_\star}^2] < \infty$. Indeed it can be verified that

$$\mathbb{E}[\|x_T^{\pi^\star}\|^2] = \sum_{t=0}^{T-1} ((A + BK_\star)^t) P_\star (A + BK_\star)^t + \mathbb{E}[x_1^\top ((A + BK_\star)^T)^\top P_\star (A + BK_\star)^T x_1].$$

Therefore, when considering the expected regret, we can take (15) as the regret definition.

C.2 Regret Decomposition

The regret decomposition for the three scenarios can be unified. To that end, let us denote

$$\forall t \geq 1, \quad \xi_t = \begin{cases} \nu_t & \text{in Scenario I} \\ \zeta_t & \text{in Scenario III (A known)} \\ 0 & \text{in Scenario II (B known)} \end{cases}$$

and note that for all $t \geq 1$, ξ_t is a zero mean gaussian random vector with variance proxy $\tilde{\sigma}_t = \sigma_t$ in Scenario I, $\tilde{\sigma}_t = 1$ in Scenario III (A known), and $\tilde{\sigma}_t = 0$ in Scenario II (B known). Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration such that $\mathcal{F}_t$ is the σ-algebra generated by $(\eta_1, \ldots, \eta_t)$ and $(\xi_1, \ldots, \xi_t)$ for all $t \geq 0$. Now, we may observe that the controller used by our algorithm is of the form

$$\forall t \geq 1, \quad u_t = \widetilde{K}_t x_t + \alpha_t \xi_t \tag{16}$$

where $(\widetilde{K}_t)_{t\geq 0}$ is a sequence of random matrices taking values in $\mathbb{R}^{d_u \times d_x}$, such that $\widetilde{K}_t$ is $\mathcal{F}_{t-1}$-measureable $\forall t \geq 1$ and $\widetilde{K}_0 = 0$, and where $(\alpha_t)_{t\geq 0}$ is a sequence of random variables taking values in $\{0, 1\}$ such that α_t is $\mathcal{F}_{t-1}$-measurable and $\alpha_0 = 0$.

Now we are ready to establish a regret decomposition that is valid for any controller of the form (16). We state this decomposition in the following result.

Lemma 1 (Exact Regret Decomposition)**.** *Let $(u_t)_{t\geq 0}$ be a sequence of control inputs that can be expressed as in (16). Define, for all $t \geq 0$,*

$$\widetilde{P}_t = \begin{cases} P_\star(\widetilde{K}_t) & \text{if} \quad \|B(\widetilde{K}_t - K_\star)\| < \frac{1}{4\|P_\star\|^{3/2}} \text{ and } \|P_\star(\widetilde{K}_t)\| \leq 2\|P_\star\| \\ P_\star & \text{otherwise} \end{cases} \tag{17}$$

$$P_{\star,t} = (A + B\widetilde{K}_t)^\top \widetilde{P}_t (A + B\widetilde{K}_t) + Q + \widetilde{K}_t^\top R \widetilde{K}_t. \tag{18}$$

Then, for all $T \geq 1$

$$\mathbb{E}[R_T(\pi)] = \mathbb{E}\left[\sum_{t=1}^{T} \|x_t\|^2_{P_{\star,t} - \widetilde{P}_{t-1}} + \|\eta_t\|^2_{\widetilde{P}_t - P_\star} + \alpha_t \|\xi_t\|^2_{B^\top \widetilde{P}_t B + R}\right] + \mathbb{E}\left[\|x_1\|^2_{\widetilde{P}_0} - \|x_{T+1}\|^2_{\widetilde{P}_T}\right].$$

Proof of Lemma 1. First, we note that the sequence $(\widetilde{P}_t)_{t\geq 1}$ is well defined. Indeed, when $\|B(\widetilde{K}_t - K_\star)\| < 1/(4\|P_\star\|^{3/2})$, $P_\star(K_t)$ exists and is the solution to the Lyapunov equation $P = (A + B\widetilde{K}_t)^\top P (A + B\widetilde{K}_t) + Q + \widetilde{K}_t^\top R \widetilde{K}_t$ (see Lemma 16).

Next, in view of our choice of control imputs $(u_t)_{t\geq 1}$, we can express the dynamics of the problem as $x_{t+1} = (A + B\widetilde{K}_t)x_t + B\alpha_t\xi_t + \eta_t$ for all $t \geq 0$. Thus, multiplying both sides of (18) by x_t, we obtain the identity

$$\|x_t\|^2_{P_{\star,t}} = \|x_{t+1} - B\alpha_t\xi_t - \eta_t\|^2_{\widetilde{P}_t} + \|x_t\|^2_Q + \|u_t - \alpha_t\xi_t\|^2_R.$$

Then, carefully expanding the above identity, leads to

$$\begin{aligned} \|x_t\|^2_Q + \|u_t\|^2_R = \|x_t\|^2_{P_{\star,t}} - \|x_{t+1}\|^2_{\widetilde{P}_t} + \|\eta_t\|^2_{\widetilde{P}_t} + \alpha_t\|\xi_t\|^2_{B^\top \widetilde{P}_t B + R} \\ + 2(B\alpha_t\xi_t + \eta_t)^\top \widetilde{P}_t (A + B\widetilde{K}_t)x_t + 2\eta_t^\top \widetilde{P}_t B \xi_t + 2\alpha_t \xi_t^\top R \widetilde{K}_t x_t. \end{aligned}$$

We note that $\mathbb{E}[\eta_t|\mathcal{F}_{t-1}] = \mathbb{E}[\xi_t|\mathcal{F}_{t-1}] = 0$, and that $\widetilde{P}_t$, x_t, α_t, and $\widetilde{K}_t$ are all $\mathcal{F}_{t-1}$-measurable. Thus, using the tower rule, and substracting $\|\eta_t\|^2_{P_\star}$ from both sides, we obtain for all $t \geq 1$,

$$\mathbb{E}\left[\|x_t\|^2_Q + \|u_t\|^2_R - \|\eta_t\|^2_{P_\star}\right] = \mathbb{E}\left[\|x_t\|^2_{P_{\star,t}} - \|x_{t+1}\|^2_{\widetilde{P}_t} + \|\eta_t\|^2_{\widetilde{P}_t - P_\star} + \alpha_t\|\xi_t\|^2_{B^\top \widetilde{P}_t B + R}\right].$$

Summing over $t \in \{1, \dots, T\}$ (we note that $x_0 = 0$ and $u_0 = 0$) we obtain

$$\begin{aligned}
\mathbb{E}[R_T(\pi)] &= \mathbb{E}\left[\sum_{t=1}^{T} \|x_t\|_Q^2 + \|u_t\|_R^2 - \|\eta_t\|_{P_\star}^2\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \|x_t\|_{P_{\star,t}}^2 - \|x_{t+1}\|_{\widetilde{P}_t}^2 + \|\eta_t\|_{\widetilde{P}_t - P_\star}^2 + \alpha_t \|\xi_t\|_{B^\top \widetilde{P}_t B + R}^2\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \|x_t\|_{P_{\star,t} - \widetilde{P}_{t-1}}^2 + \|\eta_t\|_{\widetilde{P}_t - P_\star}^2 + \alpha_t \|\xi_t\|_{B^\top \widetilde{P}_t B + R}^2\right] + \mathbb{E}\left[\|x_1\|_{\widetilde{P}_0}^2 - \|x_{T+1}\|_{\widetilde{P}_T}^2\right],
\end{aligned}$$

where we recall that $\mathrm{tr}(P_\star) = \mathcal{J}_\star$. This concludes the proof. □

C.3 Integration Lemma

We use Lemma 2 to integrate the high probability bounds on regret.

Lemma 2. *Let X be a postive random variable such that for all $\delta \in (0, 1)$*

$$\mathbb{P}\left(X > C_1 \log(e/\delta) + C2 \log(e/\delta)^{\beta_1}\right) \le C_3 \log(e/\delta)^{\beta_2} \delta$$

where $C_1, C_2, C_3, \beta_1, \beta_2 > 0$. Then

$$\mathbb{E}[X] \le (2\log(eC_3) + 2^{\beta_2+1}\Gamma(\beta_2+1))C_1 + ((2\log(eC_3))^{\beta_1} + \beta_1 2^{\beta_1+\beta_2}\Gamma(\beta_1+\beta_2))C_2$$

where, here, $\Gamma(\cdot)$ refers to the gamma function.

Proof. First, for convenience, we start by reparmetrizing $\rho = \log(e/\delta)$, so that we have $\mathbb{P}(X > C_1\rho + C_2\rho^{\beta_1}) \le C_3\rho^{\beta_2}e^{-\rho}$ for all $\rho > 1$. Additionally, we note that for $\rho > 2\log(eC_3)$, we have $C_3 e^{-\rho/2} < 1$. Thus for $\rho > 2\log(eC_3)$, we have

$$\mathbb{P}(X > C_1\rho + C_2\rho^{\beta_1}) \le \rho^{\beta_2} e^{-\rho/2}.$$

Now, we integrate and perform the change of variable $u = C_1\rho + C_2\rho^{\beta_1}$ for $u > C_1 2\log(eC_3) + C_2(2\log(eC_3))^{\beta_1}$, which yields

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty \mathbb{P}(X > u)du \\
&\le 2\log(eC_3)C_1 + (2\log(eC_3))^{\beta_1}C_2 + \int_0^\infty \mathbb{P}(X > C_1\rho + C_2\rho^{\beta_1})(C_1 + C_2\beta_1\rho^{\beta_1-1})d\rho.
\end{aligned}$$

Then observe that

$$\begin{aligned}
\int_0^\infty \mathbb{P}(X > C_1\rho + C_2\rho^{\beta_1})(C_1 + C_2\beta_1\rho^{\beta_1-1})d\rho &\le \int_0^\infty (C_1 + C_2\beta_1\rho^{\beta_1-1})\rho^{\beta_2}e^{-\rho/2}d\rho \\
&\le C_1\int_0^\infty \rho^{\beta_2}e^{-\rho/2}d\rho + \beta_1 C_2 \int_0^\infty \rho^{\beta_1+\beta_2-1}e^{-\rho/2}d\rho \\
&\le 2^{\beta_2+1}C_1\int_0^\infty \rho^{\beta_2}e^{-\rho}d\rho + \beta_1 2^{\beta_1+\beta_2}C_2\int_0^\infty \rho^{\beta_1+\beta_2-1}e^{-\rho}d\rho \\
&\le 2^{\beta_2+1}\Gamma(\beta_2+1)C_1 + \beta_1 2^{\beta_1+\beta_2}\Gamma(\beta_1+\beta_2)C_2
\end{aligned}$$

where $\Gamma(x)$ refers to the gamma function evaluated at x. To conclude, we have shown that

$$\mathbb{E}[X] \le (2\log(eC_3) + 2^{\beta_2+1}\Gamma(\beta_2+1))C_1 + ((2\log(eC_3))^{\beta_1} + \beta_1 2^{\beta_1+\beta_2}\Gamma(\beta_1+\beta_2))C_2.$$

□

C.4 Proof of Theorem 1 - Regret Analysis in Scenario I

Motivated by the regret decomposition established in Lemma 1, we define what we shall refer to from now on as *proxy regret* as follows

$$\widetilde{R}_T(\pi) = \sum_{t=1}^{T} \|x_t\|^2_{P_{\star,t}-\widetilde{P}_{t-1}} + \operatorname{tr}\left(\widetilde{P}_t - P_\star\right) + \sigma_t^2 \operatorname{tr}\left(B^\top \widetilde{P}_t B + R\right) + \|x_1\|^2_{\widetilde{P}_0} \tag{19}$$

where $(\widetilde{P}_t)_{t\geq 0}$ is definied as in Lemma 1 with $\alpha_t = 1$, and $\xi_t = \nu_t$ for all $t \geq 1$. We note by the same lemma that $\mathbb{E}[R_T(\pi)] \leq \mathbb{E}[\widetilde{R}_T(\pi)]$, thus we may restrict our attention to analysing the *proxy regret* instead of the true regret. Now, we provide a high probability bound on this *proxy regret* which holds for all confidence levels $\delta > 0$.

Step 1: (Defining the nice event) We start by defining the following event.

$$\mathcal{E}_\delta = \left\{ \begin{array}{lll} & (i) & \widetilde{K}_t = K_t, \\ & (ii) & \|B(K_t - K_\star)\| \leq (4\|P_\star\|^{3/2})^{-1} \\ \forall t \geq t(\delta), & (iii) & \|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2 \\ & (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \\ & (v) & \sum_{t=t(\delta)}^{T} r_t^2 \|x_t\|^2 \leq C_2(\delta) + C_3 \|r_{1:T}\|_2^2 \end{array} \right\} \tag{20}$$

where

$$\begin{aligned} r_t^2 &= \log(e i_t) i_t^{-1/2}, \\ i_t &= \max\{t_k \in \mathcal{T} : t_k \leq t\}, \\ t(\delta) &= \operatorname{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, d_u, \gamma_\star) \log(e/\delta)^{12\gamma_\star} \\ C_1(\delta) &= c_1 \sigma^2 C_K^2 \|P_\star\|^8 \gamma_\star d_x^{-1/2} (d_x + d_u) \log(e C_\circ \mathcal{G}_\circ d_x) \log(e/\delta) \\ C_2(\delta) &= \operatorname{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star) \log(e/\delta)^{19\gamma_\star^2} \\ C_3 &= c_3 \sigma^2 C_B^2 \|P_\star\|^{3/2} \gamma_\star d_x \end{aligned}$$

for some universal postive constants $c_1, c_3 > 0$. Furthremore, applying Theorem 5, we have

$$\mathbb{P}(\mathcal{E}_\delta) \geq 1 - t(\delta)\delta$$

for some proper choice of the universal constants defining $t(\delta), C_1(\delta), C_2(\delta), C_3$.

We can interpret the nice event $\mathcal{E}_\delta$ as follows. The first point *(i)* means that CEC($\mathcal{T}$) is playing certainty equivalence for all $t \geq t(\delta)$. The second *(ii)* means that the sequence of $(K_t)_{t\geq t(\delta)}$ is such that the resulting behaviour of the system is that of a stable system, and naturally $\rho(A + BK_t) < 1$. To see that, we refer the reader Proposition 18 (see also Lemma 16). The third point *(iii)* indicates that the error rate is decreasing as r_t^2. The final points *(iv)- (v)* are perhaps redundent since they can be deduced from points *(ii)-(iii)*, but we include them here for convenience.

Step 2: (Regret from $t(\delta)$ onwards) We bound $\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi)$ under the event $\mathcal{E}_\delta$. Note that under this event, we have

$$\begin{aligned} \widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) &\leq C_1(\delta) \sum_{t=t(\delta)}^{T} r_t^2 (\|x_t\|^2 + d_x) + d_x (2\|P_\star\|\|B\|^2 + 1) \sum_{t=t_{\mathcal{E}}(\delta)}^{T} \sigma_t^2 \\ &\leq C_1(\delta) \left(d_x \|r_{1:T}\|_2^2 + \sum_{t=t(\delta)}^{T} r_t^2 \|x_t\|^2 \right) + 3 d_x C_B^2 \|P_\star\| \|\sigma_{1:T}\|_2^2 \\ &\leq C_1(\delta)(d_x \|r_{1:T}\|_2^2 + C_2(\delta) + C_3 \|r_{1:T}\|^2) + 6 d_x^{3/2} C_B^2 \|P_\star\| \sqrt{T} \\ &\leq 2 C_3 C_1(\delta) \|r_{1:T}\|_2^2 + C_1(\delta) C_2(\delta) + 6 d_x^{3/2} C_B^2 \|P_\star\| \sqrt{T} \end{aligned}$$

where in the first inequality, we used the *(i)* to have $\widetilde{P}_t = P_\star(K_t)$ for all $t \geq t(\delta)$, then used *(iii)* to bound $\|P_\star(K_t) - P_\star(K_{t-1})\| \leq \|P_\star(K_t) - P_\star\| + \|P_\star(K_{t-1}) - P_\star\| \leq C_1(\delta) r_t^2$ for all $t \geq t(\delta)+1$. Next, we used *(iv)*, to bound $\|B^\top \tilde{P}_t B + R\| \leq (2\|P_\star\|\|B\|^2 + \|R\|) \leq (2\|P_\star\|\|B\|^2 + 1)$. Finally we used *(v)* to bound $\sum_{t=t(\delta)}^T r_t^2 \|x_t\|$, and bounded $\|\sigma_{1:T}\|_2^2 \leq 2 d_x^{1/2}\sqrt{T}$.

Since $\mathcal{T}$ satisfies (5), we can easily verify that $\|r_{1:T}\|_2^2 \lesssim \log(T)\sqrt{T}$. To see that, note that $(r_t)_{t\geq 1}$ depends on $\mathcal{T}$, and we can always find $\mathcal{T}' = \{(C')^k : k \in \mathbb{N}\}$ for C' large enough such that the corresponding sequence $(r'_t)_{t\geq 1}$ satisfies $\|r_{1,T}\|_2^2 \leq \|r'_{1:T}\|_2^2 \lesssim \log(T)\sqrt{T}$ since $\mathcal{T}$ satisfies (5).

Therefore, recalling the expressions of $t(\delta), C_1(\delta), C_2(\delta)$ and C_3, we have

$$\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) \leq C_4 \log(e/\delta)\log(T)\sqrt{T} + C_5 \log(e/\delta)^{31\gamma_\star^2} \tag{21}$$

with

$$\begin{aligned} C_4 &= c_4 \sigma^2 C_B^2 C_K^2 \|P_\star\|^{9.5} \log(e C_\circ \mathcal{G}_\circ d_x) d_x^{1/2}(d_x + d_u), \\ C_5 &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star), \end{aligned}$$

for some universal positive constant $c_4 > 0$.

Step 3: We bound $\tilde{R}_{t(\delta)}(\pi)$ under the event $\mathcal{E}_\delta$. Note we can obtain the crude upper bound

$$\begin{aligned} \widetilde{R}_{t(\delta)-1}(\pi) &\leq \max_{1\leq t\leq t(\delta)} \|P_{\star,t}\| \sum_{t=1}^{t(\delta)} \|x_t\|^2 + d_x(1 + \max(\|B\|^2, \|R\|)\sigma_t^2) \\ &\leq \max_{1\leq t\leq t(\delta)} \|P_{\star,t}\| \sum_{t=1}^{t(\delta)} \|x_t\|^2 + 5 d_x C_B^2 \sigma^2 \end{aligned}$$

where we dropped the negative terms $-\|x_t\|^2_{\widetilde{P}_{t-1}}$ for $2 \leq t \leq t(\delta)$. Considering the definition of $P_{\star,t}$, we have

$$\|P_{\star,t}\| \leq \begin{cases} 2\|P_\star\| & \text{if } \|B(\widetilde{K}_t - K_\star)\| < \frac{1}{4\|P_\star\|^{3/2}} \text{ and } \|P_\star(\widetilde{K}_t)\| \leq 2\|P_\star\|^2, \\ 4C_\circ^2 \|P_\star\| h(t) & \text{otherwise,} \end{cases}$$

where we upper bounded $\|P_{\star,t}\| \leq 4\|P_\star\| C_\circ^2 h(t)$, using the fact that under CEC($\mathcal{T}$), we have $\|\widetilde{K}_t\|^2 \leq \max(\|K_\circ\|^2, h(t))$, and the fact that $\|P_\star\| \geq \max(\|Q\|, \|R\|)$. Thus,

$$\max_{1\leq t\leq t(\delta)} \|P_{\star,t}\| \leq 4C_\circ^2 \|P_\star\| h(t).$$

Thus, we may write

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq 4C_\circ^2 \|P_\star\| h(t(\delta)) \sum_{t=1}^{t(\delta)} (\|x_t\|^2 + 5\sigma^2 d_x C_B^2).$$

Therefore under event $\mathcal{E}_\delta$, we have

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq 36\sigma^2 d_x C_\circ^4 \|P_\star\| h(t(\delta)) f(t(\delta)),$$

since when property *(i)* holds at time $t(\delta)$, then it must mean that $\ell_{t(\delta)} = 1$, which means that $\sum_{s=1}^{t(\delta)} \|x_s\|^2 \leq \sigma^2 d_x f(t(\delta))$. Hence, recalling the expression of $t(\delta)$, we obtain

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq C_5 \log(e/\delta)^{30\gamma_\star^2} \tag{22}$$

where we note that the hidden universal constants hidden in poly$(\cdot)$ may be chosen large enough so that C_5.

Step 4: (Putting everything together) Now, under the event $\mathcal{E}_\delta$, using (21) and (22) we have

$$\begin{aligned} \widetilde{R}_T(\pi) &= \widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) + \widetilde{R}_{t(\delta)-1}(\pi) \\ &\leq C_4 \log(e/\delta)\log(T)\sqrt{T} + C_5 \log(e/\delta)^{31\gamma_\star^2} \end{aligned}$$

where we note that the universal postive constants hidden in poly(.) may be chosen large enough so that $C_5 \geq C_6$. Therefore, we have established that

$$\mathcal{E}_\delta \subseteq \left\{ \widetilde{R}_T(\pi) \leq C_4 \log(e/\delta) \log(T)\sqrt{T} + 2C_5 \log(e/\delta)^{31\gamma_\star^2} \right\}$$

where $C_6 = 2C_5$. Now recalling the expression of $t(\delta)$ and that $\mathbb{P}(\mathcal{E}_\delta) \geq 1 - t(\delta)\delta$ for all $\delta \in (0,1)$, we obtain that for all $\delta \in (0,1)$, we have

$$\mathbb{P}\left(\widetilde{R}_T(\pi) \leq C_4 \log(e/\delta) \log(T)\sqrt{T} + C_6 \log(e/\delta)^{31\gamma_\star^2} \right) \geq 1 - C_6 \log(e/\delta)^{31\gamma_\star}\delta \tag{23}$$

where $C_6 = \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star)$. Now, integrating (23) using Lemma 2, yields the final result

$$\mathbb{E}[R_T(\pi)] \leq \mathbb{E}[\widetilde{R}_T(\pi)] \leq C_7 \log(T)\sqrt{T} + C_8$$

where

$$\begin{aligned} C_7 &= c_7 \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| C_B d_x d_u \gamma_\star)^2 \sigma^2 C_B^2 C_K^2 \|P_\star\|^{9.5} d_x^{1/2}(d_x + d_u) \\ C_8 &= \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star) \end{aligned}$$

and where c_7 is a positive constant that only depends polynomially on $\gamma_\star$ – the order poly$(\cdot)$ may depend on $\gamma_\star$.

C.5 Proof Of Theorem 3 - Regret Analysis In Scenario III (A known)

The proof is very similar to that of Theorem 1 (see C.4). The only difference is that now there are no input perturbation whenever CEC$(\mathcal{T})$ uses the certainty equivalence controller, and the error rates of the LSE are now better. We shall highlight these differences throughout the proof.

Again following Lemma 1, we define the *proxy regret* as follows

$$\widetilde{R}_T(\pi) = \sum_{t=1}^{T} \|x_t\|^2_{P_t - \widetilde{P}_{t-1}} + \text{tr}\left(\widetilde{P}_t - P_\star\right) + \alpha_t \widetilde{\sigma}^2 \,\text{tr}\left(B^\top \widetilde{P}_t B + R\right) + \|x_1\|^2_{\widetilde{P}_0}$$

where $(\widetilde{P}_t)_{t\geq 1}$ is defined as in Lemma 1 with $\alpha_t = 1_{\{\widetilde{K}_t \neq K_t\}}$, $\xi_t = \zeta_t$ and thus $\widetilde{\sigma}^2 \leq 1$. Note by the same lemma, we have $\mathbb{E}[R_T(\pi)] \leq \mathbb{E}[\widetilde{R}_T(\pi)]$.

Step 1: (Defining the nice event) We start by applying Theorem 5, which guarantees that the event

$$\mathcal{E}_\delta = \left\{ \forall t \geq t(\delta), \begin{array}{ll} (i) & \widetilde{K}_t = K_t, \\ (ii) & \|B(K_t - K_\star)\| \leq (4\|P_\star\|^{3/2})^{-1} \\ (iii) & \|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2 \\ (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \\ (v) & \sum_{t=t(\delta)}^{T} r_t^2 \|x_t\|^2 \leq C_2(\delta) + C_3 \|r_{1:T}\|_2^2 \end{array} \right\} \tag{24}$$

holds with probability at leat $1 - t(\delta)\delta$. In the definition $\mathcal{E}_\delta$, we have

$$\begin{aligned} r_t^2 &= i_t^{-1}, \\ i_t &= \max\{t_k \in \mathcal{T} : t_k \leq t\}, \\ t(\delta) &= \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{18\gamma_\star^2}, \\ C_1(\delta) &= \frac{c_1 \sigma^2 \|P_\star\|^8 (d_u + d_x)\gamma_\star \log\left(\frac{e\sigma C_K \|P_\star\| d_x d_u}{\mu_\star^2}\right)}{\mu_\star^2} \log(e/\delta), \\ C_2(\delta) &= \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{54\gamma_\star^2}, \\ C_3 &= c_3 \sigma^2 \|P_\star\|^{3/2} d_x, \end{aligned}$$

for some universal positive constants $c_1, c_3 > 0$.

The properties *(i)-(v)* have the same interpretations as in C.4, with the distinction that this time, the error rates are much better. We use these properties to bound the *proxy regret*.

Step 2: (Regret from $t(\delta)$ onwards under the nice event) We bound $\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi)$ under the event $\mathcal{E}_\delta$. First let us note that under this event, we have $\alpha_t = 0$ for all $t \geq t(\delta)$. Therefore, we have

$$\begin{aligned}
\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) &\leq C_1(\delta) \sum_{t=t(\delta)}^{T} r_t^2(\|x_t\|^2 + d_x) \\
&\leq C_1(\delta) \left(d_x \|r_{1:T}\|_2^2 + 2 \sum_{t=t(\delta)}^{T} r_t^2 \|x_t\|^2 \right) \\
&\leq C_1(\delta)(d_x \|r_{1:T}\|_2^2 + 2C_2(\delta) + 2C_3 \|r_{1:T}\|^2) \\
&\leq 3C_3 C_1(\delta) \|r_{1:T}\|_2^2 + 2C_1(\delta) C_2(\delta)
\end{aligned}$$

where in the first inequality, we used the *(i)* to have $\widetilde{P}_t = P_\star(K_t)$ for all $t \geq t(\delta)$, then used *(iii)* to bound $\|P_\star(K_t) - P_\star(K_{t-1})\| \leq \|P_\star(K_t) - P_\star\| + \|P_\star(K_{t-1}) - P_\star\| \leq 2C_1(\delta) r_t^2$ for all $t \geq t(\delta) + 1$.

Since $\mathcal{T}$ satisfies (5), we can easily verify that $\|r_{1:T}\|_2^2 \lesssim \log(T)$. To see that, note that $(r_t)_{t \geq 1}$ depends on $\mathcal{T}$, and we can always find $\mathcal{T}' = \{(C')^k : k \in \mathbb{N}\}$ for C' large enough such that the corresponding sequence $(r'_t)_{t \geq 1}$ satisfies $\|r_{1,T}\|_2^2 \leq \|r'_{1:T}\|_2^2 \lesssim \log(T)$ since $\mathcal{T}$ satisfies (5).

Therefore, recalling the expressions of $t(\delta), C_1(\delta), C_2(\delta)$ and C_3, we have

$$\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) \leq C_4 \log(e/\delta) \log(T) + C_5 \log(e/\delta)^{55\gamma_\star^2} \tag{25}$$

with

$$\begin{aligned}
C_4 &= \frac{c_4 \sigma^2 \|P_\star\|^{9.5}}{\mu_\star^2} \log\left(\frac{e C_K \|P\|_\star d_x d_u}{\mu_\star^2} \right) d_x(d_x + d_u)\gamma_\star, \\
C_5 &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star),
\end{aligned}$$

for some universal positive constant $c_4 > 0$.

Step 3: (Regret up to $t(\delta)$ under the nice event)s Now, we bound $\widetilde{R}_{t(\delta)}(\pi)$ under the event $\mathcal{E}_\delta$. Note we can obtain the crude upper bound

$$\begin{aligned}
\widetilde{R}_{t(\delta)-1}(\pi) &\leq \max_{1 \leq t \leq t(\delta)} \|P_{\star,t}\| \sum_{t=1}^{t(\delta)} \|x_t\|^2 + d_x \widetilde{\sigma}(1 + \max(\|B\|^2, \|R\|)) \\
&\leq \max_{1 \leq t \leq t(\delta)} \|P_{\star,t}\| \sum_{t=1}^{t(\delta)} \|x_t\|^2 + 2 d_x C_B^2
\end{aligned}$$

where we dropped the negative terms $-\|x_t\|^2_{\widetilde{P}_{t-1}}$ for $2 \leq t \leq t(\delta)$. Recalling the definition of $P_{\star,t}$, we have

$$\|P_{\star,t}\| \leq \begin{cases} 2\|P_\star\| & \text{if } \|B(\widetilde{K}_t - K_\star)\| < \frac{1}{4\|P_\star\|^{3/2}} \text{ and } \|P_\star(\widetilde{K}_t)\| \leq 2\|P_\star\|^2, \\ 4C_\circ^2 \|P_\star\| h(t) & \text{otherwise} \end{cases}$$

where we upper bounded $\|P_{\star,t}\| \leq 4\|P_\star\| C_\circ^2 h(t)$, using the fact that under CEC($\mathcal{T}$), we have $\|\widetilde{K}_t\|^2 \leq \max(\|K_\circ\|^2, h(t))$, and the fact that $\|P_\star\| \geq \max(\|Q\|, \|R\|)$. Thus,

$$\max_{1 \leq t \leq t(\delta)} \|P_{\star,t}\| \leq 4C_\circ^2 \|P_\star\| h(t).$$

Thus, we may write

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq 4C_\circ^2 \|P_\star\| h(t(\delta)) \sum_{t=1}^{t(\delta)} (\|x_t\|^2 + 2 d_x C_B^2)$$

Therefore under event $\mathcal{E}_\delta$, we have

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq 12\sigma^2 d_x C_\circ^2 C_B^2 \|P_\star\| h(t(\delta)) f(t(\delta)),$$

since when property *(i)* holds at time $t(\delta)$, then it must mean that $\ell_{t(\delta)} = 1$, which means that $\sum_{s=1}^{t(\delta)} \|x_s\|^2 \leq \sigma^2 d_x f(t(\delta))$. Hence, recalling the expression of $t(\delta)$, we obtain

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq C_6 \log(e/\delta)^{45\gamma_\star^3} \tag{26}$$

where $C_6 = \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|^2, \mu_\star^{-1}, C_B, d_x, d_u, \gamma_\star)$.

Step 4: (Putting everything together) Now, under the event $\mathcal{E}_\delta$, using (25) and (26), we have

$$\begin{aligned}\widetilde{R}_T(\pi) &= \widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) + \widetilde{R}_{t(\delta)-1}(\pi) \\ &\leq C_4 \log(e/\delta)\log(T) + C_7 \log(e/\delta)^{55\gamma_\star^3}\end{aligned}$$

where we note that the universal postive constants hidden in poly(.) may be chosen large enough so that $C_7 \geq 2C_5 + C_6$. Therefore, we have established that

$$\mathcal{E}_\delta \subseteq \left\{\widetilde{R}_T(\pi) \leq C_4 \log(e/\delta)\log(T) + C_7 \log(e/\delta)^{55\gamma_\star^3}\right\}.$$

Now recalling the expression of $t(\delta)$ and that $\mathbb{P}(\mathcal{E}_\delta) \geq 1 - t(\delta)\delta$ for all $\delta \in (0,1)$, we obtain that for all $\delta \in (0,1)$, we have

$$\mathbb{P}\left(\widetilde{R}_T(\pi) \leq C_4 \log(e/\delta)\log(T) + C_7 \log(e/\delta)^{55\gamma_\star^3}\right) \geq 1 - C_7 \log(e/\delta)^{55\gamma_\star^3}\delta \tag{27}$$

where the hidden universal constants in poly$(\cdot)$ defining C_7 may be chosen to be large enough for the above to hold. Now, integrating (27) using Lemma 2, yields the final result

$$\mathbb{E}[R_T(\pi)] \leq \mathbb{E}[\widetilde{R}_T(\pi)] \leq C_8 \log(T) + C_9$$

where

$$\begin{aligned}C_8 &= \frac{c_8 \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| \mu_\star^{-1} C_B d_x d_u \gamma_\star)^2 \sigma^2 \|P_\star\|^{9.5} d_x (d_x + d_u)}{\mu_\star^{-1}} \\ C_9 &= \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, C_B, d_x, d_u, \gamma_\star)\end{aligned}$$

where c_8 is a positive constant that only depends polynomially on $\gamma_\star$, and the order of poly$(\cdot)$ may depend on $\gamma_\star$.

C.6 Proof Of Theorem 2 - Regret Analysis In Scenario II (B known)

Again, the proof is very similar to that of Theorems 1 and 2 (see C.4 and C.5). Note that in this scenario, there are no input perturbations, and the LSE error rates are as fast as in Scenario III. We shall highlight these differences throughout the proof.

Again following Lemma 1, we define the *proxy regret* as follows

$$\widetilde{R}_T(\pi) = \sum_{t=1}^{T} \|x_t\|^2_{P_t - \widetilde{P}_{t-1}} + \text{tr}\left(\widetilde{P}_t - P_\star\right) + \|x_1\|^2_{\widetilde{P}_0}$$

where $(\widetilde{P}_t)_{t\geq 1}$ is defined as in Lemma 1 with $\alpha_t = 0$, $\xi_t = 0$. Note by the same lemma, we have $\mathbb{E}[R_T(\pi)] \leq \mathbb{E}[\widetilde{R}_T(\pi)]$.

Step 1: (Defining the nice event) We start by applying Theorem 5, which guarantees that the event

$$\mathcal{E}_\delta = \left\{ \begin{array}{lll} & (i) & \widetilde{K}_t = K_t, \\ & (ii) & \|B(K_t - K_\star)\| \leq (4\|P_\star\|^{3/2})^{-1} \\ \forall t \geq t(\delta), & (iii) & \|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2 \\ & (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \\ & (v) & \sum_{t=t(\delta)}^{T} r_t^2 \|x_t\|^2 \leq C_2(\delta) + C_3 \|r_{1:T}\|_2^2 \end{array} \right\} \tag{28}$$

holds with probability at leat $1 - t(\delta)\delta$. In the definition of $\mathcal{E}_\delta$, we have

$$\begin{aligned}
r_t^2 &= i_t^{-1}, \\
i_t &= \max\{t_k \in \mathcal{T} : t_k \le t\}, \\
t(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{6\gamma_\star^2}, \\
C_1(\delta) &= c_6 \sigma^2 \|P_\star\|^8 d_x \log\left(e\|P_\star\| d_x\right) \log(e/\delta), \\
C_2(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{15\gamma_\star^3}, \\
C_3 &= 24\sigma^2 \|P_\star\|^{3/2} d_x,
\end{aligned}$$

for some universal postive constants $c_1, c_3 > 0$.

The properties *(i)-(v)* have the same interpretations as in C.4, with the distinction that this time again, the error rates are much better. We use these properties to bound the *proxy regret*.

Step 2: (Regret from $t(\delta)$ onwards under the nice event) We bound $\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi)$ under the event $\mathcal{E}_\delta$. First let us note that under this event, we have $\alpha_t = 0$ for all $t \ge t(\delta)$. Therefore, we have

$$\begin{aligned}
\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) &\le C_1(\delta) \sum_{t=t(\delta)}^{T} r_t^2(\|x_t\|^2 + d_x) \\
&\le C_1(\delta) \left(d_x \|r_{1:T}\|_2^2 + 2 \sum_{t=t(\delta)}^{T} r_t^2 \|x_t\|^2 \right) \\
&\le C_1(\delta)(d_x \|r_{1:T}\|_2^2 + 2C_2(\delta) + 2C_3 \|r_{1:T}\|^2) \\
&\le 3C_3 C_1(\delta) \|r_{1:T}\|_2^2 + 2C_1(\delta) C_2(\delta)
\end{aligned}$$

where in the first inequality, we used the *(i)* to have $\widetilde{P}_t = P_\star(K_t)$ for all $t \ge t(\delta)$, then used *(iii)* to bound $\|P_\star(K_t) - P_\star(K_{t-1})\| \le \|P_\star(K_t) - P_\star\| + \|P_\star(K_{t-1}) - P_\star\| \le 2C_1(\delta) r_t^2$ for all $t \ge t(\delta) + 1$.

Since $\mathcal{T}$ satisfies (5), we can easily verify that $\|r_{1:T}\|_2^2 \lesssim \log(T)$ (see the proof in the previous scenario).

Therefore, recalling the expressions of $t(\delta), C_1(\delta), C_2(\delta)$ and C_3, we have

$$\widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) \le C_4 \log(e/\delta) \log(T) + C_5 \log(e/\delta)^{16\gamma_\star^2} \tag{29}$$

with

$$\begin{aligned}
C_4 &= c_4 \sigma^4 \|P_\star\|^{9.5} \log\left(e\|P\|_\star d_x\right) d_x^2 \gamma_\star, \\
C_5 &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star),
\end{aligned}$$

for some universal positive constant $c_4 > 0$.

Step 3: (Regret up to $t(\delta)$ under the nice event). Now, we bound $\widetilde{R}_{t(\delta)}(\pi)$ under the event $\mathcal{E}_\delta$. Note we can obtain the crude upper bound

$$\widetilde{R}_{t(\delta)-1}(\pi) \le \max_{1 \le t \le t(\delta)} \|P_{\star,t}\| \sum_{t=1}^{t(\delta)} \|x_t\|^2$$

where we dropped the negative terms $-\|x_t\|^2_{\widetilde{P}_{t-1}}$ for $2 \le t \le t(\delta)$. Recalling the definition of $P_{\star,t}$, we have

$$\|P_{\star,t}\| \le \begin{cases} 2\|P_\star\| & \text{if } \|B(\widetilde{K}_t - K_\star)\| < \frac{1}{4\|P_\star\|^{3/2}} \text{ and } \|P_\star(\widetilde{K}_t)\| \le 2\|P_\star\|^2, \\ 4C_\circ^2 \|P_\star\| h(t) & \text{otherwise} \end{cases}$$

where we upper bounded $\|P_{\star,t}\| \le 4\|P_\star\| C_\circ^2 h(t)$, using the fact that under CEC($\mathcal{T}$), we have $\|\widetilde{K}_t\|^2 \le \max(\|K_\circ\|^2, h(t))$, and the fact that $\|P_\star\| \ge \max(\|Q\|, \|R\|)$. Thus,

$$\max_{1 \le t \le t(\delta)} \|P_{\star,t}\| \le 4C_\circ^2 \|P_\star\| h(t).$$

Thus, we may write

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq 4C_\circ^2 \|P_\star\| h(t(\delta)) \sum_{t=1}^{t(\delta)} \|x_t\|^2.$$

Therefore, under event $\mathcal{E}_\delta$, we have

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq 4\sigma^2 d_x C_\circ^2 C_B^2 \|P_\star\| h(t(\delta)) f(t(\delta))$$

since when property *(i)* holds at time $t(\delta)$, then it must mean that $\ell_{t(\delta)} = 1$, which means that $\sum_{s=1}^{t(\delta)} \|x_s\|^2 \leq \sigma^2 d_x f(t(\delta))$. Hence, recalling the expression of $t(\delta)$, we obtain

$$\widetilde{R}_{t(\delta)-1}(\pi) \leq C_6 \log(e/\delta)^{15\gamma_\star^3} \tag{30}$$

where $C_6 = \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, \gamma_\star)$.

Step 4: (Putting everything together) Now, under the event $\mathcal{E}_\delta$, using (29) and (30) we have

$$\begin{aligned}\widetilde{R}_T(\pi) &= \widetilde{R}_T(\pi) - \widetilde{R}_{t(\delta)-1}(\pi) + \widetilde{R}_{t(\delta)-1}(\pi) \\ &\leq C_4 \log(e/\delta)\log(T) + C_7 \log(e/\delta)^{16\gamma_\star^3}\end{aligned}$$

where we note that the universal positive constants hidden in poly(.) may be chosen large enough so that $C_7 \geq C_5 + C_6$. Therefore, we have established that

$$\mathcal{E}_\delta \subseteq \left\{ \widetilde{R}_T(\pi) \leq C_4 \log(e/\delta)\log(T) + C7 \log(e/\delta)^{16\gamma_\star^3} \right\}.$$

Now recalling the expression of $t(\delta)$ and that $\mathbb{P}(\mathcal{E}_\delta) \geq 1 - t(\delta)\delta$ for all $\delta \in (0,1)$, we obtain that for all $\delta \in (0,1)$,

$$\mathbb{P}\left(\widetilde{R}_T(\pi) \leq C_4 \log(e/\delta)\log(T) + C_7 \log(e/\delta)^{16\gamma_\star^3} \right) \geq 1 - C_7 \log(e/\delta)^{16\gamma_\star^3}\delta \tag{31}$$

where the hidden universal constants in poly$(\cdot)$ defining C_7 may be chosen large enough for the above to hold. Now, integrating (31) using Lemma 2, yields the final result

$$\mathbb{E}[R_T(\pi)] \leq \mathbb{E}[\widetilde{R}_T(\pi)] \leq C_8 \log(T) + C_9$$

where

$$\begin{aligned}C_8 &= c_8 \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| C_B d_x \gamma_\star)^2 \sigma^4 \|P_\star\|^{9.5} d_x^2 \gamma_\star, \\ C_9 &= \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, \gamma_\star),\end{aligned}$$

where c_8 is a positive constant that only depends polynomially on $\gamma_\star$, and the order of poly$(\cdot)$ may depend on $\gamma_\star$.

D THE NICE EVENT AND ITS LIKELIHOOD

In Appendix C, we have seen that the regret analysis relied on the definition of a "nice" event $\mathcal{E}_\delta$, and on the fact that its occurrence probability is close enough to 1. This appendix is devoted to presenting such event and establishing its likelihood for all the three envisioned scenarios. The main results are stated in Theorem 5, Theorem 6, and Theorem 7 for Scenario I, Scenario II – A known, and Scenario II – B known, respectively. Their proofs follow the same line of reasoning and rely on the consistency of the least squares estimator (see Appendix F), perturbations bounds on Riccati equations (see Appendix I), and the fact that CEC($\mathcal{T}$) eventually just uses the certainty equivalence controller K_t (see Appendix E).

D.1 Scenario I

To analyse regret, we need to ensure the event $\mathcal{E}_\delta$ where for all $t \geq t(\delta)$

(i) $\widetilde{K}_t = K_t$

(ii) $\|B(K_t - K_\star)\| \leq (4\|P_\star\|^{3/2})^{-1}$

(iii) $\|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2$

(iv) $\|P_\star(K_t)\| \leq 2\|P_\star\|$

(v) $\sum_{s=t(\delta)}^{t} r_t^2 \|x_t\|^2 \leq C_3 \|r_{1:T}\|_2^2 + C_2(\delta)$

holds with probability at least $1 - t(\delta)\delta$ for all $\delta \in (0,1)$. We shall precise $t(\delta), C_1(\delta), C_2(\delta)$ and C_3 in Theorem 5. As for the sequence $(r_t)_{t\geq 1}$, it is defined as

$$\forall t \geq 1, \quad r_t^2 = \frac{\log(e i_t)}{i_t^{1/2}}$$

where $i_t = \max\{t_k \in \mathcal{T} : t_k \leq t\}$. We note that if $\mathcal{T} = \mathbb{N}$, then $i_t = t$, and if $\mathcal{T} = \{e^k : t \in \mathbb{N}\}$, then $i_t = e^{\lfloor \log(t) \rfloor}$.

Theorem 5. *Assume $\mathcal{T}$ satisfies (5). Then under* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$, we have*

$$\mathbb{P}(\forall t \geq t(\delta),\ \text{(i)} - \text{(v)}\ \text{hold}) \geq 1 - t(\delta)\delta$$

where

$$\begin{aligned}
t(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, d_u, \gamma_\star) \log(e/\delta)^{12\gamma_\star}, \\
C_1(\delta) &= c_1 \sigma^2 C_K^2 \|P_\star\|^8 \gamma_\star d_x^{-1/2} (d_x + d_u) \log(e C_\circ \mathcal{G}_\circ d_x) \log(e/\delta), \\
C_2(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star) \log(e/\delta)^{19\gamma_\star^3}, \\
C_3 &= c_3 \sigma^2 C_B^2 \|P_\star\|^{3/2} \gamma_\star d_x,
\end{aligned}$$

for some universal positive constant $c_1, c_3 > 0$.

Proof. The proof proceeds in the following steps.

Step 1: (Least squares estimation under CEC($\mathcal{T}$)) First, since $\mathcal{T}$ satisifes (5), we have by Proposition 2 that under CEC($\mathcal{T}$), the following holds

$$\max(\|A_t - A\|^2, \|B_t - B\|^2) \leq \frac{C_1 \sigma^2 C_K^2 \left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x t\right) + \log(e/\delta)\right)}{(d_x t)^{1/2}} \tag{32}$$

with probability at least $1 - \delta$, provided that

$$t^{1/4} \geq c\sigma^2 C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right), \tag{33}$$

for some universal positive constants $C_1, c_1 > 0$. Constraining further t to satisfy

$$t^{1/2} \geq \frac{C_1 160^2 \sigma^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x t\right) + \log(e/\delta)\right)}{(d_x)^{1/2}} \tag{34}$$

ensures in addition that

$$\max(\|A_t - A\|, \|B_t - B\|) \leq \frac{1}{160\|P_\star\|^5}. \tag{35}$$

Now, using Lemma 22, we can find a universal positive constant $c_2 > 0$ such that if

$$t^{1/4} \geq c_2 \sigma^2 C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right), \tag{36}$$

then conditions (33) and (34) also hold.

We remind here that because $\mathcal{T}$ satisfies condition (5), if the constant $c_2 > 0$ is chosen large enough, we can claim that for all $t \in \mathbb{N}$ such that (33) holds, we have $K_t = K_{t_k}$ for some t_k that also satisfies (33). Therefore, applying Lemma 23, ensures that

$$\mathbb{P}(\forall t \geq t_1(\delta), \quad (32) \text{ and } (35)) \geq 1 - \delta \tag{37}$$

with

$$t_1(\delta)^{1/4} = c_3 \sigma^2 C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right). \tag{38}$$

for some universal postive constant $c_3 > 0$. Finally, a direct application of Proposition 16 ensures that

$$\mathcal{C}_\delta = \left\{ \forall t \geq t_1(\delta), \quad \begin{array}{ll} (ii) & \|B(K_t - K_\star)\| \leq \frac{1}{4\|P_\star\|^{3/2}} \\ (iii) & \|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2 \\ (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \end{array} \right\} \tag{39}$$

holds with probability $1 - \delta$, with

$$C_1(\delta) = 140 C_1 \sigma^2 C_K^2 \|P_\star\|^8 d_x^{-1/2} (d_x + d_u) \gamma_\star \log\left(e\sigma C_\circ \mathcal{G}_\circ d_x\right) \log(e/\delta).$$

Step 2: (Commitment to the certainty equivalence controller) Using Theorem 8, we have

$$\mathcal{D}_\delta = \left\{ \forall t \geq t_2(\delta), \quad (i)\ \widetilde{K}_t = K \right\} \tag{40}$$

holds with probability at least $1 - 5t_2(\delta)\delta$, with

$$t_2(\delta) = \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, d_u, \gamma_\star) \log(e/\delta)^{12\gamma_\star} \tag{41}$$

such that $t_2(\delta) \geq t_1(\delta)$ (this can be ensured by taking the universal positive constants hidden poly$(\cdot)$ large enough).

Step 3: (Stability under the certainty equivalence controller) Let us define the events

$$E_{1,\delta,T} = \left\{ \sum_{t=t_2(\delta)}^{T} r_t^2 \|x_t\|^2 \leq 8\|P_\star\|^{3/2} \left(\|r_{1:T}\|_\infty^2 \|x_{t_2(\delta)}\|^2 + 6 C_B^2 \sigma^2 \left(\|r_{1:T}\|_2^2 + \log(e/\delta) \right) \right) \right\},$$

$$E_{2,\delta} = \left\{ \forall t \geq t_2(\delta), \quad (i) \text{ and } (ii) \text{ hold } \right\}.$$

Noting that $t_2(\delta)$ may be chosen so that $t_2(\delta) \geq d_x$, we obtain by a direct application Proposition 18,

$$\mathbb{P}\left(E_{1,\delta,T} \cup E_{2,\delta}^c\right) \geq 1 - \delta.$$

Step 4: (Putting everything together) To conclude, we note that under the event $\mathcal{C}_\delta \cap \mathcal{D}_\delta \cap (E_{1,\delta,T} \cup E_{2,\delta})$, the propoerties $(i)-(iv)$ hold for all $t \geq t_2(\delta)$. Additionally, under CEC$(\mathcal{T})$, it also holds that $\|x_{t_2(\delta)}\|^2 \leq \sigma^2 d_x f(t_2(\delta))$,

therefore we have

$$\begin{aligned}\sum_{t=t_2(\delta)}^{T} r_t^2\|x_t\|^2 &\le 8\|P_\star\|^{3/2}\left(\|r_{t_2(\delta):T}\|_\infty^2\|x_{t_2(\delta)}\|^2 + 6C_B^2\sigma^2\left(d_x\|r_{1:T}\|_2^2 + \log(e/\delta)\right)\right)\\ &\le 8\|P_\star\|^{3/2}\left(\|r_{t_2(\delta):T}\|_\infty^2\|x_{t_2(\delta)}\|^2 + 6C_B^2\sigma^2\left(d_x\|r_{1:T}\|_2^2 + \log(e/\delta)\right)\right)\\ &\le 48\sigma^2\|P_\star\|^{3/2}C_B^2 d_x\|r_{1:T}\|_2^2 + \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star)\log(e/\delta)t_2(\delta)^{1+\gamma/2}\\ &\le 48\sigma^2\|P_\star\|^{3/2}C_B^2 d_x\|r_{1:T}\|_2^2 + \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star)\log(e/\delta)^{19\gamma_\star^2}\end{aligned}$$

where we used the fact $\|r_{t:T}\|_\infty^2 \le \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, d_u, \gamma_\star)$. Thus, defining the property

$$\text{(v)}\quad \sum_{t=t_2(\delta)}^{T} r_t^2\|x_t\|^2 \le C_3\|r_{1:T}\|_2^2 + C_2(\delta),$$

where

$$\begin{aligned}C_2(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, C_B, d_x, d_u, \gamma_\star)\log(e/\delta)^{19\gamma_\star^2},\\ C_3 &= 48\sigma^2\|P_\star\|^{3/2}C_B^2 d_x,\end{aligned}$$

we have shown that

$$\mathcal{E}_\delta = \{\forall t \ge t_2(\delta),\quad (i)-(v)\text{ hold}\} \subseteq \mathcal{C}_\delta \cap \mathcal{D}_\delta \cap E_{1,\delta,T}$$

Finally, we note that $\mathcal{C}_\delta \cap \mathcal{D}_\delta \cap (E_{1,\delta,T} \cup E_{2,\delta}) = \mathcal{C}_\delta \cap \mathcal{D}_\delta \cap E_{1,\delta,T}$. Therefore, by a union bound, we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}_\delta) &\ge 1 - \mathbb{P}\left(\mathcal{C}_\delta^c \cup \mathcal{D}_\delta^c \cup (E_{1,\delta,t} \cup E_{d,\delta})^c\right)\\ &\ge 1 - 2\delta - 5t_2(\delta)\delta\\ &\ge 1 - 7t_2(\delta)\delta.\end{aligned}$$

This gives the desired result with modified universal constants. □

D.2 Scenario III – A known

To analyse regret, we need to ensure the event $\mathcal{E}_\delta$ where for all $t \ge t(\delta)$

(i) $\widetilde{K}_t = K_t$

(ii) $\|B(K_t - K_\star)\| \le (4\|P_\star\|^{3/2})^{-1}$

(iii) $\|P_\star(K_t) - P_\star\| \le C_1(\delta)r_t^2$

(iv) $\|P_\star(K_t)\| \le 2\|P_\star\|$

(v) $\sum_{s=t(\delta)}^{t} r_t^2\|x_t\|^2 \le C_3\|r_{1:T}\|_2^2 + C_2(\delta)$

holds with probability at least $1 - t(\delta)\delta$ for all $\delta \in (0,1)$. We shall precise $t(\delta), C_1(\delta), C_2(\delta)$ and C_3 in Theorem 6. As for the sequence $(r_t)_{t\ge1}$, it is defined this time as

$$\forall t \ge 1,\quad r_t^2 = \frac{1}{i_t}$$

where $i_t = \max\{t_k \in \mathcal{T} : t_k \le t\}$.

Theorem 6. *Assume $\mathcal{T}$ satisfies (5). Then under* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$, we have*

$$\mathbb{P}(\forall t \geq t(\delta), \quad \text{(i)} - \text{(v) hold}) \geq 1 - t(\delta)\delta$$

where

$$\begin{aligned}
t(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{18\gamma_\star^2}, \\
C_1(\delta) &= \frac{c_1 \sigma^2 \|P_\star\|^8 (d_u + d_x)\gamma_\star \log\left(\frac{e\sigma C_K \|P_\star\| d_x d_u}{\mu_\star^2}\right)}{\mu_\star^2} \log(e/\delta), \\
C_2(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{54\gamma_\star^3}, \\
C_3 &= c_3 \sigma^2 \|P_\star\|^{3/2} d_x,
\end{aligned}$$

for some universal positive constant $c_1, c_3 > 0$.

Proof of Theorem 6. The proof proceeds in the following steps.

Step 1: (Least squares estimation under CEC($\mathcal{T}$)) First, since $\mathcal{T}$ satisfies (5), we have by Proposition 4 that under Algorithm CEC($\mathcal{T}$), the following holds

$$\|B_t - B\|^2 \leq \frac{C_1 \sigma^2 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u t) + d_x + \log(e/\delta)\right)}{\mu_\star^2 t} \tag{42}$$

with probability at least $1 - \delta$, provided that

$$t^{1/2} \geq \frac{c\sigma^2 C_K^2 \|P_\star\|^{10}}{\mu_\star} \left(d_u \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right) \tag{43}$$

for some universal positive constants $C_1, c_1 > 0$. Constraining further t to satisfy

$$t \geq \frac{C_1 160^2 \sigma^2 \|P_\star\|^{10} \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u t) + d_x + \log(e/\delta)\right)}{\mu_\star^2} \tag{44}$$

ensures in addition that

$$\|B_t - B\| \leq \frac{1}{160 \|P_\star\|^5} \tag{45}$$

Now, using Lemma 22, we can find an universal positive constant $c_2 > 0$ such that if

$$t^{1/2} \geq \frac{c_2 \sigma^2 C_K^2 \|P_\star\|^{10}}{\mu_\star} \left(d \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right), \tag{46}$$

then conditions (43) and (44) also hold.

We remind here that because $\mathcal{T}$ satisfies condition (5), if the constant $c_2 > 0$ is chosen large enough, we can claim that for all $t \in \mathbb{N}$ such that (43) holds, we have $K_t = K_{t_k}$ for some t_k that also satisfies (43). Therefore, applying Lemma 23, ensures that

$$\mathbb{P}(\forall t \geq t_1(\delta), \quad (42) \text{ and } (45)) \geq 1 - \delta \tag{47}$$

with

$$t_1(\delta)^{1/2} = \frac{c_3 \sigma^2 C_K^2 \|P_\star\|^{10}}{\mu_\star} \left(d \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right), \tag{48}$$

for some universal postive constant $c_3 > 0$. Finally, a direct application of Proposition 16 ensures that

$$\mathcal{C}_{1,\delta} = \left\{ \forall t \geq t_1(\delta), \quad \begin{matrix} (ii) & \max(\|B(K_t - K_\star)\|, \|K_t - K_\star\|) \leq \frac{1}{4\|P_\star\|^{3/2}} \\ (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \end{matrix} \right\} \tag{49}$$

holds with probability $1 - \delta$.

Step 2: (Commitment to the certainty equivalence controller) Using Theorem 8, we have

$$\mathcal{D}_\delta = \left\{ \forall t \geq t_2(\delta), \quad (i)\ \widetilde{K}_t = K \right\} \tag{50}$$

holds with probability at least $1 - 5t_2(\delta)\delta$, with

$$t_2(\delta) = \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{6\gamma_\star} \tag{51}$$

such that $t_2(\delta) \geq t_1(\delta)$ (this can be ensured by taking the universal positive constants hidden in $\mathrm{poly}(\cdot)$ large enough).

Step 3: (Refined error rate under the certainty equivalence controller) Now, note that since the event $\mathcal{C}_{1,\delta} \cap \mathcal{D}_{1,\delta}$ holds with probability at least $1 - 6t_2(\delta)\delta$, we may apply Proposition 5 and obtain

$$\|B_t - B\| \leq \frac{c_4\sigma^2}{\mu_\star^2 t}\left((d_u + d_x)\gamma_\star \log\left(\frac{e\sigma C_K \|P_\star\| d_x d_u}{\mu_\star^2}\right) + \log(e/\delta)\right)$$

provided that

$$t \geq \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{18\gamma_\star^2}$$

for some universal positive constant $c_4 > 0$. Again, after using Lemma 23, and using the perturbation bounds via Proposition 16 we obtain that the event

$$\mathcal{C}_{2,\delta} = \left\{ \forall t \geq t_3(\delta), \quad \begin{array}{ll} (ii) & \|B(K_t - K_\star)\| \leq \frac{1}{4\|P_\star\|^{3/2}} \\ (iii) & \|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2 \\ (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \end{array} \right\} \tag{52}$$

holds with probability at least $1 - c_5 t_3(\delta)\delta$, where we define

$$t_3(\delta) = \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{18\gamma_\star^2}$$
$$C_1(\delta) = \frac{c_6\sigma^2\|P_\star\|^8(d_u + d_x)\gamma_\star \log\left(\frac{e\sigma C_K\|P_\star\| d_x d_u}{\mu_\star^2}\right)}{\mu_\star^2} \log(e/\delta)$$

for some universal positive constants $c_5, c_6 > 0$.

Step 4: (Stability under the certainty equivalence controller) Let us define the events

$$E_{1,\delta,T} = \left\{ \sum_{t=t_3(\delta)}^{T} r_t^2 \|x_t\|^2 \leq 8\|P_\star\|^{3/2}\left(\|r_{1:T}\|_\infty^2 \|x_{t_2(\delta)}\|^2 + 3\sigma^2\left(\|r_{1:T}\|_2^2 + \log(e/\delta)\right)\right) \right\},$$
$$E_{2,\delta} = \{\forall t \geq t_3(\delta), \quad (i) \text{ and } (ii) \text{ hold } \}.$$

Noting that $t_2(\delta)$ may be chosen so that $t_2(\delta) \geq d_x$, we obtain by a direct application Proposition 18, that

$$\mathbb{P}\left(E_{1,\delta,T} \cup E_{2,\delta}^c\right) \geq 1 - \delta.$$

Step 5: (Putting everything together) To conclude, we note that under the event $\mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap (E_{1,\delta,T} \cup E_{2,\delta})$, propoerties $(i) - (iv)$ hold for all $t \geq t_3(\delta)$. Additionally, under Algorithm CEC($\mathcal{T}$), it also holds that $\|x_{t_3(\delta)}\|^2 \leq \sigma^2 d_x f(t_3(\delta))$, therefore it follows that

$$\begin{aligned}
\sum_{t=t_3(\delta)}^{T} r_t^2\|x_t\|^2 &\leq 8\|P_\star\|^{3/2}\left(\|r_{t_3(\delta):T}\|_\infty^2\|x_{t_3(\delta)}\|^2 + 3\sigma^2\left(d_x\|r_{1:T}\|_2^2 + \log(e/\delta)\right)\right) \\
&\leq 8\|P_\star\|^{3/2}\left(\|r_{t_3(\delta):T}\|_\infty^2\|x_{t_3(\delta)}\|^2 + 3\sigma^2\left(d_x\|r_{1:T}\|_2^2 + \log(e/\delta)\right)\right) \\
&\leq 24\sigma^2\|P_\star\|^{3/2} d_x\|r_{1:T}\|_2^2 + \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star)\log(e/\delta) t_3(\delta)^{1+\gamma/2} \\
&\leq 24\sigma^2\|P_\star\|^{3/2} d_x\|r_{1:T}\|_2^2 + \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star)\log(e/\delta)^{54\gamma_\star^3}
\end{aligned}$$

where we used the fact $\|r_{t:T}\|_\infty^2 \leq 1$. Thus, defining the property

$$(\mathrm{v}) \quad \sum_{t=t_3(\delta)}^{T} r_t^2\|x_t\|^2 \leq C_3\|r_{1:T}\|_2^2 + C_2(\delta),$$

where

$$C_2(\delta) = \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{54\gamma_\star^3},$$
$$C_3 = 24\sigma^2 \|P_\star\|^{3/2} d_x,$$

we have shown that

$$\mathcal{E}_\delta = \{\forall t \geq t_3(\delta),\ \text{(i)} - \text{(v)}\ \text{hold}\} \subseteq \mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap E_{1,\delta,T}.$$

Finally, we note that $\mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap (E_{1,\delta,T} \cup E_{2,\delta}) = \mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap E_{1,\delta,T}$. Therefore, by a union bound, we have

$$\begin{aligned}\mathbb{P}(\mathcal{E}_\delta) &\geq 1 - \mathbb{P}\left(\mathcal{C}_\delta^c \cup \mathcal{D}_\delta^c \cup (E_{1,\delta,t} \cup E_{d,\delta})^c\right) \\ &\geq 1 - c_7 t_3(\delta)\delta\end{aligned}$$

for some universal positive constant $c_7 > 0$. This gives the desired result with modified universal constants. □

D.3 Scenario II – B known

To analyse regret, we need to ensure the event $\mathcal{E}_\delta$ where for all $t \geq t(\delta)$

(i) $\widetilde{K}_t = K_t$

(ii) $\|B(K_t - K_\star)\| \leq (4\|P_\star\|^{3/2})^{-1}$

(iii) $\|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2$

(iv) $\|P_\star(K_t)\| \leq 2\|P_\star\|$

(v) $\sum_{s=t(\delta)}^t r_t^2 \|x_t\|^2 \leq C_3 \|r_{1:T}\|_2^2 + C_2(\delta)$

holds with probability at least $1 - t(\delta)\delta$ for all $\delta \in (0,1)$. We shall precise $t(\delta), C_1(\delta), C_2(\delta)$ and C_3 in Theorem 7. As for the sequence $(r_t)_{t\geq 1}$, it is defined this time as

$$\forall t \geq 1, \quad r_t^2 = \frac{1}{i_t}$$

where $i_t = \max\{t_k \in \mathcal{T} : t_k \leq t\}$.

Theorem 7. *Assume $\mathcal{T}$ satisfies (5). Then under* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$, we have*

$$\mathbb{P}(\forall t \geq t(\delta),\quad \text{(i)} - \text{(v)}\ \text{hold}) \geq 1 - t(\delta)\delta$$

where

$$\begin{aligned}t(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{6\gamma_\star^2}, \\ C_1(\delta) &= c_6 \sigma^2 \|P_\star\|^8 d_x \log\left(e\|P_\star\| d_x\right) \log(e/\delta), \\ C_2(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{15\gamma_\star^3}, \\ C_3 &= 24\sigma^2 \|P_\star\|^{3/2} d_x,\end{aligned}$$

for some universal positive constant $c_1, c_3 > 0$.

Proof of Theorem 7. The proof proceeds in the following steps.

Step 1: (Least squares estimation under CEC($\mathcal{T}$)) First, since $\mathcal{T}$ satisfies (5), we have by Proposition 6 that under CEC($\mathcal{T}$), the following holds

$$\|A_t - A\|^2 \leq \frac{C_1 \sigma^2 \left(d_x \gamma \log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)\right)}{t} \tag{53}$$

with probability at least $1-\delta$, provided that

$$t \geq c_1(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)), \tag{54}$$

for some universal positive constants $C_1, c_1 > 0$. Constraining further t to satisfy

$$t \geq C_1 160^2 \sigma^2 \|P_\star\|^{10} \left(d_x\gamma \log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)\right) \tag{55}$$

ensures in addition that

$$\|A_t - A\| \leq \frac{1}{160\|P_\star\|^5}. \tag{56}$$

Now, using Lemma 22, we can find an universal positive constant $c_2 > 0$ such that if

$$t \geq c_2\sigma^2 \|P_\star\|^{10} \left(d_x\gamma \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x \gamma) + \log(e/\delta)\right), \tag{57}$$

then conditions (54) and (55) also hold.

We remind here that because $\mathcal{T}$ satisfies condition (5), if the constant $c_2 > 0$ is chosen large enough, we can claim that for all $t \in \mathbb{N}$ such that (54) holds, we have $K_t = K_{t_k}$ for some t_k that also satisfies (54). Therefore, applying Lemma 23 ensures that

$$\mathbb{P}(\forall t \geq t_1(\delta), \quad (53) \text{ and } (56)) \geq 1-\delta \tag{58}$$

with

$$t_1(\delta) = c_3\sigma^2 \|P_\star\|^{10} \left(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x \gamma) + \log(e/\delta)\right), \tag{59}$$

for some universal postive constant $c_3 > 0$. Finally, a direct application of Proposition 16 ensures that

$$\mathcal{C}_{1,\delta} = \left\{ \forall t \geq t_1(\delta), \quad \begin{matrix} (ii) & \max(\|B(K_t - K_\star)\|, \|K_t - K_\star\|) \leq \frac{1}{4\|P_\star\|^{3/2}} \\ (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \end{matrix} \right\} \tag{60}$$

holds with probability $1-\delta$.

Step 2: (Commitment to the certainty equivalence controller) Using Theorem 8, we have

$$\mathcal{D}_\delta = \left\{ \forall t \geq t_2(\delta), \quad (i)\ \widetilde{K}_t = K \right\} \tag{61}$$

holds with probability at least $1 - 5t_2(\delta)\delta$, with

$$t_2(\delta) = \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{3\gamma_\star} \tag{62}$$

such that $t_2(\delta) \geq t_1(\delta)$ (this can be ensured by taking the universal positive constants hidden in poly$(\cdot)$ large enough)

Step 3: (Refined error rate under the certainty equivalence controller) Now note that the event $\mathcal{C}_{1,\delta} \cap \mathcal{D}_{1,\delta}$ holds with probability at least $1 - 6t_2(\delta)\delta$. Therefore, applying Proposition 5 we obtain

$$\|A_t - A\| \leq \frac{c_4\sigma^2}{t} \left(d_x \log\left(e\|P_\star\| d_x\right) + \log(e/\delta)\right)$$

provided that

$$t \geq \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{6\gamma_\star^2}$$

for some universal positive constant $c_4 > 0$. Again, after using Lemma 23, and using the perturbation bounds via Proposition 16 we obtain that the event

$$\mathcal{C}_{2,\delta} = \left\{ \forall t \geq t_3(\delta), \quad \begin{matrix} (ii) & \|B(K_t - K_\star)\| \leq \frac{1}{4\|P_\star\|^{3/2}} \\ (iii) & \|P_\star(K_t) - P_\star\| \leq C_1(\delta) r_t^2 \\ (iv) & \|P_\star(K_t)\| \leq 2\|P_\star\| \end{matrix} \right\} \tag{63}$$

holds with probability at least $1 - c_5 t_3(\delta)\delta$, wehre we define

$$t_3(\delta) = \text{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{6\gamma_\star^2},$$
$$C_1(\delta) = c_6\sigma^2 \|P_\star\|^8 d_x \log\left(e\|P_\star\| d_x\right) \log(e/\delta),$$

for some universal positive constants $c_5, c_6 > 0$.

Step 4: (Stability under the certainty equivalence controller) Let us define the events

$$E_{1,\delta,T} = \left\{ \sum_{t=t_3(\delta)}^{T} r_t^2 \|x_t\|^2 \leq 8\|P_\star\|^{3/2} \left(\|r_{1:T}\|_\infty^2 \|x_{t_2(\delta)}\|^2 + 3\sigma^2 \left(\|r_{1:T}\|_2^2 + \log(e/\delta) \right) \right) \right\},$$
$$E_{2,\delta} = \{ \forall t \geq t_3(\delta), \quad (i) \text{ and } (ii) \text{ hold } \}.$$

Noting that $t_3(\delta)$ may be chosen so that $t_3(\delta) \geq d_x$, we obtain by a direct application Proposition 18, that

$$\mathbb{P}\left(E_{1,\delta,T} \cup E_{2,\delta}^c \right) \geq 1 - \delta.$$

Step 5: (Putting everything together) To conlcude, we note that under the event $\mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap (E_{1,\delta,T} \cup E_{2,\delta})$, the propoerties $(i)-(iv)$ hold for all $t \geq t_3(\delta)$. Additionally, under CEC($\mathcal{T}$), it also holds that $\|x_{t_3(\delta)}\|^2 \leq \sigma^2 d_x f(t_3(\delta))$, therefore it follows that

$$\begin{aligned}
\sum_{t=t_3(\delta)}^{T} r_t^2 \|x_t\|^2 &\leq 8\|P_\star\|^{3/2} \left(\|r_{t_3(\delta):T}\|_\infty^2 \|x_{t_3(\delta)}\|^2 + 3\sigma^2 \left(d_x \|r_{1:T}\|_2^2 + \log(e/\delta) \right) \right) \\
&\leq 8\|P_\star\|^{3/2} \left(\|r_{t_3(\delta):T}\|_\infty^2 \|x_{t_3(\delta)}\|^2 + 3\sigma^2 \left(d_x \|r_{1:T}\|_2^2 + \log(e/\delta) \right) \right) \\
&\leq 24\sigma^2 \|P_\star\|^{3/2} d_x \|r_{1:T}\|_2^2 + \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta) t_3(\delta)^{1+\gamma/2} \\
&\leq 24\sigma^2 \|P_\star\|^{3/2} d_x \|r_{1:T}\|_2^2 + \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{15\gamma_\star^2}
\end{aligned}$$

where we used the fact $\|r_{t:T}\|_\infty^2 \leq 1$. Thus, defining the property

$$\text{(v)} \sum_{t=t_3(\delta)}^{T} r_t^2 \|x_t\|^2 \leq C_3 \|r_{1:T}\|_2^2 + C_2(\delta),$$

where

$$\begin{aligned}
C_2(\delta) &= \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, \gamma_\star) \log(e/\delta)^{15\gamma_\star^2}, \\
C_3 &= 24\sigma^2 \|P_\star\|^{3/2} d_x,
\end{aligned}$$

we have shown that

$$\mathcal{E}_\delta = \{ \forall t \geq t_3(\delta), \ \text{(i)} - \text{(v)} \text{ hold} \} \subseteq \mathcal{C}_\delta \cap \mathcal{D}_\delta \cap E_{1,\delta,T}.$$

Finally, we note that $\mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap (E_{1,\delta,T} \cup E_{2,\delta}) = \mathcal{C}_{2,\delta} \cap \mathcal{D}_\delta \cap E_{1,\delta,T}$. Therefore, by a union bound, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_\delta) &\geq 1 - \mathbb{P}\left(\mathcal{C}_\delta^c \cup \mathcal{D}_\delta^c \cup (E_{1,\delta,t} \cup E_{d,\delta})^c \right) \\
&\geq 1 - c_7 t_3(\delta) \delta
\end{aligned}$$

for some universal positive constant $c_7 > 0$. This gives the desired result with modified universal constants. $\square$

E HYSTERESIS SWITCHING

In this appendix, we analyze the hysteresis switching scheme of CEC($\mathcal{T}$). The main result is stated in Theorem 8, and essentially says that eventually, CEC($\mathcal{T}$) just uses the certainty equivalence controller K_t. The proof of this result relies mainly on the consistency of the least squares estimator (see Appendix F), and on the perturbation bounds on Riccati equations (see Appendix I). The stability behaviour of the resulting dynamical system is also instrumental in the analysis (see Appendix J).

E.1 Main Result

The following theorem says that after time $t(\delta)$, CEC($\mathcal{T}$) only uses the certainty equivalence controller K_t. In its proof, we will use lemmas presented later in this appendix.

Theorem 8. *Assume that $\mathcal{T}$ satisfies (5). For all $\delta \in (0,1)$, there exists a stopping time $\upsilon(\delta)$, such that* CEC($\mathcal{T}$) *uses the certainty equivalence controller at $\upsilon(\delta)$, and such that the stopping time $\tau(\delta) = \inf\{t > \upsilon(\delta) :$* CEC($\mathcal{T}$) *uses stabilizing controller at time $t\}$ verifies*

$$\mathbb{P}(\upsilon(\delta) \leq t(\delta), \tau(\delta) = \infty) \geq 1 - 5t(\delta)\delta$$

where

$$t(\delta) = \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) \log(e/\delta)^{3\gamma_\star \beta}$$

and where the order of the polynomial only depends on γ, and $\beta = 4$, $\alpha = 0$ in Scenario I, $\beta = 2$ in Scenario III (A known), and $\beta = 1$ in Scenario II (B known). We note that in Scenarios I and II, we have $\mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, \mu_\star^{-1}, d_x, d_u, \gamma_\star) = \mathrm{poly}(\sigma, C_\circ, \mathcal{G}_\circ, \|P_\star\|, d_x, d_u, \gamma_\star)$ *since we have no more dependency on $\mu_\star$.*

Proof of Theorem 8. The proof for the three scenarios are very similar. We provide the proof for Scenario I, as for the other two scenarios, we simply highlight the differences in the proof.

Scenario I. By Lemma 4, we have for all $\delta \in (0,1)$, $\mathbb{P}(E_{1,\delta}) \geq 1 - \delta$, provided that

$$t_1(\delta)^{1/4} = c\sigma C_\circ^2 C_K^2 \|P_\star\|^{10} d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) \log(e/\delta)$$

for some universal positive constant $c_1 > 0$ large enough so that $t_1(\delta) \geq d_x + \log(e/\delta)$. We note here that we may have the crude upper bound $C_K \leq C_\circ \|P_\star\|$. Now defining

$$\upsilon(\delta) = \inf\{t > t_1(\delta) : \ \ell_t = 1\},$$
$$\tau(\delta) = \inf\{t > \upsilon(\delta) \ : \ \ell_t = 0\}.$$

and denoting

$$t(\delta) = c_2 (C_\circ \mathcal{G}_\circ)^{8/\gamma} \|P_\star\|^{3/\gamma} t_1(\delta)^{3\gamma_\star})$$

where $c_2 > 0$ is a universal positive constant. Note that $t(\delta) \geq c_2 (C_\circ \mathcal{G}_\circ)^{8/\gamma} t_1(\delta)^{3\gamma_\star}$ (recall that $\|P_\star\| \geq 1$). Thus, provided c_2 is large enough, we may apply Lemma 3 and obtain

$$\mathbb{P}(\upsilon(\delta) \leq t(\delta)) \geq 1 - \delta.$$

Furthermore, note that $t(\delta) \geq c_2 \|P_\star\|^{3/2} t_1(\delta)$. Hence, provided c_2 is large enough, we may apply Lemma 7, and obtain

$$\mathbb{P}(\upsilon(\delta) \leq t(\delta), \tau(\delta) < \infty) \leq 4t(\delta)\delta.$$

Therefore, we have

$$\begin{aligned}
\mathbb{P}(\upsilon(\delta) \leq t(\delta), \tau(\delta) = \infty) &= \mathbb{P}(\upsilon(\delta) \leq t(\delta)) - \mathbb{P}(\upsilon(\delta) \leq t(\delta), \tau(\delta) < \infty) \\
&\leq 1 - \delta - 4t(\delta)\delta \\
&\geq 1 - 5t(\delta)\delta,
\end{aligned}$$

where c_2 may be chosen sufficiently large so that $t(\delta) \geq 1$.

Scenario III (A known). We start by applying Lemma 6, to obtain that $\mathbb{P}(E_{2,\delta}) \geq 1-\delta$, provided that

$$t_1(\delta)^{1/2} = \frac{c\sigma^2 C_K \|P_\star\|^{10}}{\mu_\star^2}\left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ)\|P_\star\|\mu_\star^{-1} d_x d_u \gamma_\star) + \log(e/\delta)\right).$$

The remaining of the proof follows similarly as in Scenario I.

Scenario II (B known). We start by applying Lemma 5, to obtain that $\mathbb{P}(E_{3,\delta}) \geq 1-\delta$, provided that

$$t_1(\delta) = c\|P_\star\|^{10}(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x\gamma_\star) + \log(e/\delta)).$$

The remaining of the proof follows similarly as in Scenario I. □

E.2 The Time It Takes For CEC($\mathcal{T}$) To Use The Certainty Equivalence Controller

Lemma 3 quantifies the probability of switching to the certainty equivalence controller (i.e., $\widetilde{K}_t = K_t$) at some time, say between i and j. The proof of Lemma 3 relies on Lemma 15) and Proposition 15. The stabilizing controller (i.e., $\widehat{K}_t = K_\circ$) will eventually bring the system to a stable behaviour, thus to a state where it can attempt the certainty equivalence controller.

Lemma 3. *Under* CEC($\mathcal{T}$)*, for all* $\delta \in (0,1)$ *we have*

$$\mathbb{P}\left(\exists k \in \{i,\dots,j\}, \ell_k = 1\right) \geq 1-\delta$$

provided that

$$i \geq d_x + \log(e/\delta) \quad \text{and} \quad j \geq c(C_\circ \mathcal{G}_\circ)^{8/\gamma} i^{3\gamma_\star}$$

for some universal positive constant $c > 0$*. Refer to the pseudo-code of* CEC($\mathcal{T}$) *for the definition of* ℓ_k*.*

Proof of Lemma 3. We start by defining the following events.

$$\mathcal{E}_{i,j} = \{\exists k \in \{i,\dots,j\}, \quad \ell_k = 1\},$$
$$\mathcal{A}_{\delta,t} = \left\{\sum_{s=0}^{t} \|x_s\|^2 \leq C_1\sigma^2\mathcal{G}_\circ^2 C_\circ^2 (d_x t^{1+2\gamma} + \log(e/\delta))\right\}.$$

We have, by Proposition 15, that the event $\mathcal{A}_{\delta,t}$ holds with probability at least $1-\delta$ for some universal positive constant $C_1 > 1$. Our analysis is essentially the same for all the envisioned scenarios. Therefore, to avoid unnecessary rewriting, we denote

$$\forall s \geq 0: \quad \xi_s = \begin{cases} \nu_s & \text{if in Scenario I} \\ \zeta_s & \text{if in Scenario III} \\ 0 & \text{if in Scenario II} \end{cases}$$

where we remark that for all $s \geq 0$, ξ_s is a zero-mean, sub-gaussian random vector that has variance proxy at worst σ^2 for all $s \geq d_x$.

Now, provided that $i \geq \log(e/\delta)$, observe that under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{i,j}^c$ the following holds

(a) $\sum_{s=0}^{j} \|x_s\|^2 > \sigma^2 d_x f(j) = \sigma^2 d_x j^{1+\gamma/2}$,

(b) $\sum_{s=0}^{i} \|x_s\|^2 \leq C_1\sigma^2\mathcal{G}_\circ^2 C_\circ^2 (d_x i^{1+2\gamma} + \log(e/\delta)) \leq 2C_1\sigma^2\mathcal{G}_\circ^2 C_\circ^2 d_x i^{3\gamma_\star}$,

(c) $\forall s \in \{i,\dots,j\}$, $u_s = K_\circ x_s + \xi_s$.

Consider the following dynamical system

$$\forall s \geq 0, \qquad y_{s+1} = (A + BK_\circ)y_s + B\xi_{i+s} + \eta_{i+s} \quad \text{with} \quad y_0 = x_i$$

and note that when *(c)* holds, we have $(x_i, \dots, x_j) = (y_0, \dots, y_{j-i})$, thus, $\sum_{s=i}^{j} \|x_s\|^2 = \sum_{s=0}^{j-i} \|y_s\|^2$. On the other hand, the event

$$\mathcal{E}_{1,\delta,i,j} = \left\{ \sum_{s=0}^{j-i} \|y_s\|^2 \leq 2\mathcal{G}_\circ^2 \left(\|x_i\|^2 + 2\sigma^2 C_\circ^2 (d_x(j-i) + \log(e/\delta)) \right) \right\}$$

holds with probability at least $1 - \delta$ provided $i \geq d_x$. This follows from Lemma 15. Therefore, provided $i \geq d_x + \log(e/\delta)$, under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{i,j}^c \cap \mathcal{E}_{1,\delta,i,j}$, we have

$$\begin{aligned}
\sum_{s=0}^{j} \|x_s\|^2 &\leq \sum_{s=0}^{i} \|x_s\|^2 + \sum_{s=i}^{j} \|x_s\|^2 \\
&\leq 2C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x i^{3\gamma_\star} + \sum_{s=0}^{j-i} \|y_s\|^2 && \text{(Because (b) and (c) hold)} \\
&\leq 2C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x i^{3\gamma_\star} + 2\mathcal{G}_\circ^2 \|x_i\|^2 + 2\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x j && \text{(Under } \mathcal{E}_{1,\delta,i,j}\text{)} \\
&\leq 4C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^4 d_x (i^{3\gamma_\star} + j). && \text{(Because (b) holds)}
\end{aligned}$$

After some elementary calculations, we obtain that if $j \geq (8C_1 C_\circ^2 \mathcal{G}_\circ^4)^{2/\gamma}, j \geq i^{3\gamma_\star}$, and $i \geq d_x + \log(e/\delta)$, then, under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{i,j}^c \cap \mathcal{E}_{1,\delta,i,j}$, we have

$$\sum_{s=0}^{j} \|x_s\|^2 \leq \sigma^2 d_x f(j).$$

But this cannot hold under $\mathcal{E}_{i,j}^c$ otherwise it would contradict property *(a)*, therefore it must be that $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,\delta,i,j} \subseteq \mathcal{E}_{i,j}$. Hence by union bound, we have

$$\mathbb{P}\left(\mathcal{E}_{i,j}\right) \geq 1 - \mathbb{P}(\mathcal{A}_{\delta,t}^c \cup \mathcal{E}_{1,\delta,i,j}^c) \geq 1 - 2\delta$$

provided that

$$i \geq d_x + \log(e/\delta) \quad \text{and} \quad j \geq c(C_\circ \mathcal{G}_\circ)^{8/\gamma} i^{3\gamma_\star}$$

for some universal positive constant $c > 0$. Reparametrizing by $\delta' = 2\delta$ yields the desired result with modified positive constants. □

E.3 Consistency Of LSE Leads To Commitment

The event that leads to commitment. We note that the conditions under which CEC($\mathcal{T}$) switches to the certainty equivalence controller vary depending on which scenario we are in. Therefore, we are constrained to define the event that leads to commitment in each of the three envisioned scenarios.

For scenario I, we define for all $\delta \in (0,1)$, the event of interest as

$$E_{1,\delta} = \left\{ \forall t \geq t_1(\delta), \quad \begin{array}{ll} (i) & \|B(K_t - K_\star)\| \leq \frac{1}{\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t) \\ (iii) & \lambda_{\min}\left(\sum_{s=0}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top \right) \geq t^{1/4} \end{array} \right\}. \tag{64}$$

The definition of the time $t_1(\delta)$ is made precise in the following result, proved in E.5.

Lemma 4. *For all $\delta \in (0,1)$, let $E_{1,\delta}$ be defined as in (64). Under Algorithm* CEC($\mathcal{T}$)*, assuming that $\mathcal{T}$ satisifes (5), then, for all $\delta \in (0,1)$,*

$$\mathbb{P}\left(E_{1,\delta}\right) \geq 1-\delta,$$

provided that

$$t_1(\delta)^{1/4} = c\sigma C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ)\|P_\star\| d_x d_u \gamma_\star\right) + \log(e/\delta)\right).$$

For Scenario III, we define for all $\delta \in (0,1)$, the event of interest as

$$E_{2,\delta} = \left\{ \forall t \geq t_2(\delta), \quad \begin{array}{ll} (i) & \|B(K_t - K_\star)\| \leq \frac{1}{\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t) \\ (iii) & \lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq t^{1/2} \end{array} \right\}. \tag{65}$$

Lemma 5. *For all $\delta \in (0,1)$, let $E_{2,\delta}$ be defined as in (65). Under Algorithm* CEC($\mathcal{T}$)*, assuming that $\mathcal{T}$ satisifes (5), then, for all $\delta \in (0,1)$,*

$$\mathbb{P}\left(E_{2,\delta}\right) \geq 1-\delta,$$

provided that

$$t_2(\delta)^{1/2} = \frac{c\sigma^2 C_K \|P_\star\|^{10}}{\mu_\star^2} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ)\|P_\star\| \mu_\star^{-1} d_x d_u \gamma_\star\right) + \log(e/\delta)\right),$$

For Scenario II,

$$E_{3,\delta} = \left\{ \forall t \geq t_3(\delta), \quad \begin{array}{ll} (i) & \|B(K_t - K_\star)\| \leq \frac{1}{\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t) \end{array} \right\}. \tag{66}$$

Lemma 6. *For all $\delta \in (0,1)$, let $E_{3,\delta}$ be defined as in (66). Under Algorithm* CEC($\mathcal{T}$)*, assuming that $\mathcal{T}$ satisifes (5), then, for all $\delta \in (0,1)$,*

$$\mathbb{P}\left(E_{3,\delta}\right) \geq 1-\delta,$$

provided that

$$t_3(\delta) = c\|P_\star\|^{10}(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x \gamma_\star) + \log(e/\delta)).$$

The proofs of Lemma 4, Lemma 5, and Lemma 6 are presented in E.5. They rely on the consistency of the Least squares algorithm under CEC($\mathcal{T}$) when $\mathcal{T}$ satisfies (5).

In the definitions of the events $E_{1,\delta}$, $E_{2,\delta}$, and $E_{3,\delta}$, property (i) is the most important for establishing the commitment lemma that we shall provide shortly.

E.4 The Commitment Lemma

Lemma 7 states that CEC($\mathcal{T}$) will eventually only use the certainty equivalence conotroller, provided that K_t is sufficiently close to $K_\star$ (this is captured by property (i) in the aforementioned events).

Lemma 7. *Assume that $\mathcal{T}$ satisfies (5). Assume that for all $\delta \in (0,1)$,*

$$\mathbb{P}(E_{p,\delta}) \geq 1-\delta$$

for some $t_p(\delta) \geq d_x + \log(e/\delta)$ for $p \in \{1,2,3\}$ and where $E_{1,\delta}$, $E_{2,\delta}$, and $E_{3,\delta}$ are defined in (64), (66), and (65), respectively. Define the following stopping times

$$\upsilon(\delta) = \inf\left\{t > t_p(\delta) : \ell_t = 1\right\},$$
$$\tau(\delta) = \inf\left\{t > \upsilon(\delta) : \ell_t = 0\right\}.$$

Then for all $\delta \in (0,1)$, we have

$$\mathbb{P}\left(\upsilon(\delta) \leq k, \tau(\delta) < \infty\right) \leq 4k\delta$$

provided that $k \geq 18^{2/\gamma}\|P_\star\|^{3/\gamma} t_p(\delta)$ for some universal positive constant $c > 0$.

Proof of Lemma 7. For ease of notation, we drop the dependency of $\upsilon(\delta)$, and $\tau(\delta)$ on δ, and simply write υ, and τ. Furthermore, we shall refer to

$$\forall s \geq 0 \quad \xi_s = \begin{cases} \nu_s & \text{if in Scenario I} \\ 0 & \text{if in Scenarios II or III.} \end{cases}$$

We note that ξ_s is a zero-mean, sub-gaussian random vector with variance proxy at most σ^2 provided $s \geq d_x$. Fix $p \in \{1, 2, 3\}$

We have

$$\begin{aligned}
\mathbb{P}\left(\upsilon \leq k, \tau < \infty\right) &\leq \mathbb{P}\left(\{\upsilon \leq k, \tau < \infty\} \cap E_{p,\delta}\right) + \mathbb{P}\left(\mathcal{E}_{1,\delta}^c\right) \\
&\leq \mathbb{P}\left(\{\upsilon \leq k, \tau < \infty\} \cap E_{p,\delta}\right) + \delta \\
&\leq \sum_{i=t(\delta)+1}^{k} \mathbb{P}(\{\upsilon = i, \tau < \infty\} \cap E_{p,\delta}) + \delta \\
&\leq \sum_{i=t(\delta)+1}^{k} \sum_{j=i+1}^{\infty} \mathbb{P}(\{\upsilon = i, \tau = j\} \cap E_{p,\delta}) + \delta.
\end{aligned}$$

Computing the probabilities $\mathbb{P}(\{\upsilon = i, \tau = j\} \cap E_{p,\delta})$**.** Let $j > i$. First, let us note that under the event $\{\upsilon = i, \tau = j\} \cap E_{p,\delta}$, the following must hold

(a) For all $i \leq t < j$, $u_t = K_t x_t + \xi_t$ (since $\ell_t = 0$ and conditions *(ii)-(iii)* hold),

(b) $\sum_{s=0}^{i} \|x_s\|^2 < \sigma^2 d_x f(i) = \sigma^2 d_x i^{1+\gamma/2}$,

(c) $\sum_{s=0}^{j} \|x_s\|^2 > \sigma^2 d_x g(j) = \sigma^2 d_x j^{1+\gamma}$.

First, we use a truncation trick, and define

$$\forall t \geq 0, \quad \widetilde{K}_t = (K_t - K_\star) 1_{\left\{\|K_{i+s} - K_\star\| \leq \frac{1}{4\|P_\star\|^{3/2}}\right\}} + K_\star$$

and note that under the event $E_{p,\delta}$, we have $K_t = \widetilde{K}_t$ for all $t \geq t(\delta)$. Now consider the following dynamical system

$$\forall s \geq 1, \quad y_{s+1} = (A + B\widetilde{K}_{i+s}) y_s + B\xi_{i+s} + \eta_{i+s} \quad \text{with} \quad y_0 = x_i.$$

We note that under the event $E_{p,\delta}$, we have $(x_i, \ldots, x_j) = (y_0, \ldots, y_{j-i})$. On the other hand, the event

$$\mathcal{E}_{2,\delta,i,j} = \left\{ \sum_{s=0}^{j-i} \|y_s\|^2 \leq 8\|P_\star\|^{3/2} \left(\|x_i\|^2 + 6\sigma^2 C_\circ^2 (d_x(j-i) + \log(j^2/\delta)) \right) \right\}$$

holds with probability at least $1 - \delta/j^2$, provided that $i \geq d_x$. This follows by Lemma 16 (see also Proposition 18). Therefore, under the event $E_{p,\delta} \cap \mathcal{E}_{2,\delta,i,j} \cap \{\upsilon = i, \tau = j\}$, we have

$$\begin{aligned}
\sum_{s=0}^{j} \|x_s\|^2 &\leq \sum_{s=0}^{i} \|x_s\|^2 + \sum_{s=i}^{j} \|x_s\|^2 \\
&\leq \sigma^2 d_x i^{1+\gamma/2} + \sum_{s=i}^{j} \|x_s\|^2 && \text{(Because (b) holds)} \\
&\leq \sigma^2 d_x i^{1+\gamma/2} + 8\|P_\star\|^{3/2} \|x_i\|^2 + 6\sigma^2 C_\circ^2 (d_x + 2) j && \text{(Under } \mathcal{E}_{2,\delta,i,j} \cap E_{p,\delta}\text{)} \\
&\leq 9\sigma^2 \|P_\star\|^{3/2} d_x i^{1+\gamma/2} + 3\sigma^2 C_\circ^2 d_x j && \text{(Because (a) holds)}
\end{aligned}$$

provided that $i \geq \log(e/\delta)$.

After some elementary computations, we obtain that if $j > 18^{2/\gamma}\|P_\star\|^{3/\gamma}$, and $i \geq \log(e/\delta)$, then under the event $E_{p,\delta} \cap \mathcal{E}_{2,\delta,i,j} \cap \{\upsilon = i, \tau = j\}$, we have

$$\sum_{s=0}^{j} \|x_s\|^2 \leq \sigma^2 d_x j^{1+\gamma}.$$

But this cannot be under the event $\{\upsilon = i, \tau = j\} \cap E_{p,\delta}$, otherwise it would contradict *(c)*. Therefore, it must be that $\{\upsilon = i, \tau = j\} \cap E_{p,\delta} \subseteq \mathcal{E}^c_{2,\delta,i,j}$. Hence

$$\mathbb{P}(\{\upsilon = i, \tau = j\} \cap E_{p,\delta}) \leq \mathbb{P}(\mathcal{E}^c_{2,\delta,i,j}) \leq \frac{\delta}{j^2}$$

provided that

$$i \geq d_x + \log(e/\delta) \qquad \text{and} \qquad j > 18^{2/\gamma}\|P_\star\|^{3/\gamma}.$$

Let us remind here that $j > i$ and $i > t(\delta)$.

Concluding step. To conclude, provided that $k \geq 18^{2/\gamma}\|P_\star\|^{3/\gamma}(d_x + \log(e/\delta))$, we have

$$\begin{aligned}
\mathbb{P}\left(\upsilon \leq k, \tau(\delta) < \infty\right) &\leq \sum_{i=t(\delta)+1}^{k} \sum_{j=i+1}^{\infty} \mathbb{P}(\upsilon = i, \tau = j, E_{p,\delta}) + \delta \\
&\leq \sum_{i=t(\delta)+1}^{k} \sum_{j=i+1}^{\infty} \frac{\delta}{j^2} + \delta \\
&\leq \frac{\pi^2 k\delta}{6} + \delta \leq 4k\delta.
\end{aligned}$$

□

E.5 Remaining Proofs

Proof of Lemma 4. Using Lemma 8, we have already established that:

$$\mathbb{P}\left(\forall t \geq t_4(\delta) : \|B(K_t - K_\star)\| \leq \frac{1}{5\|P_\star\|^{3/2}} \quad \text{and} \quad \|K_t\|^2 \leq h(t)\right) \geq 1 - \delta \tag{67}$$

with

$$t_4(\delta)^{1/4} = c_1 \sigma C_\circ^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right) \tag{68}$$

for some universal positive constant $c_1 > 0$. Now we may use Proposition 11 to obtain that:

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right) \geq \frac{C_2 \sigma^2 \sqrt{d_x t}}{C_K^2}\right) \geq 1 - \delta$$

provided that $t^{1/4} \geq c_2 \sigma C_\circ^2 C_K^2 \|P_\star\|^{10}(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$ for some universal positive constants $C_2, c_2 > 0$. From which we may conclude that

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right) \geq t^{1/4}\right) \geq 1 - \delta$$

provided that $t^{1/4} \geq c_3 \sigma C_\circ^2 C_K^2 \|P_\star\|^{10}(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$, for some universal positive constant $c_3 > 0$ that is chosen to be large enough so that $\frac{C_2 \sigma^2 \sqrt{d_x t}}{C_K^2} \geq t^{1/4}$. Next, we apply Lemma 23 to obtain the following bound that holds uniformly over time.

$$\mathbb{P}\left(\forall t \geq t_5(\delta), \quad \lambda_{\min}\left(\sum_{s=0}^{t-1} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right) > t^{1/4}\right) \geq 1 - \delta \tag{69}$$

where

$$t_5(\delta)^{1/4} = c_4 \sigma C_\circ^2 C_K^2 \|P_\star\|^{10} (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)) \tag{70}$$

for some universal constant $c_4 > 0$. Now, using a union bound, we obtain from (72), and (74) that

$$\mathbb{P}\left(E_{1,\delta}\right) \geq 1 - 2\delta \tag{71}$$

with

$$t_1(\delta)^{1/4} = c\sigma C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ) \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right)$$

for some universal positive constant $c > 0$ chosen large enough so that $t_1(\delta) \geq \max(t_4(\delta), t_5(\delta))$. □

Proof of Lemma 5. Using Lemma 8, we have already established that:

$$\mathbb{P}\left(\forall t \geq t_4(\delta) : \|B(K_t - K_\star)\| \leq \frac{1}{5\|P_\star\|^{3/2}} \quad \text{and} \quad \|K_t\|^2 \leq h(t)\right) \geq 1 - \delta \tag{72}$$

with

$$t_4(\delta)^{1/2} = c\|P_\star\|^{10} (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)) \tag{73}$$

for some universal positive constant $c_1 > 0$. Now we may use Proposition 13 to obtain that:

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq \frac{\mu_\star^2 t}{10}\right) \geq 1 - \delta$$

provided that $t^{1/2} \geq c_2 \frac{\sigma^2 C_K \|P_\star\|^{10}}{\mu_\star} (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$ for some universal positive constants $c_2 > 0$. From which we may conclude that

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq t^{1/2}\right) \geq 1 - \delta$$

provided that $t^{1/4} \geq \frac{c_3 \sigma^2 C_K \|P_\star\|^{10}}{\mu_\star^2} (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$, for some universal positive constant $c_3 > 0$ that is chosen to be large enough so that $\frac{\mu_\star^2 t}{10} \geq t^{1/2}$. Next, we apply Lemma 23 to obtain the following bound that holds uniformly over time.

$$\mathbb{P}\left(\forall t \geq t_5(\delta) : \quad \lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) > t^{1/2}\right) \geq 1 - \delta \tag{74}$$

where

$$t_5(\delta)^{1/2} = \frac{c_4 \sigma^2 C_K \|P_\star\|^{10}}{\mu_\star^2} (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| \mu_\star^{-1} d_x d_u \gamma_\star) + \log(e/\delta)) \tag{75}$$

for some universal constant $c_4 > 0$. Now, using a union bound we obtain from (72), and (74) that

$$\mathbb{P}\left(E_{2,\delta}\right) \geq 1 - 2\delta \tag{76}$$

with

$$t_2(\delta)^{1/2} = \frac{c\sigma^2 C_K \|P_\star\|^{10}}{\mu_\star^2} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ) \|P_\star\| \mu_\star^{-1} d_x d_u \gamma_\star) + \log(e/\delta)\right)$$

for some universal positive constant $c > 0$ chosen large enough so that $t_2(\delta) \geq \max(t_4(\delta), t_5(\delta))$. □

Proof of Lemma 6. Let $t \geq 0$ and $\delta \in (0,1)$. Assume the following condition holds.

$$t^{1/2} \geq C160^2 \|P_\star\|^{10} \left(d_x \gamma \log\left(\sigma C_\circ \mathcal{G}_\circ d_x t\right) + \log(e/\delta)\right). \tag{77}$$

Then, by Proposition 6, if the condition

$$t \geq c_1 (d_x \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)), \tag{78}$$

also holds, then we have

$$\mathbb{P}\left(\|A_t - A\|^2 \leq \frac{1}{160^2 \|P_\star\|^{10}} \right) \geq 1 - \delta. \tag{79}$$

Note that $\mathcal{T}$ satisfies condition (5). Therefore, provided that the constant $c_1 > 0$ is chosen large enough, we can claim that for all $t \in \mathbb{N}$ such that (77) and (78) hold, we have $K_t = K_{t_k}$ for some t_k that also satisfies (77) and (86).

Using Propostion 16, we may conclude that provided (77) and (78) hold, we have

$$\mathbb{P}\left(\max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}} \right) \geq 1 - \delta.$$

By Lemma 22, we can find a universal positive constant $c_2 > 0$ such that

$$t \geq c_2 \|P_\star\|^{10} (d_x \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x \gamma_\star) + \log(e/\delta))$$

implies that the conditions (77) and (78) hold, and $\|K_t\|^2 \leq h(t)$. Now using Lemma 23 gives that

$$\mathbb{P}(E_{3,\delta}) \geq 1 - \delta$$

where

$$t_3(\delta) = c\|P_\star\|^{10} (d_x \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x \gamma_\star) + \log(e/\delta))$$

for some universal positive constant $c > c_2$. □

F THE LEAST SQUARES ESTIMATOR AND ITS ERROR RATE

In this appendix, we study the performance of the Least Squares Estimator (LSE) of (A, B). We first provide the pseudo-code of the algorithm giving this estimator, see Algorithm 2. The error rate of the LSE in Scenario I is characterized in Propositions 1 and 2. The error rate is analyzed for Scenario III in Proposition 3, 4, and 5. Propositions 6 and 7 upper bound the error rate in Scenario II. It is worth mentioning that these results are established without the use of a doubling trick that would essentially mean that the controller is fixed. Here, in CEC($\mathcal{T}$), we allow the controller to change over time. Further note that the results hold under CEC($\mathcal{T}$) regardless of how $\mathcal{T}$ is chosen provided it satisifes (5). To derive upper bounds on the LSE error rate, we extensively exploit the results related to the smallest eigenvalue of the covariates matrix, presented in the next appendix.

The performance of CEC($\mathcal{T}$) depends on the error rate of the LSE, but also on how well the certainty equivalence controller K_t approximates the optimal controller $K_\star$. In Lemmas 8 and 9, we present upper bounds of $\|K_t - K_\star\|$ in the different scenarios. These lemmas are also used to refine the error rates of the LSE.

F.1 Pseudo-code Of The LSE

Algorithm 2: Least Squares Estimation (LSE)

input : Sample path $(x_0, u_0, \dots, x_{t-1}, u_{t-1}, x_t)$ and the cost matrices Q and R
output : Estimator (A_t, B_t) of (A, B)

if *B known* **then**
$A_t \leftarrow \left(\sum_{s=0}^{t-1}(x_{s+1} - Bu_s)x_s^\top\right)\left(\sum_{s=0}^{t-1} x_s x_s^\top\right)^\dagger$;
end
if *A known* **then**
$B_t \leftarrow \left(\sum_{s=0}^{t-2}(x_{s+1} - Ax_s)u_s^\top\right)\left(\sum_{s=0}^{t-2} u_s u_s^\top\right)^\dagger$;
end
if *(A, B) are unkown* **then**
$\begin{bmatrix} A_t & B_t \end{bmatrix} \leftarrow \left(\sum_{s=0}^{t-2} x_{s+1}\begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right)\left(\sum_{s=0}^{t-2}\begin{bmatrix} x_s \\ u_s \end{bmatrix}\begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right)^\dagger$;
end

F.2 Error Rate In Scenario I

Proposition 1. *Under* CEC($\mathcal{T}$)*, for all* $\delta \in (0, 1)$*,*

$$\max(\|A_t - A\|^2, \|B_t - B\|^2) \le \frac{C\sigma^2\left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x d_u t\right) + \log(e/\delta)\right)}{t^{1/4}}$$

holds with probability at least $1 - \delta$*, provided that* $t \ge c\sigma^4 C_\circ^8 (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta))$ *for some positive constants* $C, c > 0$*.*

Proof of Proposition 1. Fix $t \ge 2$. We start by writing the LSE error

$$\begin{bmatrix} A_t - A & B_t - B \end{bmatrix} = \left(\sum_{s=0}^{t-2} \eta_s \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right)\left(\sum_{s=0}^{t-2}\begin{bmatrix} x_s \\ u_s \end{bmatrix}\begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top\right)^\dagger.$$

For ease of notation, let $y_s = \begin{bmatrix} x_s \\ u_s \end{bmatrix}$ for all $s \ge 1$. Provided the $\sum_{s=0}^{t-2} y_s y_s^\top$ is invertible, we can decompose the error as

$$\left\|\begin{bmatrix} A_t - A & B_t - B \end{bmatrix}\right\|^2 \le \left\|\left(\sum_{s=0}^{t-2} y_s y_s^\top\right)^{-1/2}\left(\sum_{s=0}^{t-2} y_s \eta_s^\top\right)\right\|^2 \frac{1}{\lambda_{\min}\left(\sum_{s=0}^{t-2} y_s y_s^\top\right)}.$$

Step 1: (Bounding the smallest eigenvalue) Define the event

$$\mathcal{B}_t = \left\{ \lambda_{\min}\left(\sum_{s=0}^{t-2} y_s y_s^\top\right) \geq \left(\frac{t}{4}\right)^{1/4} \right\}.$$

Using Proposition 10, we know that the event $\mathcal{B}_t$ holds with probability at least $1-\delta$ provided

$$(\star) \quad t \geq c_1 \sigma^4 C_\circ^8 (\gamma_\star d \log(e\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta)),$$

for some $c_1 > 0$. Under the event $\mathcal{B}_t$, the estimation error may be upper bounded as

$$\begin{aligned}
\left\| \begin{bmatrix} A_t - A & B_t - B \end{bmatrix} \right\|^2 &\leq \frac{\sqrt{2}}{t^{1/4}} \left\| \left(\sum_{s=0}^{t-2} y_s y_s^\top\right)^{-1/2} \left(\sum_{s=0}^{t-2} y_s \eta_s^\top\right) \right\|^2, \\
&\leq \frac{2\sqrt{2}}{t^{1/4}} \left\| \left(\sum_{s=0}^{t-2} y_s y_s^\top + \frac{t^{1/4}}{\sqrt{2}} I_{d_x+d_u}\right)^{-1/2} \left(\sum_{s=0}^{t-2} y_s \eta_s^\top\right) \right\|^2,
\end{aligned}$$

where we used, for the last inequality, the fact that $2\sum_{s=0}^{t-2} y_s y_s^\top \succeq \sum_{s=0}^{t-2} y_s y_s^\top + \frac{t^{1/4}}{\sqrt{2}} I_{d_x+d_u}$.

Step 2: (Bounding the self-normalized term) Define the events

$$\begin{aligned}
\mathcal{E}_{1,t,\delta} &= \left\{ \left\| \left(\sum_{s=0}^{t-2} y_s y_s^\top + \frac{t^{1/4}}{\sqrt{2}} I_{d_x+d_u}\right)^{-1/2} \left(\sum_{s=0}^{t-2} y_s \eta_s^\top\right) \right\|^2 \right. \\
&\qquad \left. \leq 7\sigma^2 \log\left(\frac{e^d \det\left(\frac{\sqrt{2}}{t^{1/4}} \sum_{s=0}^{t-2} y_s y_s^\top + I_d\right)}{\delta} \right) \right\}, \\
\mathcal{E}_{2,t,\delta} &= \left\{ \sum_{s=0}^{t-2} \|\nu_s\|^2 \leq \sigma^2 \sqrt{d_x}(2d_u t + 3\log(e/\delta)) \right\}, \\
\mathcal{A}_{t,\delta} &= \left\{ \sum_{s=0}^{t} \|x_s\|^2 \leq c_2 \sigma^2 C_\circ^2 \mathcal{G}_\circ^2 (d_x t^{1+2\gamma} + \log(1/\delta)) \right\}.
\end{aligned}$$

Then by Proposition 15, the event $\mathcal{A}_{t,\delta}$ holds with probability at least $1-\delta$ for some absolution positive constant $c_2 > 0$. By proposition 9, the event $\mathcal{E}_{1,t}$ holds with probability at least $1-\delta$, and by Hanson-Wright inequality (see Proposition 19), the event $\mathcal{E}_{2,t,\delta}$ holds with probability $1-\delta$. Under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t,\delta} \cap \mathcal{E}_{2,t,\delta}$, we have

$$\begin{aligned}
\left(\det\left(\frac{\sqrt{2}}{t^{1/4}} \sum_{s=0}^{t-2} y_s y_s^\top + I_d \right) \right)^{1/d} &\leq \frac{\sqrt{2}}{t^{1/4}} \sum_{s=0}^{t-2} \|y_s\|^2 + 1 \\
&\leq \frac{\sqrt{2}}{t^{1/4}} \sum_{s=0}^{t-2} \|x_s\|^2 + \|u_s\|^2 + 1 \\
&\leq \frac{\sqrt{2}}{t^{1/4}} \sum_{s=0}^{t-2} \|x_s\|^2 + \|\widetilde{K}_s\|^2 \|x_s\|^2 + \|\nu_s\|^2 + 1 \\
&\leq \sqrt{2}(4c_2+6)\sigma^2 C_\circ^4 \mathcal{G}_\circ^2 d_x d_u t^{3/4+5\gamma/2} \\
&\leq \sqrt{2}(4c_2+6)\sigma^2 C_\circ^4 \mathcal{G}_\circ^2 d_x d_u t^{4\gamma_\star},
\end{aligned}$$

where we used $\|\widetilde{K}_s\|^2 \leq C_K^2 h(s)$, assumed that

$$(\star\star) \quad t \geq \log(e/\delta),$$

and used $\gamma_\star = \max(1, \gamma)$ to obtain the last two inequalities. Therefore, assuming $(\star\star)$, we have under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{t,\delta}$ that

$$\log\left(\frac{e^d \det\left(\frac{\sqrt{2}}{t^{1/4}}\sum_{s=0}^{t-2} y_s y_s^\top + I_d\right)}{\delta}\right) \leq c_3\sigma^2\left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x d_u t\right) + \log(e/\delta)\right),$$

for some universal positive constants.

Step 3: (Putting everything together) To conclude, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{B}_t \cap \mathcal{E}_{t,\delta}$,

$$\max(\|A_t - A\|^2, \|B_t - B\|^2) \leq \frac{C\sigma^2\left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x d_u t\right) + \log(e/\delta)\right)}{t^{1/4}}, \tag{80}$$

where $C = 2\sqrt{2}c_3$. Therefore, the upper bound (80) holds with probability $1 - 3\delta$ when $(\star)$ and $(\star\star)$ hold. These two conditions hold whenever

$$t \geq c\sigma^4 C_\circ^8 (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta)),$$

for c is large enough. Reparametrizing by $\delta' = \delta/3$ yields the desired guarantee with modified universal positive constants. □

Lemma 8. *Assume that $\mathcal{T}$ satisfies (5). Then, under Algorithm* $\mathrm{CEC}(\mathcal{T})$*, we have, for all $\delta \in (0,1)$,*

$$\mathbb{P}\left(\forall t \geq t(\delta), \quad \begin{matrix}(i) & \max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t)\end{matrix}\right) \geq 1 - \delta$$

where $t(\delta)^{1/4} = c\sigma C_\circ^2 \|P_\star\|^{10}(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$, for some universal positive constant $c > 0$.

Proof of Lemma 8. Let $t \geq 0$ and $\delta \in (0,1)$. Assume that:

$$t^{1/4} \geq C160^2\sigma^2\|P_\star\|^{10}\left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x d_u t\right) + \log(e/\delta)\right). \tag{81}$$

Then, by Proposition 1, if the condition

$$t \geq c_1\sigma^4 C_\circ^8 (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta)), \tag{82}$$

also holds, then we have

$$\mathbb{P}\left(\max(\|A_t - A\|^2, \|B_t - B\|^2) \leq \frac{1}{160^2\|P_\star\|^{10}}\right) \geq 1 - \delta. \tag{83}$$

Since $\mathcal{T}$ satisfies condition (5), if the constant $c_1 > 0$ is chosen large enough, we can claim that for all $t \in \mathbb{N}$ such that (81) and (82) hold, we have $K_t = K_{t_k}$ for some t_k that also satisfies (81) and (82).

Using the perturbation bounds of Propostion 16, we conclude that, provided (81) and (82) hold, we have

$$\mathbb{P}\left(\max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}}\right) \geq 1 - \delta.$$

By Lemma 22, we can find a universal positive constant $c_2 > 0$ such that

$$t^{1/4} \geq c_2\sigma^2 C_\circ^2 \|P_\star\|^{10}(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$$

implies that the conditions (81) and (82) hold, and $\|K_t\|^2 \leq h(t)$. Finally, to obtain a bound that holds uniformly over time, we use Lemma 23 and obtain

$$\mathbb{P}\left(\forall t \geq t(\delta), \quad \begin{matrix}(i) & \max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t)\end{matrix}\right) \geq 1 - \delta,$$

where $t(\delta)^{1/4} = c\sigma^2 C_\circ^2 \|P_\star\|^{10}(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$ for some universal positive constant $c > c_2$. □

Proposition 2. *Assume that $\mathcal{T}$ satisfies (5). Under algorithm* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$,*

$$\max(\|A_t - A\|^2, \|B_t - B\|^2) \leq \frac{CC_K^2 \left(d\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x t\right) + \log(e/\delta)\right)}{(d_x t)^{1/2}}$$

holds with probability at least $1-\delta$, when

$$t^{1/4} \geq c\sigma C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right),$$

for some positive constants $C, c > 0$.

Proof. The proof differs from that of Proposition 1 only in the second step, where instead we define the event

$$\mathcal{B}_t = \left\{ \lambda_{\min}\left(\sum_{s=0}^{t-2} \begin{bmatrix} x_s \\ u_s \end{bmatrix} \begin{bmatrix} x_s \\ u_s \end{bmatrix}^\top \right) \geq \frac{C\sigma^2 \sqrt{d_x t}}{C_K^2} \right\},$$

where $C > 0$ is some universal constant to be defined through Proposition 11. Indeed, using Lemma 8, we may apply Proposition 11 which guarantees that $\mathcal{B}_t$ holds with probability at least $1-\delta$, provided that

$$t^{1/4} \geq c\sigma^2 C_\circ^2 C_K^2 \|P_\star\|^{10} \left(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right)$$

for some positive constant $c > 0$. The remaining steps are identical to those of the proof of Proposition 1. □

F.3 Error Rate In Scenario III

Proposition 3. *Under* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$,*

$$\|B_t - B\| \leq \frac{C\sigma^2 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u t) + d_x + \log(e/\delta)\right)}{t^{1/2}}$$

with probability at least $1-\delta$, provided that

$$t \geq c_1 \sigma^2 (d_x \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta))$$

for some universal positive constants $C, c > 0$.

Proof of Proposition 3. Fix $t \geq 2$. We start by writing the estimation error as

$$B_t - B = \left(\sum_{s=0}^{t-2} \eta_s u_s^\top \right) \left(\sum_{s=0}^{t-2} u_s u_s^\top \right)^\dagger.$$

When $\sum_{s=0}^{t-2} u_s u_s^\top$ is invertible, we can decompose the estimation error as

$$\|B_t - B\|^2 \leq \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2 \frac{1}{\lambda_{\min}\left(\sum_{s=0}^{t-2} u_s u_s^\top \right)}.$$

Step 1: (Bounding the smallest eigenvalue) Define the event

$$\mathcal{B}_t = \left\{ \lambda_{\min}\left(\sum_{s=0}^{t} u_s u_s^\top \right) \geq \left(\frac{t}{2} \right)^{1/2} \right\}.$$

Using Proposition 12, we have the event $\mathcal{B}_t$ holds with probability at least $1-\delta$, if the following condition $(\star)$ hold:

$$(\star) \quad t \geq c_1 (d_u \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)).$$

Under the event $\mathcal{B}_t$, the estimation error may be upper bounded as

$$\begin{aligned}\|B_t - B\|^2 &\leq \frac{\sqrt{2}}{t^{1/2}} \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2 \\ &\leq \frac{2\sqrt{2}}{t^{1/2}} \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top + \frac{t^{1/2}}{\sqrt{2}} \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2,\end{aligned}$$

where we used the fact that $2\sum_{s=0}^{t-2} u_s u_s^\top \succeq \sum_{s=0}^{t-2} u_s u_u^\top + \frac{t^{1/2}}{\sqrt{2}}$.

Step 2: (Bounding the self-normalized term) Consider the events,

$$\begin{aligned}\mathcal{E}_{1,t,\delta} &= \left\{ \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top + \frac{t^{1/2}}{\sqrt{2}} I_{d_u} \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2 \right. \\ &\qquad \left. \leq 7\sigma^2 \log \left(\frac{e^d \det \left(\frac{\sqrt{2}}{t^{1/2}} \sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u} \right)}{\delta} \right) \right\}, \\ \mathcal{E}_{2,t,\delta} &= \left\{ \sum_{s=0}^{t-2} \|\zeta_s\|^2 \leq \sigma^2 (2d_u t + 3\log(1/\delta)) \right\}, \\ \mathcal{A}_{t,\delta} &= \left\{ \sum_{s=0}^{t} \|x_s\|^2 \leq c_2 \sigma^2 C_\circ^2 \mathcal{G}_\circ^2 (d_x t^{1+2\gamma} + \log(1/\delta)) \right\}.\end{aligned}$$

By Proposition 15, the event $\mathcal{A}_{t,\delta}$ holds with probability at least $1-\delta$ for some universal positive constant $c_2 > 0$. By Proposition 9, the event $\mathcal{E}_{1,t,\delta}$ holds with probability at least $1-\delta$, and by Hanson-Wright inequality (See Proposition 19) the event $\mathcal{E}_{2,t,\delta}$ holds with probability at least $1-\delta$. Under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t,\delta}$ we have

$$\begin{aligned}\left(\det \left(\frac{\sqrt{2}}{t^{1/2}} \sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u} \right) \right)^{\frac{1}{d}} &\leq \frac{\sqrt{2}}{t^{1/2}} \sum_{s=0}^{t-2} \|u_s\|^2 + 1 \\ &\leq \frac{\sqrt{2}}{t^{1/2}} \sum_{s=0}^{t-2} \|\widetilde{K}_s\|^2 \|x_s\|^2 + \|\zeta_s\|^2 + 1 \\ &\leq \sqrt{2} C_\circ^2 t^{(\gamma-1)/2} \sum_{s=0}^{t-2} \|x_s\|^2 + \|\zeta_s\|^2 + 1 \\ &\leq 2\sqrt{2}(c_2+3)\sigma^2 C_\circ^4 \mathcal{G}_\circ^2 d_x d_u t^{(1+3\gamma)/2},\end{aligned}$$

where in the last two inequalities, we used the fact $\|\widetilde{K}_s\|^2 \leq \|K_\circ\|^2 h(t)$, we assumed that

$$(\star\star) \quad t \geq \log(e/\delta)$$

holds and used $\gamma_\star = \max(\gamma, 1)$. Therefore, assuming that $(\star\star)$ holds, under the event $\mathcal{E}_{1,t,\delta} \cap \mathcal{E}_{2,t,\delta} \cap \mathcal{A}_{t,\delta}$, we have

$$\log \left(\frac{e^d \det \left(\frac{\sqrt{2}}{t^{1/2}} \sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u} \right)}{\delta} \right) \leq c_3 \sigma^2 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u t) + \log(e/\delta) \right)$$

for some universal constant $c_3 > 0$.

Step 3: (Putting everything together) To conclude, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{B}_t \cap \mathcal{E}_{1,t,\delta} \cap \mathcal{E}_{2,t,\delta}$, we have

$$\|B_t - B\| \leq \frac{C\sigma^2 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u t) + \log(e/\delta) \right)}{t^{1/2}} \tag{84}$$

for some universal constant $C > 0$. Therefore, the upper bound (88) holds with probability at least $1-\delta$, provided that $(\star)$ and $(\star\star)$ hold. These two conditions hold whenever

$$t \geq c\sigma^2(d_u\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta))$$

for some universal constant $c > 0$. This concludes the proof. □

Lemma 9. *Assume that $\mathcal{T}$ satisfies (5). Under* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$, we have*

$$\mathbb{P}\left(\forall t \geq t(\delta), \quad \begin{matrix}(i) & \max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t)\end{matrix}\right) \geq 1-\delta,$$

where $t(\delta)^{1/2} = c\sigma^2\|P_\star\|^{10}(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$ for some universal positive constant $c > 0$.

Proof. Let $t \geq 0$ and $\delta \in (0,1)$. Assume that the following condition holds.

$$t^{1/2} \geq C160^2\sigma^2\|P_\star\|^{10}\left(d_u\gamma_\star \log\left(\sigma C_\circ \mathcal{G}_\circ d_x d_u t\right) + d_x + \log(e/\delta)\right). \tag{85}$$

Then, by proposition 3, if the condition

$$t \geq c_1(d_u\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta)), \tag{86}$$

also holds, then we have

$$\mathbb{P}\left(\|B_t - B\|^2 \leq \frac{1}{160^2\|P_\star\|^{10}}\right) \geq 1-\delta. \tag{87}$$

Since $\mathcal{T}$ satisfies condition (5), if the constant $c_1 > 0$ is chosen large enough, we can claim that for all $t \in \mathbb{N}$ such that (85) and (86) hold, we have $K_t = K_{t_k}$ for some t_k that also satisfies (85) and (86).

Using Propostion 16, we may conclude that when (85) and (86) hold, we have

$$\mathbb{P}\left(\max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}}\right) \geq 1-\delta.$$

By Lemma 22, we can find a universal positive constant $c_2 > 0$ such that

$$t^{1/2} \geq c_2\sigma^2\|P_\star\|^{10}(d_u\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))$$

implies that the conditions (85) and (86) hold, and $\|K_t\|^2 \leq h(t)$. Now using Lemma 23 yields

$$\mathbb{P}\left(\forall t \geq t(\delta), \quad \begin{matrix}(i) & \max(\|K_t - K_\star\|, \|B(K_t - K_\star)\|) \leq \frac{1}{5\|P_\star\|^{3/2}} \\ (ii) & \|K_t\|^2 \leq h(t)\end{matrix}\right) \geq 1-\delta,$$

where $t(\delta)^{1/2} = c\sigma^2\|P_\star\|^{10}(d_u\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + d_x + \log(e/\delta))$ for some universal positive constant $c > c_2$. □

Proposition 4. *Assume that $\mathcal{T}$ satisfies (5). Under* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$,*

$$\|B_t - B\| \leq \frac{C\sigma^2\left(d_u\gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u t) + d_x + \log(e/\delta)\right)}{\mu_\star^2 t}$$

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

$$t^{1/2} \geq \frac{c\sigma^2 C_K^2\|P_\star\|^{10}}{\mu_\star}\left(d_u\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta)\right)$$

for some positive constants $C, c > 0$.

Proof. The proof differs from that of Proposition 3 only in the second step, where instead we define the event

$$\mathcal{B}_t = \left\{ \lambda_{\min}\left(\sum_{s=0}^{t-2} u_s u_s^\top \right) \geq C\mu_\star^2 t \right\},$$

where $\mu_\star^2 = \min(\lambda_{\min}(K_\star K_\star^\top), 1)$, and $C > 0$ is some universal constant to be defined through Proposition 13. Indeed, using Lemma 9, we may apply Proposition 13 which guarantees that $\mathcal{B}_t$ holds with probability at least $1-\delta$, provided that

$$t^{1/2} \geq c \frac{\sigma^2 C_K^2 \|P_\star\|^{10}}{\mu_\star} \left(d_u \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta) \right)$$

for some positive constant $c > 0$. The remaining steps are identical to those of the proof of Proposition 3. □

Proposition 5. *Assume that $\mathcal{T}$ satisfies (5), and that under $\mathrm{CEC}(\mathcal{T})$, for all $\delta \in (0,1)$, we have*

$$\mathbb{P}\left(\forall t \geq t(\delta), \begin{array}{ll} (i) & \widetilde{K}_t = K_t \\ (ii) & \max(\|B(K_t - K_\star), \|K_t - K_\star\|) \leq (4\|P_\star\|^{3/2})^{-1} \end{array} \right) \geq 1 - t(\delta)(\delta)$$

for some $t(\delta) \geq \log(e/\delta)$. Then for all $\delta \in (0,1)$, the following

$$\|B_t - B\| \leq \frac{c\sigma^2}{\mu_\star^2 t} \left((d_u + d_x)\gamma_\star \log\left(\frac{e\sigma C_K \|P_\star\| d_x d_u}{\mu_\star^2} \right) + \log(e/\delta) \right)$$

holds with probability at least $1 - c_1 t(\delta)\delta$ provided that

$$t \geq c_2 \max(t(\delta)^{3\gamma_\star}, \sigma^4 (d_u \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta))^2$$

for some universal constants $C, c_1, c_2 > 0$.

Proof of Proposition 5. Fix $t \geq 2$. We start by writing the estimation error as

$$B_t - B = \left(\sum_{s=0}^{t-2} \eta_s u_s^\top \right) \left(\sum_{s=0}^{t-2} u_s u_s^\top \right)^\dagger.$$

When $\sum_{s=0}^{t-2} u_s u_s^\top$ is invertible, we can decompose the estimation error as

$$\|B_t - B\|^2 \leq \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2 \frac{1}{\lambda_{\min}\left(\sum_{s=0}^{t-2} u_s u_s^\top \right)}.$$

Step 1: (Bounding the smallest eigenvalue) We define the event

$$\mathcal{B}_t = \left\{ \lambda_{\min}\left(\sum_{s=0}^{t-2} u_s u_s^\top \right) \geq C_1 \mu_\star^2 t \right\},$$

where $\mu_\star^2 = \min(\lambda_{\min}(K_\star K_\star^\top), 1)$, and $C_1 > 0$ is some universal constant to be defined through Proposition 13. Indeed, using Lemma 9, we may apply Proposition 13 which guarantees that $\mathcal{B}_t$ holds with probability at least $1-\delta$, provided that

$$(\star) \quad t^{1/2} \geq c_1 \frac{\sigma^2 C_K^2 \|P_\star\|^{10}}{\mu_\star} \left(d_u \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta) \right)$$

for some positive constant $c_1 > 0$. Under the event $\mathcal{B}_t$, the estimation error may be upper bounded as

$$\begin{aligned} \|B_t - B\|^2 &\leq \frac{1}{C_1 \mu_\star^2 t} \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2 \\ &\leq \frac{2}{C_1 \mu_\star^2 t} \left\| \left(\sum_{s=0}^{t-2} u_s u_s^\top + C_1 \mu_\star^2 t I_{d_u} \right)^{-1/2} \left(\sum_{s=0}^{t-2} u_s \eta_s^\top \right) \right\|^2, \end{aligned}$$

where we used the fact that $2\sum_{s=0}^{t-2} u_s u_s^\top \succeq \sum_{s=0}^{t-2} u_s u_u^\top + C_1\mu_\star^2 t I_{d_u}$.

Step 2: (Bounding the self-normalized term) Consider the events,

$$\mathcal{E}_{1,t,\delta} = \Bigg\{ \Bigg\| \Bigg(\sum_{s=0}^{t-2} u_s u_s^\top + C_1\mu_\star^2 t I_{d_u} \Bigg)^{-1/2} \Bigg(\sum_{s=0}^{t-2} u_s \eta_s^\top \Bigg) \Bigg\|^2 \leq 7\sigma^2 \log\left(\frac{e^{d_x} \det\left(\frac{1}{C_1\mu_\star^2 t} \sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u} \right)}{\delta} \right) \Bigg\},$$

$$\mathcal{E}_{2,t,\delta} = \left\{ \sum_{s=0}^{t-2} \|\zeta_s\|^2 \leq \sigma^2 (2d_u t + 3\log(1/\delta)) \right\},$$

$$E_{1,\delta,t} = \left\{ \sum_{t(\delta)}^{t} \|x_t\|^2 \leq 8\|P_\star\|^{3/2} (\|x_{t(\delta)}\| + 3\sigma^2(d_x t + \log(e/\delta))) \right\}$$

$$\mathcal{C}_{1,\delta} = \left\{ \forall t \geq t(\delta), \begin{array}{ll} (i) & \widetilde{K}_t = K_t \\ (ii) & \max(\|B(K_t - K_\star), \|K_t - K_\star\|)\| \leq (4\|P_\star\|^{3/2})^{-1} \end{array} \right\}$$

By Proposition 15, the event $\mathcal{A}_{t,\delta}$ holds with probability at least $1-\delta$ for some universal positive constant $c_2 > 0$. By Proposition 9, the event $\mathcal{E}_{1,t,\delta}$ holds with probability at least $1-\delta$, and by Hanson-Wright inequality (See Proposition 19) the event $\mathcal{E}_{2,t,\delta}$ holds with probability at least $1-\delta$. By Proposition 18 we have $\mathbb{P}(E_{1,\delta,t} \cup \mathcal{C}_{1,\delta}^c) \geq 1-\delta$, and by assumption the event $\mathcal{C}_{1,\delta}^c$) holds with probability at least $1 - t(\delta)\delta$. Under the event $\mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$ we have

$$\begin{aligned}
\left(\det\left(\frac{1}{C\mu_\star^2 t} \sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u} \right) \right)^{\frac{1}{d_u}} &\leq \frac{1}{C\mu_\star^2 t} \sum_{s=0}^{t-2} \|u_s\|^2 + 1 \\
&\leq \frac{2}{C\mu_\star^2 t} \sum_{s=0}^{t-2} \|\widetilde{K}_s\|^2 \|x_s\|^2 + \|\zeta_s\|^2 + 1 \\
&\leq \frac{2}{C\mu_\star^2 t} \left(\sum_{s=0}^{t(\delta)} \widetilde{K}_s \|x_s\|^2 + \sum_{s=t(\delta)}^{t-2} 4C_K^2 \|x_s\|^2 + \sum_{s=0}^{t-2} \|\zeta_s\|^2 \right) + 1.
\end{aligned}$$

Now always under $\mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$, provided $t \geq t(\delta)^{1+3\gamma/2}$ we have

$$\sum_{s=0}^{t(\delta)} \widetilde{K}_s \|x_s\|^2 \leq \|K_\circ\|^2 h(t(\delta)) \sum_{s=0}^{t(\delta)} \|x_s\|^2 \leq \sigma^2 d_x \|K_\circ\| t(\delta)^{1+3\gamma/2} \leq \sigma^2 d_x \|K_\circ\| t$$

where we used the fact $\|\widetilde{K}_s\|^2 \leq \|K_\circ\|^2 h(t)$, and the fact at time $t = t(\delta)$ $\widetilde{K}_t = K_t$, so that $\sum_{s=0}^{t(\delta)} \|x_s\|^2 \leq \sigma^2 d_x f(t(\delta))$. Furthermore, we have

$$\begin{aligned}
\sum_{s=t(\delta)}^{t-2} 4C_K^2 \|x_s\|^2 &\leq 32 C_K^2 \|P_\star\|^{3/2} (\|x_{t(\delta)}\| + 3\sigma^2(d_x t + \log(e/\delta))) \\
&\leq 96 C_K^2 \|P_\star\|^{3/2} \sigma^2 d_x (2t(\delta)^{1+\gamma/2} + t) \\
&\leq 200 C_K^2 \|P_\star\|^{3/2} \sigma^2 d_x t.
\end{aligned}$$

Thus, provided that $t \geq t(\delta)^{1+3/2\gamma}$, we obtain that under the event $\mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$ we have

$$\left(\det\left(\frac{1}{C\mu_\star^2 t} \sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u} \right) \right)^{\frac{1}{d}} \leq \frac{c_1 \sigma^2 C_K^2 \|P_\star\|^{3/2} d_x d_u}{\mu_\star^2}$$

for some universal positive constant $c_1 > 0$. Denote

$$(\star\star) \quad t \geq t(\delta)^{1+3/2\gamma}$$

Therefore, assuming that $(\star\star)$ holds, under the event $\mathcal{E}_{1,\delta,t} \cap \mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$, we have

$$\log\left(\frac{e^{d_x}\det\left(\frac{1}{C_1\mu_\star^2 t}\sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u}\right)}{\delta}\right) \leq c_4\left(d\log\left(\frac{e\sigma C_K\|P_\star\|d_x d_u}{\mu_\star^2}\right) + \log(e/\delta)\right)$$

for some universal constant $c_4 > 0$.

Step 3: (Putting everything together) To conclude, under the event $\mathcal{B}_t \cap \mathcal{E}_{1,\delta,t} \cap \mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$, we have

$$\|B_t - B\| \leq \frac{C\sigma^2}{\mu_\star^2 t}\left(d\log\left(\frac{e\sigma C_K\|P_\star\|d_x d_u}{\mu_\star^2}\right) + \log(e/\delta)\right) \tag{88}$$

for some universal constant $C > 0$. Therefore, the upper bound (88) holds with probability at least $1 - Ct(\delta)\delta$, provided that $(\star)$ and $(\star\star)$ hold. These two conditions hold whenever

$$t \geq c\max(t(\delta)^{3\gamma_\star}, \sigma^4(d_u\gamma_\star\log(e\sigma C_\circ\mathcal{G}_\circ d_x\gamma_\star) + \log(e/\delta))^2)$$

for some universal constants $C, c > 0$. This concludes the proof. □

F.4 Error Rate In Scenario II

Proposition 6. *Under* CEC($\mathcal{T}$)*, for all* $\delta \in (0,1)$*,*

$$\|A_t - A\|^2 \leq \frac{C\sigma^2\left(d_x\gamma\log(e\sigma C_\circ\mathcal{G}_\circ d_x t) + \log(e/\delta)\right)}{t}$$

with probability at least $1-\delta$*, when* $t \geq c(d_x\gamma_\star\log(e\sigma C_\circ\mathcal{G}_\circ d_x\gamma_\star) + \log(e/\delta))$*, for some universal positive constants* $C, c > 0$*.*

Proof of Proposition 6. Fix $t \geq 2$. We start by writing the estimation error as

$$A_t - A = \left(\sum_{s=0}^{t-2}\eta_s x_s^\top\right)\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)^\dagger.$$

When $\sum_{s=0}^{t-2} x_s x_s^\top$ is invertible, we can decompose the estimation error as

$$\|A_t - A\|^2 \leq \left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s\eta_s^\top\right)\right\|^2 \frac{1}{\lambda_{\min}\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)}.$$

Step 1: (Bounding the smallest eigenvalue) Define the event

$$\mathcal{B}_t = \left\{\lambda_{\min}\left(\sum_{s=0}^{t} x_s x_s^\top\right) \geq C_1 t\right\}.$$

By Proposition 14, the event $\mathcal{B}_t$ holds with probability at least $1-\delta$, provided that the condition $(\star)$ holds:

$$(\star) \quad t \geq c_1\sigma^2(d_x\gamma_\star\log(e\sigma C_\circ\mathcal{G}_\circ d_x\gamma_\star) + \log(e/\delta)),$$

for some universal constants $C_1, c_1 > 0$. Under the event $\mathcal{B}_t$, we have

$$\begin{aligned}\|A_t - A\|^2 &\leq \frac{1}{C_1 t}\left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s \eta_s^\top\right)\right\|^2 \\ &\leq \frac{2}{C_1 t}\left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top + C_1 t I_{d_x}\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s \eta_s^\top\right)\right\|^2,\end{aligned}$$

where we used the fact that $2\sum_{s=0}^{t-2} x_s x_s^\top \succeq \sum_{s=0}^{t-2} x_s x_s^\top + C_1 t$.

Step 2: (Bounding the self-normalized term) Consider the events,

$$\begin{aligned}\mathcal{E}_{t,\delta} = \Bigg\{ &\left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top + C_1 t I_{d_u}\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s \eta_s^\top\right)\right\|^2 \\ &\leq 7\sigma^2 \log\left(\frac{e^{d_x}\det\left(\frac{1}{C_1 t}\sum_{s=0}^{t-2} x_s x_s^\top + I_{d_x}\right)}{\delta}\right)\Bigg\}, \\ \mathcal{A}_{t,\delta} = &\left\{\sum_{s=0}^{t} \|x_s\|^2 \leq c_2 \sigma^2 C_\circ^2 \mathcal{G}_\circ^2 (d_x t^{1+2\gamma} + \log(1/\delta))\right\}.\end{aligned}$$

By Proposition 15, the event $\mathcal{A}_{t,\delta}$ holds with probability at least $1-\delta$ for some universal positive constant $c_2 > 0$ (involved in the definition of $\mathcal{A}_{t,\delta}$). By Proposition 9, the event $\mathcal{E}_{t,\delta}$ holds with probability at least $1-\delta$. Under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{t,\delta}$, we have

$$\begin{aligned}\left(\det\left(\frac{1}{C_1 t}\sum_{s=0}^{t-2} x_s x_s^\top + I_{d_u}\right)\right)^{\frac{1}{d_x}} &\leq \frac{1}{C_1 t}\sum_{s=0}^{t-2}\|x_s\|^2 + 1 \\ &\leq \frac{2c_1}{C_1}\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{2\gamma},\end{aligned}$$

where we assumed that $(\star\star)$: $t \geq \log(e/\delta)$. Therefore, assuming that $(\star\star)$ holds, we have, under the event $\mathcal{E}_{t,\delta} \cap \mathcal{E}_{2,t,\delta} \cap \mathcal{A}_{t,\delta}$,

$$\log\left(\frac{e_x^d \det\left(\frac{1}{C_1 t}\sum_{s=0}^{t-2} x_s x_s^\top + I_{d_x}\right)}{\delta}\right) \leq c_3\sigma^2\left(d_x \gamma \log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)\right),$$

for some universal constant $c_3 > 0$.

Step 3: (Putting everything together) To conclude, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{B}_t \cap \mathcal{E}_{t,\delta}$, we have

$$\|A_t - A\| \leq \frac{C\sigma^2\left(d_x \gamma \log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)\right)}{t} \tag{89}$$

for some universal constant $C > 0$. Therefore the upper bound (88) holds with probability at least $1-2\delta$, provided that $(\star)$ and $(\star\star)$ hold. These two conditions hold whenever

$$t \geq c\sigma^2(d_x \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)),$$

for some universal constant $c > 0$. This concludes the proof.

□

Proposition 7. *Assume that $\mathcal{T}$ satisfies (5), and that under* CEC($\mathcal{T}$), *for all $\delta \in (0,1)$, we have*

$$\mathbb{P}\left(\forall t \geq t(\delta), \begin{array}{ll}(i) & \widetilde{K}_t = K_t \\ (ii) & \max(\|B(K_t - K_\star), \|K_t - K_\star\|)\| \leq (4\|P_\star\|^{3/2})^{-1}\end{array}\right) \geq 1 - t(\delta)\delta$$

for some $t(\delta) \geq \log(e/\delta)$. Then for all $\delta \in (0,1)$, the following

$$\|A_t - A\| \leq \frac{C\sigma^2}{t}\left(d_x \log\left(e\|P_\star\| d_x\right) + \log(e/\delta)\right)$$

holds with probability at least $1 - c_1 t(\delta)\delta$

$$t \geq c_2 \max(t(\delta)^{2\gamma_\star}, \sigma^2(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)))$$

for some universal constants $C, c_1, c_2 > 0$.

Proof of Proposition 7. Fix $t \geq 2$. We start by writing the estimation error as

$$A_t - A = \left(\sum_{s=0}^{t-2} \eta_s x_s^\top\right)\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)^\dagger.$$

When $\sum_{s=0}^{t-2} x_s x_s^\top$ is invertible, we can decompose the estimation error as

$$\|A_t - A\|^2 \leq \left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s \eta_s^\top\right)\right\|^2 \frac{1}{\lambda_{\min}\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)}.$$

Step 1: (Bounding the smallest eigenvalue) Define the event

$$\mathcal{B}_t = \left\{\lambda_{\min}\left(\sum_{s=0}^{t} x_s x_s^\top\right) \geq C_1 t\right\}.$$

By Proposition 14, the event $\mathcal{B}_t$ holds with probability at least $1-\delta$, provided that the condition $(\star)$ holds:

$$(\star) \quad t \geq c_1 \sigma^2 (d_x \gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)),$$

for some universal constants $C_1, c_1 > 0$. Under the event $\mathcal{B}_t$, we have

$$\begin{aligned}\|A_t - A\|^2 &\leq \frac{1}{C_1 t}\left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s \eta_s^\top\right)\right\|^2 \\ &\leq \frac{2}{C_1 t}\left\|\left(\sum_{s=0}^{t-2} x_s x_s^\top + C_1 t I_{d_x}\right)^{-1/2}\left(\sum_{s=0}^{t-2} x_s \eta_s^\top\right)\right\|^2,\end{aligned}$$

where we used the fact that $2\sum_{s=0}^{t-2} x_s x_s^\top \succeq \sum_{s=0}^{t-2} x_s x_s^\top + C_1 t$.

Step 2: (Bounding the self-normalized term) Consider the events,

$$\begin{aligned}\mathcal{E}_{1,t,\delta} = \Bigg\{&\left\|\left(\sum_{s=0}^{t-1} x_s x_s^\top + C_1 t I_{d_u}\right)^{-1/2}\left(\sum_{s=0}^{t-1} x_s \eta_s^\top\right)\right\|^2, \\ &\leq 7\sigma^2 \log\left(\frac{e^{d_x}\det\left(\frac{1}{C_1 t}\sum_{s=0}^{t-1} x_s x_s^\top + I_{d_u}\right)}{\delta}\right)\Bigg\}, \\ E_{1,\delta,t} = &\left\{\sum_{t(\delta)}^{t} \|x_t\|^2 \leq 8\|P_\star\|^{3/2}(\|x_{t(\delta)}\| + 3\sigma^2(d_x t + \log(e/\delta)))\right\}, \\ \mathcal{C}_{1,\delta} = &\left\{\forall t \geq t(\delta), \begin{array}{ll}(i) & \widetilde{K}_t = K_t \\ (ii) & \max(\|B(K_t - K_\star), \|K_t - K_\star\|)\| \leq (4\|P_\star\|^{3/2})^{-1}\end{array}\right\}.\end{aligned}$$

By Proposition 9, the event $\mathcal{E}_{1,t,\delta}$ holds with probability at least $1-\delta$. By Proposition 18 we have $\mathbb{P}(E_{1,\delta,t} \cup \mathcal{C}_{1,\delta}^c) \geq 1-\delta$, and by assumption the event $\mathcal{C}_{1,\delta}^c$ holds with probability at least $1 - t(\delta)\delta$. Under the event $E_{1,\delta} \cap \mathcal{C}_{1,\delta}$ we have

$$\begin{aligned}\left(\det\left(\frac{1}{C_1\mu_\star^2 t}\sum_{s=0}^{t-2} x_s x_s^\top + I_{d_u}\right)\right)^{\frac{1}{d_x}} &\leq \frac{1}{C_1 t}\sum_{s=0}^{t-1}\|x_s\|^2 + 1 \\ &\leq \frac{2}{C_1 t}\left(\sum_{s=0}^{t(\delta)}\|x_s\|^2 + \sum_{s=t(\delta)}^{t-2}\|x_s\|^2\right) + 1.\end{aligned}$$

Now always under $\mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$, provided $t \geq t(\delta)^{1+\gamma/2}$ we have

$$\sum_{s=0}^{t(\delta)} \widetilde{K}_s \|x_s\|^2 \leq \sigma^2 d_x t(\delta)^{1+\gamma/2} \leq \sigma^2 d_x t$$

where we used the fact at time $t = t(\delta)$, $\widetilde{K}_t = K_t$, so that $\sum_{s=0}^{t(\delta)}\|x_s\|^2 \leq \sigma^2 d_x f(t(\delta))$. Furthermore, we have

$$\begin{aligned}\sum_{s=t(\delta)}^{t-1}\|x_s\|^2 &\leq 8\|P_\star\|^{3/2}(\|x_{t(\delta)}\| + 3\sigma^2(d_x t + \log(e/\delta))) \\ &\leq 21\|P_\star\|^{3/2}\sigma^2 d_x(t(\delta)^{1+\gamma/2} + t) \\ &\leq 42\|P_\star\|^{3/2}\sigma^2 d_x t.\end{aligned}$$

Thus, provided that $t \geq t(\delta)^{1+1/2\gamma}$, we obtain that under the event $E_{1,\delta} \cap \mathcal{C}_{1,\delta}$ we have

$$\left(\det\left(\frac{1}{C_1 t}\sum_{s=0}^{t-1} x_s x_s^\top + I_{d_u}\right)\right)^{\frac{1}{d_x}} \leq c_1\sigma^2\|P_\star\|^{3/2} d_x$$

for some universal positive constant $c_1 > 0$. Denote

$$(\star\star) \quad t \geq t(\delta)^{1+1/2\gamma}$$

Therefore, assuming that $(\star\star)$ holds, under the event $\mathcal{E}_{1,\delta,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$, we have

$$\log\left(\frac{e^{d_x}\det\left(\frac{1}{C_1\mu_\star^2 t}\sum_{s=0}^{t-2} u_s u_s^\top + I_{d_u}\right)}{\delta}\right) \leq c_4\left(d_x \log\left(e\sigma\|P_\star\| d_x\right) + \log(e/\delta)\right)$$

for some universal constant $c_4 > 0$.

Step 3: (Putting everything together) To conclude, under the event $\mathcal{B}_t \cap \mathcal{E}_{1,\delta,t} \cap \mathcal{E}_{2,t} \cap E_{1,\delta} \cap \mathcal{C}_{1,\delta}$, we have

$$\|A_t - A\| \leq \frac{C\sigma^2}{t}\left(d_x \log\left(e\|P_\star\| d_x\right) + \log(e/\delta)\right) \tag{90}$$

for some universal constant $C > 0$. Therefore, the upper bound (90) holds with probability at least $1 - Ct(\delta)\delta$, provided that $(\star)$ and $(\star\star)$ hold. These two conditions hold whenever

$$t \geq c\max(t(\delta)^{2\gamma_\star}, \sigma^2(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)))$$

for some universal constants $C, c > 0$. This concludes the proof. □

G Smallest Eigenvalue of the Cumulative Covariates Matrix

This appendix is devoted to the analysis of the smallest eigenvalue of the cumulative covariates matrix. This eigenvalue should exhibit an appropriate scaling so that the LSE performs well. We first provide a generic recipe for the analysis of this eigenvalue, and then apply it to the three scenarios. For Scenario I, the results are stated in Proposition 10 and 11. Our analysis of Scenario III is summarized in Propositions 12 and 13. Finally, for Scenario II, we establish Proposition 14.

G.1 A generic recipe

In the three scenarios, we will have to obtain high probability bounds on the smallest eigenvalue of a matrix of the form $\sum_{s=1}^t y_s y_s^\top$ where $y_s = z_s + M_s\xi_s$ where ξ_s is a random variable independent of $z_1, \dots, z_s$ and $M_1, \dots, M_s$ for all $s \geq 1$. The need for such guarantee stems mainly from the analysis of the least squares estimator. Because of this structure, common to the three settings, our proofs for the different scenarios will be similar in spirit up to some technical details that are mainly related to the nature of the sequence of matrices $(M_s)_{s\geq 1}$. We shall now sketch a generic recipe for our proofs.

Sketch of the recipe. The first step is to use Lemma 10, which will allow us to lower bound[5] $\sum_{s=1}^t y_s y_s^\top$. We obtain, for all $\lambda > 0$,

$$\sum_{s=1}^t y_s y_s^\top \succeq \underbrace{\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top}_{\textit{Random Matrix}} - \underbrace{\left(\sum_{s=1}^t z_s(M_s\xi_s)^\top\right)^\top \left(\sum_{s=1}^t z_s z_s^\top + \lambda I_d\right)^{-1} \left(\sum_{s=1}^t z_s(M_s\xi_s)^\top\right)}_{\textit{Self-Normalized Matrix Valued Process}} - \lambda I_d.$$

Then, we bound the random matrix (first term) using conditional independence via Proposition 8. Finally, we also bound the Self-Normalized Matrix Process (the second term) using Proposition 9.

Ingredients of the recipe. Let us now list the main lemmas and propositions used above. Their proofs are presented in G.5.

Lemma 10. *Let $(y_t)_{t\geq 1}$, $(z_t)_{t\geq 1}$, and $(\xi_t)_{t\geq 1}$ be three sequences of vectors in $\mathbb{R}^d$ satisfying, for all $s \geq 0$, the linear relation $y_s = z_s + \xi_s$. Then, for all $\lambda > 0$, all $t \geq 1$ and all $\varepsilon \in (0, 1]$, we have*

$$\sum_{s=1}^t y_s y_s^\top \succeq \sum_{s=1}^t \xi_s\xi_s^\top + (1-\varepsilon)\sum_{s=1}^t z_s z_s^\top - \frac{1}{\varepsilon}\left(\sum_{s=1}^t z_s\xi_s\right)^\top \left(\sum_{s=1}^t z_s z_s^\top + \lambda I_d\right)^{-1} \left(\sum_{s=1}^t z_s\xi_s^\top\right) - \varepsilon\lambda I_d.$$

Proposition 8. *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration over the underlying probability space. Let $(\xi_t)_{t\geq 1}$ be a sequence of independent, zero-mean, σ^2-sub-gaussian, isotropic random vectors taking values in $\mathbb{R}^p$ and such that ξ_t is $\mathcal{F}_t$-measurable for all $t \geq 1$. Let $(M_t)_{t\geq 1}$ be a sequence of random matrices taking values in $\mathbb{R}^{d\times p}$, such that M_t is $\mathcal{F}_{t-1}$-measurable and its norm $\|M_s\|$ is bounded a.s.. Let $m = (m_t)_{t\geq 1}$ refer to the sequence of such bounds (e.g. $\|M_s\| \leq m_s$). Then*

$$\mathbb{P}\left(\left\|\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - \sum_{s=1}^t M_s M_s^\top\right\| > 8\sigma^2\|m_{1:t}\|_2^2 \max\left(\sqrt{\frac{2\rho+5d}{r_t^2}}, \frac{2\rho+5d}{r_t^2}\right)\right) \leq 2e^{-\rho}$$

where $\|m_{1:t}\|_\infty = \max_{1\leq s\leq t}|m_s|$, $\|m_{1:t}\|_2 = \sqrt{\sum_{s=1}^t |m_s|^2}$, and $r_t = \|m_{t:t}\|_2/\|m_{1:t}\|_\infty$.

Remark 1. *For our purposes, $\|m_{1:t}\|_2^2$ and r_t^2 will be either of order $\mathcal{O}(1)$ and $\mathcal{O}(t)$ respectively, or of order $\mathcal{O}(\log(t))$ and $\mathcal{O}(\sqrt{t})$ respectively. These scalings will depend on the scenario considered.*

An immediate consequence of Proposition 8 is:

[5] Here we mean lower bound in the Lwner partial order over symmetric matrices.

Corollary 1. *Under the same assumptions on $(\xi_t)_{t\geq 1}$ and $(M_s)_{t\geq 1}$ as in Proposition 8, if we further assume that $\sup_{s\geq 1}|m_s| \leq m$, then we have for all $\rho > 0$, $\varepsilon \in (0,1)$, and for all $t \geq \min\left(\frac{8^2(\sigma m)^4}{\varepsilon^2}, \frac{8(\sigma m)^2}{\varepsilon}\right)(5d+2\rho)$,*

$$\mathbb{P}\left(\sum_{s=1}^{t} M_s M_s^\top - \varepsilon t I_d \preceq \sum_{s=1}^{t-1}(M_s\xi_s)(M_s\xi_s)^\top \preceq \sum_{s=1}^{t} M_s M_s^\top + \varepsilon t I_d\right) \geq 1 - 2e^{-\rho}.$$

In particular if $M_s = I_d$, $(\xi_t)_{t\geq 1}$ are now taking values in $\mathbb{R}^d$, we have for all $\rho > 0$, $\varepsilon \in (0,1)$, and for all $t \geq \min\left(\frac{8^2\sigma^4}{\varepsilon^2}, \frac{8\sigma^2}{\varepsilon}\right)(5d+2\rho)$,

$$\mathbb{P}\left((1-\varepsilon)tI_d \preceq \sum_{s=1}^{t-1}\xi_s\xi_s^\top \preceq (1+\varepsilon)tI_d\right) \geq 1 - 2e^{-\rho}.$$

Proposition 9 (Self-normalized matrix processes)**.** *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration over the underlying probability space. Let $(\xi_t)_{t\geq 1}$ be a sequence of independent, zero-mean, σ^2-sub-gaussian, isotropic random vectors taking values in $\mathbb{R}^p$ and such that ξ_t is $\mathcal{F}_t$-measurable for all $t \geq 1$. Let $(M_t)_{t\geq 1}$ be a sequence of random matrices taking values in $\mathbb{R}^{d\times p}$, such that M_t is $\mathcal{F}_{t-1}$-measurable and its norm $\|M_s\|$ is bounded a.s.. Let $m = (m_t)_{t\geq 1}$ refer to the sequence of such bounds. Let $(z_t)_{t\geq 1}$ be a sequence of random vectors taking values in $\mathbb{R}^d$, such that z_t is $\mathcal{F}_{t-1}$-measurable for all $t \geq 1$. Then for all positive definite matrix $V \succ 0$, the following self-normalized matrix process defined by*

$$\forall t \geq 1, \quad S_t(z, M\xi) \triangleq \left(\sum_{s=1}^{t} z_s(M_s\xi_s)^\top\right)^\top \left(\sum_{s=1}^{t} z_s z_s^\top + V\right)^{-1} \left(\sum_{s=1}^{t} z_s(M_s\xi_s)^\top\right)$$

satisfies, for all $\rho \geq 1$ and $t \geq 1$,

$$\mathbb{P}\left[\|S_t(z, M\xi)\| > \sigma^2\|m_{1:t}\|_\infty^2\left(2\log\det\left(V^{-1}\sum_{s=1}^{t} z_s z_s^\top + I_d\right) + 7d + 4\rho\right)\right] \leq e^{-\rho}.$$

G.2 Application to Scenario I

We now apply the recipe described in the previous subsection to lower bound the smallest eigenvalue of the cumulative covariates matrix in Scenario I. We first prove the following result, that will then be refined.

Proposition 10 (Sufficent exploration)**.** *Under Algorithm* CEC($\mathcal{T}$)*, for all $\delta \in (0,1)$,*

$$\lambda_{\min}\left(\sum_{s=0}^{t-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \geq \left(\frac{t}{2}\right)^{1/4}$$

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

$$t \geq c\sigma^4 C_\circ^8((d_x+d_u)\gamma_\star\log(e\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta))$$

for some universal positive constant $c > 0$.

Proof of Proposition 10. Define for all $t \geq 1$, the event

$$\mathcal{E}_{1,t} = \left\{\exists i \in \{t/2, \dots, t-1\} : \quad \lambda_{\min}\left(\sum_{s=0}^{i-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \geq i^{1/4}\right\}.$$

Let us recall that under CEC($\mathcal{T}$), we have

$$u_t \leftarrow \begin{cases} K_t x_t + \nu_t & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \text{ and } \lambda_{\min}\left(\sum_{s=0}^{t-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \geq t^{1/4} \\ K_\circ x_t + \nu_t & \text{otherwise.}\end{cases}$$

Define for all $s \geq 1$,

$$y_s = \begin{bmatrix} x_s \\ u_s \end{bmatrix}, \quad z_s = \begin{bmatrix} Ax_{s-1} + Bu_{s-1} \\ K_\circ(Ax_{s-1} + Bu_{s-1}) \end{bmatrix}, \quad M_\circ = \begin{bmatrix} I_{d_x} & O \\ K_\circ & I_{d_u} \end{bmatrix}, \quad \text{and} \quad \xi_s = \begin{bmatrix} \eta_{s-1} \\ \nu_s \end{bmatrix}.$$

Note that under the event $\mathcal{E}_{1,t}^c$ we have $y_s = z_s + M_\circ \xi_s$ for all $s \in \{t/2, \dots, t-1\}$. Applying Lemma 10, we obtain

$$\sum_{s=t/2}^{t} y_s y_s^\top \succeq \sum_{s=t/2}^{t} (M_\circ \xi_s)(M_\circ \xi_s)^\top - I_{d_u} - \left\| \left(\sum_{s=t/2}^{t} z_s z_s^\top + I_{d_u s} \right)^{-1/2} \left(\sum_{s=t/2}^{t} z_s (M_\circ \xi_s)^\top \right) \right\|^2 I_{d_u}.$$

For all $\delta \in (0,1)$ and $t \geq 1$, we define the following events

$$\begin{aligned}
\mathcal{A}_{\delta,t} &= \left\{ \sum_{s=0}^{t} \|x_s\| \leq C_1 \sigma^2 C_\circ^2 \mathcal{G}_\circ^2 (d_x t^{1+2\gamma} + \log(e/\delta)) \right\}, \\
\mathcal{E}_{2,\delta,t} &= \left\{ \left\| \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_d \right)^{-1/2} \left(\sum_{s=t/2}^{t-1} z_s (M_\circ \xi_s)^\top \right) \right\|^2 \right. \\
&\qquad \left. \leq 7\sqrt{d_x}\sigma^2 \log \left(\frac{e^d \det \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_{d_x} \right)}{\delta} \right) \right\}, \\
\mathcal{E}_{3,t} &= \left\{ \sum_{s=t/2}^{t-1} \xi_s \xi_s^\top \succeq \begin{bmatrix} (t/2) I_{d_x} & O \\ O & \sigma^2 \sqrt{d_x t}/2 I_{d_u} \end{bmatrix} \right\}, \\
\mathcal{E}_{4,t} &= \left\{ \lambda_{\max} \left(\sum_{s=0}^{t-1} \eta_s \eta_s^\top \right) \leq \frac{3t}{2} \right\}.
\end{aligned}$$

In view of Proposition 15, $\mathbb{P}(\mathcal{A}_{t,\delta}) \geq 1 - \delta$. From Proposition 9, we have $\mathbb{P}(\mathcal{E}_{2,\delta,t}) \geq 1 - \delta$. From Proposition 8, $\mathbb{P}(\mathcal{E}_{3,t}) \geq 1 - \delta$ provided that $t \geq c_1 \sigma^2 (d + \log(e/\delta))$, where we first normalize to obtain $\sum_{s=t/2}^{t} (\mathbb{E}[\xi_t \xi_t^\top]^{-1/2} \xi_s)(\mathbb{E}[\xi_t \xi_t^\top]^{-1/2} \xi_s)^\top$ then apply the proposition to get the high probability bound. We have by Proposition 8, that $\mathbb{P}(\mathcal{E}_{4,t}) \geq 1 - \delta$ provided that $t \geq \sigma^2 (d_x + \log(e/\delta))$.

Provided that $t \geq \log(e/\delta)$, we have under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{4,t}$ that

$$\begin{aligned}
\det \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_{d_x} \right)^{1/d} &\leq \sum_{s=t/2}^{t} \|z_s\|^2 + 1 \\
&\leq \sum_{s=t/2}^{t} 2C_\circ^2 \|x_s - \eta_{s-1}\|^2 + 1 \\
&\leq 4C_\circ^2 \sum_{s=0}^{t} \|x_s\|^2 + \|\eta_{s-1}\|^2 + 1 \\
&\leq 21 C_1 \sigma^2 C_\circ^4 \mathcal{G}_\circ^2 d_x t^{3\gamma^\star}.
\end{aligned}$$

Therefore, provided that $t \geq \log(e/\delta)$, we have under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{4,t}$ that

$$\left\| \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_d \right)^{-1/2} \left(\sum_{s=t/2}^{t-1} z_s (M_\circ \xi_s)^\top \right) \right\|^2 \leq C_2 \sqrt{d_x} \sigma^2 (d \gamma_\star \log(e \sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta))$$

for some universal positive constant $C_2 > 0$. Furthermore, if $t \geq \sigma^4 d_x$, under the event $\mathcal{E}_{3,t}$, we have

$$\begin{aligned}
\sum_{t/2}^{t} (M_\circ \xi_s)(M_\circ \xi_s)^\top &\succeq M_\circ \begin{bmatrix} (t/2) I_{d_x} & O \\ O & \sigma^2 \sqrt{d_x t}/2 I_{d_u} \end{bmatrix} M_\circ^\top \\
&\succeq \frac{1}{2} \begin{bmatrix} t I_{d_x} & t K_\circ^\top \\ t K_\circ & t K_\circ^\top K_\circ + \sigma^2 \sqrt{d_x t} I_{d_u} \end{bmatrix} \\
&\succeq \frac{t}{2} \min \left(\frac{\sigma^2 \sqrt{d_x}}{2\|K_\circ\|^2 \sqrt{t} + \sigma^2 \sqrt{d_x}}, \frac{\sigma^2 \sqrt{d_x}}{2\sqrt{t}} \right) I_d \\
&\succeq \frac{\sigma^2 \sqrt{d_x}}{2} \min \left(\frac{t}{2\|K_\circ\|^2 \sqrt{t} + \sigma^2 \sqrt{d_x}}, \frac{\sqrt{t}}{2} \right) I_d \\
&\succeq \frac{\sigma^2 \sqrt{d_x t}}{6 C_\circ^2} I_d,
\end{aligned}$$

where we used Lemma 11 (with $\alpha = 1/2$, and $\beta = 1$).

Therefore, provided that $t \geq \sigma^2 d_x$ and $t \geq \log(e/\delta)$, we have under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{3,t} \cap \mathcal{E}_{4,t}$ that

$$\lambda_{\min} \left(\sum_{s=0}^{t-1} u_s u_s^\top \right) \geq \frac{\sigma^2 \sqrt{d_x t}}{6 C_\circ^2} - 1 - C_2 \sqrt{d_x} \sigma^2 (d \gamma_\star \log(e \sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)).$$

Using Lemma 22, there exists $c_3 > 0$ such that if

$$t \geq c_3 \sigma^4 C_\circ^8 (d \gamma_\star \log(e \sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta)) \tag{91}$$

then

- $\frac{\sigma^2 \sqrt{d_x t}}{6 C_\circ^2} - 1 - C_2 \sqrt{d_x} \sigma^2 (d \gamma_\star \log(e \sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)) \geq \frac{\sigma^2 \sqrt{d_x t}}{10 C_\circ^2} > t^{1/4}$,
- $t \geq \log(e/\delta)$,
- $t \geq \sigma^4 d_x$.

Therefore, if condition (91) holds, we have under $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{3,t} \cap \mathcal{E}_{4,t}$

$$\lambda_{\min} \left(\sum_{s=0}^{t-1} u_s u_s^\top \right) > t^{1/4}.$$

But this cannot hold under the event $\mathcal{E}_{1,t}^c$, therefore it must be that $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{3,t} \cap \mathcal{E}_{4,t} \subseteq \mathcal{E}_{1,t}$ which in turns impies that

$$\mathbb{P}(\mathcal{E}_{1,t}) \geq 1 - \mathbb{P}(\mathcal{A}_{\delta,t}^c \cup \mathcal{E}_{2,\delta,t}^c \cup \mathcal{E}_{3,t}^c \cup \mathcal{E}_{4,t}^c) \geq 1 - 4\delta.$$

reparametrizing by $\delta' = 4\delta$ gives the desired result with modified universal postive constants. □

Proposition 11 (Sufficient exploration with refined rates). *Under* CEC($\mathcal{T}$), *assume that for all* $\delta \in (0,1)$, *we have*

$$\mathbb{P}\left(\forall t \geq t(\delta), \quad \|K_t - K_\star\| \leq 1\right) \geq 1-\delta$$

for some $t(\delta) \geq 1$. *Then for all* $\delta \in (0,1)$,

$$\lambda_{\min}\left(\sum_{s=0}^{t-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \geq \frac{C\sigma^2\sqrt{d_x t}}{C_K^2}$$

with probability at least $1-\delta$, *provided that*

$$t^{1/2} \geq c\max(t(\delta)^{1/2}, C_K^4(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))) \tag{92}$$

for some universal positive constants $C, c > 0$.

Proof of Proposition 11. Let us start by defining

$$\mathcal{E}_{1,\delta} = \{\forall t \geq t(\delta),\ \|K_t - K_\star\| \leq 1\}.$$

Now, we recall that under CEC($\mathcal{T}$), we have $u_t = (1-\alpha_t)(K_t x_t + \nu_t) + \alpha_t(K_\circ x_t + \zeta_t)$ for all $t \geq 1$, where we defined

$$\forall t \geq 1, \quad \alpha_t = \begin{cases} 0 & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \text{ and } \lambda_{\min}\left(\sum_{s=0}^{t-1}\begin{bmatrix}x_s\\u_s\end{bmatrix}\begin{bmatrix}x_s\\u_s\end{bmatrix}^\top\right) \geq t^{1/4} \\ 1 & \text{otherwise.}\end{cases} \tag{93}$$

Let $\widetilde{K}_s = (1-\alpha_s)K_s 1_{\{\|K_s - K_\star\| \leq 1\}} + \alpha_s K_\circ$, and note that $\|\widetilde{K}_s\| \leq 2C_K$. Thus, under the event $\mathcal{E}_{1,\delta}$, we have

$$\begin{bmatrix}x_s\\u_s\end{bmatrix} = \begin{bmatrix}Ax_{s-1} + Bu_s\\ \widetilde{K}_s A x_{s-1} + \widetilde{K}_s B u_{s-1}\end{bmatrix} + \begin{bmatrix}I_{d_x} & O\\ \widetilde{K}_s & I_{d_u}\end{bmatrix}\begin{bmatrix}\eta_{s-1}\\ \nu_s\end{bmatrix}. \tag{94}$$

Denote for all $s \geq 1$,

$$y_s = \begin{bmatrix}x_s\\u_s\end{bmatrix}, \quad z_s = \begin{bmatrix}Ax_{s-1} + Bu_s\\ \widetilde{K}_s A x_{s-1} + K_s B u_s\end{bmatrix}, \quad M_s = \begin{bmatrix}I_{d_x} & O\\ \widetilde{K}_s & I_{d_u}\end{bmatrix}, \quad \xi_s = \begin{bmatrix}\eta_{s-1}\\ \nu_s\end{bmatrix},$$
$$\xi_{1,s} = \begin{bmatrix}\eta_{s-1}\\ 0\end{bmatrix}, \quad \text{and} \quad \xi_{2,s} = \begin{bmatrix}0\\ \nu_s\end{bmatrix}.$$

Now, under the event $\mathcal{E}_{1,\delta}$, we may simply write $y_s = z_s + M_s\xi_s = z_s + M_s\xi_{1,s} + \xi_{2,s}$ for all $s \geq 1$. Let us note that ξ_s is independent of $(M_0, \ldots, M_s)$ and $(z_0, \ldots, z_s)$, and that $\|M_s\| \leq \sqrt{5}C_K$. Lemma 10 ensures that

$$\sum_{s=0}^{t-1} y_s y_s^\top \succeq \sum_{s=t(\delta)}^{t-1}(M_s\xi_s)(M_s\xi_s)^\top - \left(\sum_{s=t(\delta)}^{t-1} z_s(M_s\xi_s)^\top\right)^\top\left(\sum_{s=t(\delta)}^{t} z_s z_s^\top + \lambda I_d\right)^{-1}\left(\sum_{s=t(\delta)}^{t-1} z_s(M_s\xi_s)^\top\right) - I_d.$$

Upper bounding the self-normalized term. Define the following events

$$\mathcal{A}_{\delta,t} = \left\{\sum_{s=0}^{t}\|x_s\|^2 \leq C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^2(d_x t^{1+2\gamma} + \log(e/\delta))\right\},$$

$$\mathcal{E}_{2,\delta,t} = \left\{\left\|\left(\sum_{s=t(\delta)}^{t-1} z_s z_s^\top + I_d\right)^{-1/2}\left(\sum_{s=t(\delta)}^{t-1} z_s(M_s\xi_s)^\top\right)\right\|^2 \right.$$
$$\left. \leq 35 C_K^2\sqrt{d_x}\sigma^2 \log\left(\frac{e^d \det\left(\sum_{s=t(\delta)}^{t-1} z_s z_s^\top + I_d\right)}{\delta}\right)\right\},$$

$$\mathcal{E}_{3,t} = \left\{\lambda_{\max}\left(\sum_{s=0}^{t-1}\eta_s\eta_s^\top\right) \leq \frac{3t}{2}\right\}.$$

We have by Proposition 15 that the event $\mathcal{A}_{\delta,t}$ holds with probability at least $1-\delta$ for some universal positive constant $C_1 > 1$. By Proposition 9, the event $\mathcal{E}_{2,\delta,t}$ holds with probability at least $1-\delta$. Finally the event $\mathcal{E}_{3,t}$ holds with probability at least $1-\delta$ provided that $t \geq c_1\sigma^2(d_x + \log(e/\delta))$.

Provided that $t \geq \log(e/\delta)$, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{\delta,t} \cap \mathcal{E}_{\delta,t}$ we have

$$\begin{aligned}\det\left(\sum_{s=t(\delta)}^{t-1} z_s z_s^\top + I_d\right)^{1/d} &\leq \sum_{s=t(\delta)}^{t-1} \|z_s\|^2 + 1 \\ &\leq \sum_{s=0}^{t-1} 4C_K^2\|x_{s+1} - \eta_s\|^2 + 1 \\ &\leq 8C_K^2 \sum_{s=0}^{t-1} \|x_{s+1}\|^2 + \|\eta_s\|^2 + 1 \\ &\leq 8(3C_1+4)\sigma^2 C_K^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{3\gamma_\star}.\end{aligned}$$

Thus provided that $t \geq \log(e/\delta)$, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{3,t}$ we have that

$$\begin{aligned}&\left\|\left(\sum_{s=t(\delta)}^{t-1} z_s z_s^\top + I_d\right)^{-1/2}\left(\sum_{s=t(\delta)}^{t-1} z_s (M_s\xi_s)^\top\right)\right\|^2 \\ &\qquad \leq 35 C_K^2 \sqrt{d_x}\sigma^2(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x t) + \log(e/\delta)).\end{aligned}$$

Lower bounding $\sum_{s=t(\delta)}^{t-1}(M_s\xi_s)(M_s\xi_s)^\top$. This is the most challenging task, and we shall break it into several steps. First, we note that $\xi_{1,s}$ and $\xi_{2,s}$ are independent (by design of CEC($\mathcal{T}$)), and $M_s\xi_s = M_s\xi_{1,s} + \xi_{2,s}$. We use Lemma 10 to write:

$$\sum_{s=t(\delta)}^{t} (M_s\xi_s)(M_s\xi_s)^\top \succeq \xi_2^\top\xi_2 + \frac{1}{2}\xi_1^\top\xi_1 - 2\xi_1^\top\xi_2(\xi_2^\top\xi_2 + I_d)^{-1}\xi_2^\top\xi_1 - \frac{1}{2}I_d$$

where we define, for ease of notations, the tall matrices $\xi_1^\top = \begin{bmatrix} M_1\xi_{1,1} & \dots & M_t\xi_{1,t}\end{bmatrix}$ and $\xi_2^\top = \begin{bmatrix}\xi_{2,t} & \dots & \xi_{2,t}\end{bmatrix}$, so that we have

$$\begin{aligned}\xi_1^\top\xi_1 &= \sum_{s=t(\delta)}^{t-1}(M_s\xi_{1,s})(M_s\xi_{1,s})^\top, \\ \xi_2^\top\xi_2 &= \sum_{s=t(\delta)}^{t-1}\xi_{2,s}\xi_{2,s}^\top, \\ \xi_1^\top\xi_2(\xi_2^\top\xi_2 + I_d)^{-1}\xi_2^\top\xi_1 &= \\ &\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_{1,s})\xi_{2,s}^\top\right)^\top\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_{1,s})(M_s\xi_{1,s})^\top + I_d\right)^{-1}\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_{1,s})\xi_{2,s}^\top\right).\end{aligned}$$

Step 1: We first derive a lower bound on the smallest eigenvalue of $\xi_1\xi_1^\top$. We have

$$\xi_1\xi_1^\top = \sum_{s=t(\delta)}^{t-1}\xi_{1,s}\xi_{1,s}^\top \succeq \begin{bmatrix} \frac{\lambda}{\|\sum_{s=0}^{t-1}(\widetilde{K}_{s+1}\eta_s)(\widetilde{K}_{s+1}\eta_s)^\top\| + \lambda}\sum_{s=0}^{t-1}\eta_s\eta_s^\top & O \\ O & -\lambda I_{d_u}\end{bmatrix}.$$

Indeed, we have for all $\lambda > 0$,

$$\begin{aligned}\xi_1\xi_1^\top &= \sum_{s=1}^{t} \xi_{1,s}\xi_{1,s}^\top \\ &= \sum_{s=1}^{t-1} \begin{bmatrix} \eta_{s-1}\eta_{s-1}^\top & \eta_{s-1}(\widetilde{K}_s\eta_{s-1})^\top \\ \widetilde{K}_s\eta_{s-1}\eta_{s-1}^\top & (\widetilde{K}_s\eta_{s-1})(\widetilde{K}_s\eta_{s-1})^\top \end{bmatrix} \\ &= \begin{bmatrix} \eta^\top\eta & \eta^\top X \\ X\eta & X^\top X \end{bmatrix} \\ &\succeq \begin{bmatrix} \frac{\lambda}{\|X\|^2+\lambda}\eta^\top\eta & 0 \\ O & -\lambda I_{d_u} \end{bmatrix},\end{aligned}$$

where we defined, for ease of notations, the tall matrices $\eta^\top = \begin{bmatrix}\eta_0 & \dots & \eta_{t-1}\end{bmatrix}$, $X^\top = \begin{bmatrix}K_1\eta_0 & \dots & K_t\eta_{t-1}\end{bmatrix}$, and used Lemma 12 to obtain the last inequality. We may apply Corollary 1 to bound from below the above inequality. First define the events

$$\begin{aligned}\mathcal{E}_{4,t} &= \left\{ 2(t-t(\delta))I_{d_x} \succeq \sum_{s=t(\delta)}^{t-1} \eta_s\eta_s^\top \succeq \frac{t-t(\delta)}{2} I_{d_x} \right\}, \\ \mathcal{E}_{5,t} &= \left\{ \left\| \sum_{s=t(\delta)}^{t-1} (\widetilde{K}_{s+1}\eta_s)(\widetilde{K}_{s+1}\eta_s)^\top \right\| \le 8C_K^2(t-t(\delta)) \right\}.\end{aligned}$$

By proposition 8, the event $\mathcal{E}_{4,t}$ holds with probability at least $1-\delta$, provided that $t \ge c_2\sigma^2(d_x + \log(e/\delta))$ for some universal positive constant $c_2 > 0$. By Propostion 8, the event $\mathcal{E}_{5,t}$ holds with probability $1-\delta$, provided that $t \ge c_3\sigma^2(d_x + \log(e/\delta))$ for some universal positive constants $C_3, c_3 > 0$. Therefore, provided that $t \ge 2t(\delta)$ and $\lambda = \epsilon\sqrt{t}$ under the event $\mathcal{E}_{4,t} \cap \mathcal{E}_{5,t}$ we have

$$\begin{aligned}\xi_1^\top\xi_1 &\succeq \begin{bmatrix} \frac{\epsilon(t-t(\delta))\sqrt{t}}{8C_K^2(t-t(\delta))+\epsilon\sqrt{t}} I_{d_x} & O \\ O & -\lambda I_{d_u} \end{bmatrix} \\ &\succeq \begin{bmatrix} \frac{\epsilon t\sqrt{t}}{8C_K^2 t + 2\epsilon\sqrt{t}} I_{d_x} & O \\ O & -\epsilon\sqrt{t} I_{d_u} \end{bmatrix}.\end{aligned}$$

Step 2: Next, we find a lower bound on the smallest eigenvalue of the random matrix $\xi_2\xi_2^\top$. Consider the event

$$\mathcal{E}_{6,t} = \left\{ \lambda_{\min}\left(\sum_{s=t(\delta)}^{t-1} \nu_s\nu_s^\top \right) \ge \frac{\sigma^2\sqrt{d_x(t-t(\delta))}}{2} \right\}.$$

By Proposition 8, the event $\mathcal{E}_{6,t}$ holds with probability at least $1-\delta$, provided that we have $t \ge c_4(d_u + \log(e/\delta)) + t(\delta)$ for some universal positive constant $c_4 > 0$. Note that we need to apply Proposition 8 to the normalized random matrix $\frac{1}{\sigma^2\sqrt{d_x t}}\sum_{s=t(\delta)}^{t}\nu_s\nu_s^\top$. Thus, provided that $t \ge 2t(\delta)$, under the event $\mathcal{E}_{4,t}$, we have

$$\xi_1\xi_1^\top \succeq \begin{bmatrix} O & O \\ O & \frac{\sigma^2\sqrt{d_x t}}{2\sqrt{2}} I_{d_u} \end{bmatrix}.$$

Step 3: We now upper bound the norm of the self-normalized matrix process $\xi_1^\top\xi_2(\xi_2^\top\xi_2 + I_d)^{-1}\xi_2^\top\xi_1$. Consider the event

$$\mathcal{E}_{7,\delta,t} = \left\{ \left\| (\xi_2^\top\xi_2 + I_d)^{-1/2}\xi_2^\top\xi_1 \right\|^2 \le 7\sigma^2\sqrt{d_x}\log\left(\frac{e^d \det\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_{1,s})(M_s\xi_{1,s})^\top + I_d \right)}{\delta} \right) \right\}.$$

By Proposition 9, the event $\mathcal{E}_{7,t,\delta}$ holds with probability at least $1-\delta$. Therefore, provided that $t \geq \log(e/\delta)$, under the event $\mathcal{E}_{7,t,\delta} \cap \mathcal{E}_{4,t}$ we have

$$\begin{aligned}\det\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_{1,s})(M_s\xi_{1,s})^\top + I_d\right)^{1/d} &\leq \sum_{s=t(\delta)}^{t-1}\|M_s\xi_{1,s}\|^2 + 1 \\ &\leq \sum_{s=t(\delta)}^{t-1} 2\|\widetilde{K_s}\|^2\|\eta_s\|^2 + 1 \\ &\leq 8C_K^2 t.\end{aligned}$$

Therefore, for $t \geq \log(e/\delta)$, we have, under $\mathcal{E}_{7,t,\delta} \cap \mathcal{E}_{4,t}$, that

$$\left\|(\xi_2^\top\xi_2 + I_d)^{-1/2}\xi_2^\top\xi_1\right\|^2 \leq C_4\sigma^2\sqrt{d_x}(d\log(eC_Kt) + \log(e/\delta)),$$

for some universal positive constant $C_4 > 0$.

Step 4: *(Putting everything together)* From the first and second step, provided that $t \geq 2t(\delta)$ and $t \geq \frac{\sigma^4 d_x}{32}$, under the event $\mathcal{E}_{4,t} \cap \mathcal{E}_{5,t} \cap \mathcal{E}_{6,t}$, we have

$$\begin{aligned}\xi_2\xi_2^\top + \frac{1}{2}\xi_1\xi_1^\top &\succeq \begin{bmatrix} \frac{\frac{\sigma^2\sqrt{d_x}}{8\sqrt{2}}t\sqrt{t}}{8C_K^2 t + \frac{\sigma^2\sqrt{d_x}}{4\sqrt{2}}\sqrt{t}} I_{d_x} & O \\ O & \frac{\sigma^2\sqrt{d_x}}{8\sqrt{2}}\sqrt{t} I_{d_u}\end{bmatrix} \\ &\succeq \frac{\sigma^2\sqrt{d_x t}}{8\sqrt{2}}\begin{bmatrix} \frac{1}{8C_K^2 + \frac{\sigma^2\sqrt{d_x}}{4\sqrt{2t}}} I_{d_x} & O \\ O & I_{d_u}\end{bmatrix} \\ &\succeq \frac{\sigma^2\sqrt{d_x t}}{81\sqrt{2}C_K^2} I_d\end{aligned}$$

where we chose $\epsilon = \frac{\sigma^2\sqrt{d_x}}{8\sqrt{2}}$. Therefore provided $t \geq 2t(\delta)$, $t \geq \frac{\sigma^4 d_x}{32}$ and $t \geq \log(e/\delta)$, we have under the event $\mathcal{E}_{4,t} \cap \mathcal{E}_{5,t} \cap \mathcal{E}_{6,t} \cap \mathcal{E}_{7,\delta,t}$ that

$$\lambda_{\min}\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_s)(M_s\xi_s)^\top\right) \geq \frac{\sigma^2\sqrt{d_x t}}{81\sqrt{2}C_K^2} - C_4\sigma^2\sqrt{d_x}(\log(eC_Kt) + \log(e/\delta)) - \frac{1}{2}$$

Now, using Lemma 22, there exists an universal positive constant $c_6 > 0$, such that under the following condition

$$t^{1/2} \geq c_5 \max\left(t(\delta)^{1/2}, \sigma^2 C_K^2(d\log(eC_k d_x d_u) + \log(e/\delta))\right) \tag{95}$$

then the following conditions also hold

- $\frac{\sigma^2\sqrt{d_x t}}{81\sqrt{2}C_K^2} - C_4\sigma^2\sqrt{d_x}(\log(eC_Kt) + \log(e/\delta)) - \frac{1}{2} \geq \frac{\sigma^2\sqrt{d_x t}}{100\sqrt{2}C_K^2}$
- $t \geq \log(e/\delta)$
- $t \geq 2t(\delta)$
- $t \geq \frac{\sigma^4 d_x}{32}$
- $t \geq c_2\sigma^2(d_x + \log(e/\delta))$
- $t \geq c_3\sigma^2(d_x + \log(e/\delta))$
- $t \geq c_4(d_u + \log(e/\delta))$

Hence, if condition (95) holds, we have under the event $\mathcal{E}_{4,t} \cap \mathcal{E}_{5,t} \cap \mathcal{E}_{6,t} \cap \mathcal{E}_{7,\delta,t}$ that

$$\lambda_{\min}\left(\sum_{s=t(\delta)}^{t-1}(M_s\xi_s)(M_s\xi_s)^\top\right) \geq \frac{\sigma^2\sqrt{d_x t}}{100\sqrt{2}C_K^2}.$$

The concluding step. To conclude, provided that condition (95) holds, we have under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,\delta} \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{3,t} \cap \mathcal{E}_{4,t} \cap \mathcal{E}_{5,t} \cap \mathcal{E}_{6,t} \cap \mathcal{E}_{7,\delta,t}$,

$$\lambda_{\min}\left(\sum_{s=0}^{t} y_s y_s^\top\right) \geq \frac{\sigma^2\sqrt{d_x t}}{100\sqrt{2}C_K^2} - 1 - 35C_K^2\sqrt{d_x}\sigma^2(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x t) + \log(e/\delta)).$$

Using again Lemma 22, there exists an universal positive constant $c > 0$, such that if

$$t^{1/2} \geq c\max(t(\delta)^{1/2}, C_K^4(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x d_u \gamma_\star) + \log(e/\delta))) \tag{96}$$

then

- $\frac{\sigma^2\sqrt{d_x t}}{100\sqrt{2}C_K^2} - 1 - 35C_K^2\sqrt{d_x}\sigma^2(d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x t) + \log(e/\delta)) \geq \frac{\sigma^2\sqrt{d_x t}}{150C_K^2}$,
- condition (95) holds.

Therefore provided condition (96) holds, we have

$$\begin{aligned}\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t-1} y_s y_s^\top\right) \geq \frac{\sigma^2\sqrt{d_x t}}{150C_K^2}\right) &\geq 1 - \mathbb{P}(\mathcal{A}_{\delta,t}^c \cup \mathcal{E}_{1,\delta}^c \cup \mathcal{E}_{2,\delta,t}^c \cup \mathcal{E}_{3,t}^c \cup \mathcal{E}_{4,t}^c \cup \mathcal{E}_{5,t}^c \cup \mathcal{E}_{6,t}^c \cup \mathcal{E}_{7,\delta,t}^c) \\ &\geq 1 - 8\delta\end{aligned}$$

Hence reparametrizing by $\delta' = 8\delta$ gives the desired result with modified universal constants. □

G.3 Scenario III (A known)

In this scenario, the cumulative covariates matrix is $\sum_{s=0}^{t-1} u_s u_s^\top$. We present two results about its smallest eigenvalue. In the first result, we show that this eigenvalue scales at least as $\sqrt{t}$. In the second result, we obtain a linear growth rate, when the certainty equivalence controller K_t has become close to the true optimal controller $K_\star$.

Proposition 12 (Sufficent exploration)**.** *Under Algorithm* CEC($\mathcal{T}$)*, we have for all $t \geq 1$, and $\delta \in (0,1)$,*

$$\lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq \sqrt{\frac{t}{2}}$$

holds with probability at least $1-\delta$, provided that $t \geq c(d_u\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta))$ for some universal positive constant $c > 0$.

Proof of Proposition 12. Recall that under CEC($\mathcal{T}$), we have

$$u_t \leftarrow \begin{cases} K_t x_t & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \text{ and } \lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq \sqrt{t} \\ K_\circ x_t + \zeta_t & \text{otherwise.} \end{cases}$$

For ease of notation, for all $s \geq 0$, we denote $z_s = K_\circ x_s$. Consider the event

$$\mathcal{E}_{1,t} = \left\{\exists i \in \{t/2, \ldots, t-1\} : \quad \lambda_{\min}\left(\sum_{s=0}^{i-1} u_i u_i^\top\right) \geq \sqrt{i}\right\}.$$

Under the event $\mathcal{E}_{1,t}^c$, for all $s \in \{t/2, \ldots, t-1\}$, $u_s = K_\circ x_s + \zeta_s = z_s + \zeta_s$. Thus, by Lemma 10 (with $\lambda = 1$),

$$\sum_{s=0}^{t} u_s u_s^\top \succeq \sum_{s=t/2}^{t} u_s u_s^\top \succeq \sum_{s=t/2}^{t} \zeta_s \zeta_s^\top - I_{d_u} - \left\|\left(\sum_{s=t/2}^{t} z_s z_s^\top + I_{d_u}\right)^{-1/2}\left(\sum_{s=t/2}^{t} z_s \zeta_s^\top\right)\right\|^2 I_{d_u}.$$

For all $\delta \in (0,1)$ and $t \geq 1$, define the following events

$$\mathcal{A}_{t,\delta} = \left\{ \sum_{s=0}^{t} \|x_s\| \leq C_1 \mathcal{G}_\circ^2 C_\circ^2 \sigma^2 (d_x t^{1+2\gamma} + \log(e/\delta)) \right\},$$

$$\mathcal{E}_{2,t,\delta} = \left\{ \left\| \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_d \right)^{-1/2} \left(\sum_{s=t/2}^{t-1} z_s \zeta_s^\top \right) \right\|^2 \leq 7 \log \left(\frac{e^{d_u} \det \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_{d_u} \right)}{\delta} \right) \right\},$$

$$\mathcal{E}_{3,t} = \left\{ \lambda_{\min} \left(\sum_{s=t/2}^{t-1} \zeta_s \zeta_s^\top \right) \geq \frac{t}{3} \right\}.$$

In view of Proposition 15, $\mathbb{P}(\mathcal{A}_{\rho,t}) \geq 1-\delta$. By Proposition 9, we have $\mathbb{P}(\mathcal{E}_{2,t,\delta}) \geq 1-\delta$, and by Proposition 8, $\mathbb{P}(\mathcal{E}_{3,t}) \geq 1-\delta$ when $t \geq c_1(d_u + \log(e/\delta))$ for some universal positive constant $c_1 > 0$. Under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,t,\delta}$, when $t \geq \log(e/\delta)$, we have

$$\begin{aligned}
\det \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_{d_u} \right)^{1/d_u} &\leq \sum_{s=t/2}^{t} \|z_s\|^2 + 1 \\
&\leq \|K_\circ\|^2 \sum_{s=0}^{t} \|x_s\|^2 + 1 \\
&\leq 3C\sigma^2 \|K_\circ\|^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{1+2\gamma} \\
&\leq 3C\sigma^2 C_\circ^4 \mathcal{G}_\circ^2 d_x t^{3\gamma_\star}.
\end{aligned}$$

Thus, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,t,\delta}$, we have

$$\left\| \left(\sum_{s=t/2}^{t-1} z_s z_s^\top + I_{d_u} \right)^{-1/2} \left(\sum_{s=t/2}^{t-1} z_s \zeta_s^\top \right) \right\|^2 \leq C_2 (d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)).$$

Hence, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,t,\delta} \cap \mathcal{E}_{3,t}$,

$$\lambda_{\min} \left(\sum_{s=0}^{t-1} u_s u_s^\top \right) \geq \frac{t}{3} - 1 - C_3 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta) \right),$$

when $t \geq \log(e/\delta)$. Now, Lemma 22 ensures that there exists some universal constant $c_1 > 0$ such that if

$$t \geq c_1 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x d_u \gamma_\star) + \log(e/\delta) \right) \tag{97}$$

then

- $\frac{t}{3} - 1 - C_3 \left(d_u \gamma_\star \log(\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta) \right) \geq \frac{t}{6} > \sqrt{t}$,
- $t \geq c_1(d_u + \log(e/\delta))$,
- $t \geq \log(e/\delta)$.

Therefore when the condition (97) holds, then under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t}^c \cap \mathcal{E}_{2,t,\delta} \cap \mathcal{E}_{3,t}$ it must hold that

$$\lambda_{\min} \left(\sum_{s=0}^{t-1} u_s u_s^\top \right) > \sqrt{t}$$

but this cannot hold under $\mathcal{E}_{t,\delta}^c$, therefore it must be that $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{2,t,\delta} \cap \mathcal{E}_{3,t} \subseteq \mathcal{E}_{1,t}$ which in turn implies that

$$\mathbb{P}(\mathcal{E}_{1,t}) \geq 1 - \mathbb{P}(\mathcal{A}_{t\delta}^c) - \mathbb{P}(\mathcal{E}_{2,t,\delta}^c) - \mathbb{P}(\mathcal{E}_{3,t}^c) \geq 1 - 3\delta.$$

Reparametrizing $\delta' = 3\delta$ yields the desired result with modified universal constants. □

Proposition 13 (Sufficient exploration with refined rates). *Under* CEC($\mathcal{T}$), *assume that* $\mu_\star^2 = \min(\lambda_{\min}(K_\star K_\star^\top), 1) > 0$ *and that for all* $\delta \in (0,1)$ *we have*

$$\mathbb{P}\left(\forall t \geq t(\delta), \quad \|K_t - K_\star\| < \frac{\mu_\star}{2}\right) \geq 1 - \delta$$

for some $t(\delta) \geq 1$. *Then for all* $\delta \in (0,1)$,

$$\lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq \frac{\mu_\star^2 t}{10}$$

holds with probability at least $1-\delta$, *provided that*

$$t \geq c \max\left(t(\delta), \frac{\sigma^4 C_K^2}{\mu_\star^2}\left((d_u + d_x)\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_u d_x \gamma_\star) + \log(e/\delta)\right)\right),$$

for some universal positive constants $C, c > 0$.

Proof of Proposition 13. We start by defining the event

$$\mathcal{E}_{1,\delta} = \left\{\forall t \geq t(\delta) : \|K_t - K_\star\| \leq \frac{\mu_\star}{2}\right\}.$$

We note that under the event $\mathcal{E}_{1,\delta}$, we have $(2\|K_\star\|)^2 \succeq K_t K_t^\top \succeq \left(\frac{\mu_\star}{2}\right)^2 I_{d_u}$. Now, recall that under CEC($\mathcal{T}$), we have $u_t = (1-\alpha_t)(K_t x_t) + \alpha_t(K_\circ x_t + \zeta_t)$ for all $t \geq 1$, where

$$\forall t \geq 1: \quad \alpha_t = \begin{cases} 0 & \text{if } \ell_t = 1 \text{ and } \|K_t\|^2 \leq h(t), \text{ and } \lambda_{\min}\left(\sum_{s=0}^{t-1} u_s u_s^\top\right) \geq \sqrt{t} \\ 1 & \text{otherwise.} \end{cases}$$

Define $\widetilde{K}_t = (1-\alpha_t) K_t 1_{\{\|K_t - K_\star\| \leq \mu_\star\}} + \alpha_t K_\circ$ for all $t \geq 1$, and

$$z_t = \widetilde{K}_t(Ax_t + Bu_t), \quad M_t = \begin{bmatrix} \widetilde{K}_t & \alpha_t I_{d_u} \end{bmatrix}, \quad \text{and} \quad \xi_t = \begin{bmatrix} \eta_{t-1} \\ \zeta_t \end{bmatrix}.$$

Note that $u_t = z_t + M_t \xi_t$ for all $t \geq t(\delta)$ under the event $\mathcal{E}_{1,\delta}$. We may also use Lemma 10, and obtain always under $\mathcal{E}_{1,\delta}$,

$$\sum_{s=t(\rho)}^{t} u_s u_s^\top \succeq \sum_{s=t(\rho)}^{t} (M_t\xi_t)(M_t\xi_t)^\top - I_{d_u} - \left\| \left(\sum_{s=t(\delta)}^{t} z_s z_s^\top + I_{d_u}\right)^{-1/2} \left(\sum_{s=t(\delta)}^{t} z_s (M_s\xi_s)^\top\right) \right\|^2 I_{d_u}.$$

It is worth mentioning here that ξ_t is independent of $(M_0, \ldots, M_t)$ and $(\alpha_0, \ldots, \alpha_t)$. Furthermore, we can easily verify that $\|M_t\| \leq 2C_K$ and ξ_t is zero-mean and σ^2-sub-subgaussian. Let us consider the events

$$\mathcal{E}_{2,\delta,t} = \left\{ \sum_{s=t(\delta)}^{t} (M_s\xi_s)(M_s\xi_s)^\top \succeq \sum_{s=t(\delta)}^{t} M_s M_s^\top - \frac{\mu_\star^2}{8}(t - t(\delta) + 1) I_{d_u} \right\},$$

$$\mathcal{E}_{3,\delta,t} = \left\{ \left\| \left(\sum_{s=t(\delta)}^{t} z_s z_s^\top + I_{d_u}\right)^{-1/2} \left(\sum_{s=t(\delta)}^{t} z_s (M_s\xi_s)^\top\right) \right\|^2 \right.$$
$$\left. \leq 14\sigma^2 C_K^2 \log\left(\frac{e^{d_u} \det\left(\sum_{s=t(\delta)}^{t} z_s z_s^\top + I_{d_u}\right)}{\delta}\right) \right\},$$

$$\mathcal{E}_{4,t} = \left\{ \lambda_{\max}\left(\sum_{s=0}^{t} \eta_t \eta_t^\top\right) \leq \frac{3t}{2} \right\},$$

$$\mathcal{A}_{\delta,t} = \left\{ \sum_{s=0}^{t} \|x_s\|^2 \leq C_1 \sigma^2 \mathcal{G}_\circ^2 C_\circ^2 (d_x t^{1+2\gamma} + \log(e/\delta)) \right\}.$$

We have, by Corollary 1, that $\mathbb{P}\left(\mathcal{E}_{2,\delta,t}\right) \geq 1-\delta$ provided that $t \geq \frac{c_1(\sigma C_K)^4}{\mu_\star^2}(d_u + \log(e/\delta))$ for some universal positive constant $c_1 > 0$. By Proposition 9, $\mathbb{P}(\mathcal{E}_{3,\delta,t}) \geq 1-\delta$. Applying Corollary 1, we get $\mathbb{P}(\mathcal{E}_{4,t}) \geq 1-\delta$ provided that $t \geq c_2\sigma^4(d_x + \log(e/\delta))$. Finally we have under CEC($\mathcal{T}$), by Proposition 15, $\mathbb{P}(\mathcal{A}_{\delta,t}) \geq 1-\delta$ for some universal positive constant $C_1 > 1$ (used in the definition of $\mathcal{A}_{\delta,t}$).

Under the event $\mathcal{E}_{1,\delta,t} \cap \mathcal{E}_{2,\delta,t}$, we have

$$\begin{aligned}
\sum_{s=t(\delta)}^{t} (M_s\xi_s)(M_s\xi_s)^\top &\succeq \sum_{s=t(\delta)}^{t} M_s M_s^\top - \frac{\mu_\star^2}{8}(t - t(\delta) + 1) I_{d_u} \\
&\succeq \left(\sum_{s=t(\delta)}^{t} (1-\alpha_t) 1_{\{\|K_t - K_\star\| < \mu_\star\}} K_s K_s^\top + \alpha_t K_\circ K_\circ^\top + \alpha_t I_{d_u} \right) \\
&\qquad - \frac{\mu_\star^2}{8}(t - t(\delta) + 1) I_{d_u} \\
&\succeq \left(\sum_{s=t(\delta)}^{t} (1-\alpha_t) \frac{\mu_\star^2}{4} I_{d_u} + \alpha_t I_{d_u} \right) - \frac{\mu_\star^2}{8}(t - t(\delta) + 1) I_{d_u} \\
&\succeq \frac{\mu_\star^2}{8}(t - t(\delta) + 1) I_{d_u}.
\end{aligned}$$

When $t \geq \log(e/\delta)$, under the event $\mathcal{E}_{3,\delta,t} \cap \mathcal{E}_{4,t} \cap \mathcal{A}_{\delta,t}$, we obtain

$$\begin{aligned}
\det\left(\sum_{s=t(\delta)}^{t} z_s z_s^\top + I_{d_u} \right)^{1/d_u} &\leq \sum_{s=t(\delta)}^{t} \|z_s\|^2 + 1 \\
&\leq 4C_K^2 \sum_{s=t(\delta)}^{t} \|x_{s+1} - \eta_s\|^2 + +1 \\
&\leq 4C_K^2 \sum_{s=0}^{t} 2\|x_{s+1}\|^2 + 2\|\eta_s\|^2 + 1 \\
&\leq 4C_K^2 (4C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{1+2\gamma} + 3t) + 1 \\
&\leq C_2 \sigma^2 C_K^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{3\gamma^\star},
\end{aligned}$$

for some universal positive constant C_2 that is large enough for the last inequality to hold. Therefore, when $t \geq \log(e/\delta)$, under the event $\mathcal{E}_{3,\delta,t} \cap \mathcal{E}_{4,t} \cap \mathcal{A}_{\delta,t}$ it holds that

$$\left\| \left(\sum_{s=t(\delta)}^{t} z_s z_s^\top + I_{d_u} \right)^{-1/2} \left(\sum_{s=t(\delta)}^{t} z_s (M_s\xi_s)^\top \right) \right\|^2 \leq C_3 \sigma^2 C_K^2 \left(d \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x t) + \log(e/\delta) \right),$$

for some universal positive constant $C_3 > 0$ where we used the crude upper bound $C_K \leq C_\circ \|P_\star\|$ and denoted $d = d_x + d_u$. Therefore, when $t \geq \log(e/\delta)$, under the event $\mathcal{A}_{\delta,t} \cap \mathcal{E}_{1,\delta,t} \cap \mathcal{E}_{2,\delta,t} \cap \mathcal{E}_{3,\delta,t} \cap \mathcal{E}_{4,t}$, we have

$$\sum_{s=t(\delta)}^{t} u_s u_s^\top \succeq \frac{\mu_\star^2}{8}(t - t(\delta) + 1) - 1 - C_3 \sigma^2 C_K^2 \left(d \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x t) + \log(e/\delta) \right).$$

Using Lemma 22, there exists a constant $c_3 > 0$ such that if

$$t \geq c_3 \left(t(\delta) + \frac{\sigma^4 C_K^2}{\mu_\star^2} (d\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d \gamma_\star) + \log(e/\delta) \right), \tag{98}$$

then the following holds

- $\frac{\mu_\star^2}{8}(t - t(\delta) + 1) - 1 - C_3\sigma^2 C_K^2 \left(d \log(e\sigma C_\circ \mathcal{G}_\circ \|P_\star\| d_x t) + \log(e/\delta)\right) \geq \frac{\mu_\star^2 t}{10}$,
- $t \geq \log(e/\delta)$,
- $t \geq \frac{c_1(\sigma C_K)^2}{\mu_\star^2}(d_u + \log(e/\delta))$,
- $t \geq c_2\sigma^4(d_x + \log(e/\delta))$.

Therefore, provided (98) holds,

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=t(\delta)}^{t} u_s u_s^\top\right) \geq \frac{\mu_\star^2 t}{10}\right) \geq 1 - \mathbb{P}(\mathcal{A}_{\delta,t}^c \cup \mathcal{E}_{1,\delta}^c \cup \mathcal{E}_{2,\delta,t}^c \cup \mathcal{E}_{3,\delta,t}^c \cup \mathcal{E}_{4,t}^c) \geq 1 - 5\delta.$$

Reparametrizing by $\delta' = 5\delta$ gives the desired result with modified universal positive constant. This concludes the proof. □

G.4 Scenario II (B known)

In this scenario, the cumulative covariates matrix is $\sum_{s=0}^{t} x_s x_s^\top$. We establish that its smallest eigenvalue scales linearly with time.

Proposition 14. *Under Algorithm* CEC($\mathcal{T}$)*, for all* $\delta \in (0,1)$*,*

$$\lambda_{\min}\left(\sum_{s=0}^{t} x_s x_s^\top\right) \geq \frac{t}{4}$$

holds with probability at least $1-\delta$*, when* $t \geq c\sigma^2\left(d_x\gamma_\star \log(\sigma\mathcal{G}_\circ C_\circ d_x\gamma_\star) + \log(e/\delta)\right)$ *for some universal positive constant* $c > 1$*.*

Proof of Proposition 14. Recall that for all $s \geq 0$, we have $x_{s+1} = Ax_s + Bu_s + \eta_s$. For ease of notation, we define for all $s \geq 0$, $z_s = Ax_s + Bu_s$. Thus, we may write $x_{s+1} = z_s + \eta_s$ and note that η_s is independent of z_s. Now, a direct application of Lemma 10 (with the choice $\lambda = d_x$) gives, for all $t \geq 1$,

$$\sum_{s=0}^{t} x_s x_s^\top \succeq \sum_{s=0}^{t-1} \eta_s \eta_s^\top - I_{d_x} - \left\| \left(\sum_{s=0}^{t-1} z_s z_s^\top + I_{d_x}\right)^{-1/2} \left(\sum_{s=0}^{t-1} z_s \eta_s^\top\right) \right\|^2 I_{d_x}. \tag{99}$$

We shall see that the terms appearing on the right hand side of the above inequality can be bounded adequatly when certain events hold. Let $\delta \in (0,1)$ and $t \geq 1$, and define the following events

$$\mathcal{A}_{\delta,t} = \left\{ \sum_{s=0}^{t} \|x_s\| \leq C_1 \mathcal{G}_\circ^2 C_\circ^2 \sigma^2 (d_x h(t) g(t) + \log(e/\delta)) \right\},$$

$$\mathcal{E}_{1,\delta,t} = \left\{ \left\| \left(\sum_{s=0}^{t-1} z_s z_s^\top + I_d\right)^{-1/2} \left(\sum_{s=0}^{t-1} z_s \eta_s^\top\right) \right\|^2 \leq 7\sigma^2 \log\left(\frac{e^{d_x} \det\left(\sum_{s=0}^{t-1} z_s z_s^\top + I_{d_x}\right)}{\delta}\right) \right\},$$

$$\mathcal{E}_{2,\rho,t} = \left\{ \frac{2t}{3} \leq \lambda_{\min}\left(\sum_{s=0}^{t-1} \eta_s \eta_s^\top\right) \quad \text{and} \quad \lambda_{\min}\left(\sum_{s=0}^{t-1} \eta_s \eta_s^\top\right) \leq \frac{4t}{3} \right\}.$$

We have:

$$\mathbb{P}(\mathcal{A}_{\rho,t}) \geq 1 - \delta, \tag{100}$$

$$\mathbb{P}(\mathcal{E}_{1,\rho,t}) \geq 1 - \delta, \tag{101}$$

$$\mathbb{P}(\mathcal{E}_{2,\rho,t}) \geq 1 - \delta, \quad \text{if} \quad t \geq c_1\sigma^2(d_x + \log(e/\delta)). \tag{102}$$

(100) follows from Proposition 15 with the constant C_1 defined in the statement of the proposition, (101) follows from Proposition 9, and finally (102) follows from Corollary 1 (with the choice $\varepsilon = 1/3$, and using $4\sigma^2 \geq 1$ by isotropy of noise).

Under the event $\mathcal{A}_{\rho,t} \cap \mathcal{E}_{2,\rho,t}$, we have

$$\begin{aligned}\det\left(\lambda^{-1}\sum_{s=0}^{t-1} z_s z_s^\top + I_{d_x}\right)^{1/d_x} &\leq \frac{1}{\lambda}\sum_{s=0}^{t-1}\|z_s\|^2 + 1 \\ &\leq \sum_{s=0}^{t-1}\|x_{s+1}\|^2 + \|\eta_s\|^2 + 1 \\ &\leq 6C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{1+2\gamma} \\ &\leq 6C_1\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 d_x t^{3\gamma_\star}\end{aligned}$$

where we assumed that $t \geq \log(e/\delta)$. Thus under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t,\delta} \cap \mathcal{E}_{2,t,\delta}$, we have

$$\left\|\left(\sum_{s=0}^{t-1} z_s z_s^\top + I_d\right)^{-1/2}\left(\sum_{s=0}^{t-1} z_s \eta_s^\top\right)\right\|^2 \leq C_2\sigma^2(d_x\gamma_\star(\log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)).$$

Therefore, under the event $\mathcal{A}_{t,\delta} \cap \mathcal{E}_{1,t,\delta} \cap \mathcal{E}_{2,t,\delta}$, in view of (99), it must hold that

$$\lambda_{\min}\left(\sum_{s=0}^{t} x_s x_s^\top\right) \geq \frac{2t}{3} - 1 - C_2\sigma^2(d_x\gamma_\star(\log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)).$$

Using Lemma 22, we can find an universal positive constant $c > 0$ such that when the following condition holds

$$t \geq c_2\sigma^2(d_x\gamma_\star \log(e\sigma C_\circ \mathcal{G}_\circ d_x \gamma_\star) + \log(e/\delta)), \tag{103}$$

then it must also hold that

$$\frac{2t}{3} - 1 - C_2\sigma^2(d_x\gamma_\star(\log(e\sigma C_\circ \mathcal{G}_\circ d_x t) + \log(e/\delta)) \geq \frac{t}{4}, \quad t \geq \log(e/\delta), \quad \text{and} \quad t \geq c_1\sigma^2(d_x + \log(e/\delta)).$$

Hence, when condition (103) holds,

$$\mathbb{P}\left(\lambda_{\min}\left(\sum_{s=0}^{t} x_s x_s^\top\right) \geq \frac{t}{4}\right) \geq 1 - \mathbb{P}(\mathcal{A}_{t,\delta}^c \cup \mathcal{E}_{1,t,\delta}^c \cup \mathcal{E}_{2,t,\delta}^c) \geq 1 - 3\delta.$$

Reparametrizing $\delta' = 3\delta$ gives the desired bound with different universal constants. □

G.5 Proofs of the main ingredients

Proof of Lemma 10. Let $\lambda, \varepsilon > 0$, $t \geq 1$, and $u \in S^{d-1}$. We have

$$\begin{aligned}\sum_{s=1}^{t}|u^\top y_s|^2 &= \sum_{s=1}^{t}|u^\top \xi_s|^2 + 2(u^\top \xi_s)(u^\top z_s) + |u^\top z_s|^2 \\ &\geq \sum_{s=1}^{t}|u^\top \xi_s|^2 + (1-\varepsilon)|u^\top z_s|^2 - \varepsilon\lambda + \inf_{v\in\mathbb{R}^d} \varepsilon\lambda v^\top v + \sum_{s=1}^{t} 2(u^\top \xi_s)(v^\top z_s) + \varepsilon|v^\top z_s|^2 \\ &\geq \sum_{s=1}^{t}|u^\top \xi_s|^2 + (1-\varepsilon)|u^\top z_s|^2 - \varepsilon\lambda - \sup_{v\in\mathbb{R}^d} -\varepsilon\lambda v^\top v - \sum_{s=1}^{t} 2(u^\top \xi_s)(v^\top z_s) + \varepsilon|v^\top z_s|^2\end{aligned}$$

where the first inequality follows by adding then substracting $\lambda u^\top u = \lambda$, and by taking the infimum over $v \in \mathbb{R}^d$. Next, we can easily verify that

$$\begin{aligned}\sup_{v\in\mathbb{R}^d} -2\,v^\top\left(\sum_{s=1}^{t} z_s(u^\top \xi_s)\right) - \varepsilon v^\top\left(\sum_{s=1}^{t} z_s z_s^\top + \lambda I_d\right) v \\ = \frac{1}{\varepsilon}\left\|\left(\sum_{s=1}^{t} z_s z_s^\top + \lambda I_d\right)^{-1/2}\left(\sum_{s=1}^{t} z_s(u^\top \xi_s)\right)\right\|^2.\end{aligned}$$

Thus if follows that

$$\sum_{s=1}^{t}|u^\top y_s|^2 \geq \sum_{s=1}^{t}|u^\top \xi_s|^2 + (1-\varepsilon)|u^\top z_s|^2 - \varepsilon\lambda u^\top u - \frac{1}{\varepsilon}\left\|\left(\sum_{s=1}^{t} z_s z_s^\top + \lambda I_d\right)^{-1/2}\left(\sum_{s=1}^{t} z_s \xi_s^\top\right) u\right\|^2$$

which implies that

$$\sum_{s=1}^{t} y_s y_s^\top \succeq \sum_{s=1}^{t}\xi_s\xi_s^\top + (1-\varepsilon)\sum_{s=1}^{t} z_s z_s^\top - \frac{1}{\varepsilon}\left(\sum_{s=1}^{t} z_s\xi_s\right)^\top\left(\sum_{s=1}^{t} z_s z_s^\top + \lambda I_d\right)^{-1}\left(\sum_{s=1}^{t} z_s\xi_s^\top\right) - \varepsilon\lambda I_d$$

□

Proof of Proposition 8. The vectors $M_s\xi_s$ are zero-mean, $(\sigma m_s)^2$-sub-gaussian, conditionally on $\mathcal{F}_{s-1}$:

$$\mathbb{E}\left[\exp(\theta^\top M_s\xi_s)|\mathcal{F}_{s-1}\right] \leq \exp\left(\frac{\|M_s\theta\|^2\sigma^2}{2}\right) \leq \exp\left(\frac{\|\theta\|^2(\sigma m_s)^2}{2}\right).$$

Thus, for all $x \in S^{d-1}$ and $s \geq 1$, we have, by Lemma 20, that the random variable $(x^\top M_s\xi_s)^2 - \mathbb{E}[(x^\top M_s\xi_s)^2|\mathcal{F}_{s-1}]$ is $(4\sigma m_s)^2$-sub-exponential conditionally on $\mathcal{F}_{s-1}$. Therefore, fixing $x \in S^{d-1}$, with a peeling argument, we immediately obtain for all $|\lambda| < \frac{1}{(4\sigma)^2\|m_{1:t}\|_\infty^2}$,

$$\mathbb{E}\left[\exp\left(\lambda\sum_{s=1}^{t}(x^\top M_s\xi_s)^2 - \mathbb{E}\left[(x^\top M_s\xi_s)^2|\mathcal{F}_{s-1}\right]\right)\right] \leq \exp\left(\frac{\lambda^2(4\sigma)^4\|m_{1:t}\|_4^4}{2}\right).$$

Now, Markov inequality yields for all $\rho > 0$, and all $|\lambda| < \frac{1}{(4\sigma)^2\|m_{1:t}\|_\infty^2}$,

$$\mathbb{P}\left(\sum_{s=1}^{t}(x^\top M_s\xi_s)^2 - \mathbb{E}\left[(x^\top M_s\xi_s)^2|\mathcal{F}_{s-1}\right] > \rho\right) \leq \exp\left(\frac{1}{2}\left(\lambda^2(4\sigma)^4\|m_{1:t}\|_4^4 - 2\lambda\rho\right)\right).$$

Using $\lambda = \min\left(\frac{\rho}{(4\sigma m)^4\|m_{1:t}\|_4^4}, \frac{1}{(4\sigma)^2\|m_{1:t}\|_\infty^2}\right)$ gives

$$\mathbb{P}\left(\sum_{s=1}^{t}(x^\top M_s\xi_s)^2 - \mathbb{E}\left[(x^\top M_s\xi_s)^2|\mathcal{F}_{s-1}\right] > \rho\right) \leq \exp\left(-\frac{1}{2}\min\left(\frac{\rho^2}{(4\sigma)^4\|m_{1:t}\|_4^4}, \frac{\rho}{(4\sigma)^2\|m_{1:t}\|_\infty^2}\right)\right).$$

In a similar way we can establish

$$\mathbb{P}\left(\sum_{s=1}^{t}\mathbb{E}\left[(x^\top M_s\xi_s)^2|\mathcal{F}_{s-1}\right] - (x^\top M_s\xi_s)^2 > \rho\right) \leq \exp\left(-\frac{1}{2}\min\left(\frac{\rho^2}{(4\sigma)^4\|m_{1:t}\|_4^4}, \frac{\rho}{(4\sigma)^2\|m_{1:t}\|_\infty^2}\right)\right).$$

Therefore, by union bound we obtain

$$\mathbb{P}\left(\left|\sum_{s=1}^{t}(x^\top M_s\xi_s)^2 - \mathbb{E}\left[(x^\top M_s\xi_s)^2|\mathcal{F}_{s-1}\right]\right| > \rho\right) \leq 2\exp\left(-\frac{1}{2}\min\left(\frac{\rho^2}{(4\sigma)^4\|m_{1:t}\|_4^4}, \frac{\rho}{(4\sigma)^2\|m_{1:t}\|_\infty^2}\right)\right).$$

By isotropy of the noise vectors $(\xi_t)_{t\geq 1}$, we have $\sum_{s=1}^t \mathbb{E}[(x^\top M_s\xi_s)|\mathcal{F}_{s-1}] = \sum_{s=1}^t x^\top M_s M_s^\top x$. Now, applying an ϵ-net argument with $\epsilon = 1/4$, we get, by Lemma 21, that

$$\mathbb{P}\left(\left\|\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - \sum_{s=1}^t M_s M_s^\top\right\| > \rho\right) \leq 2\cdot 9^d \exp\left(-\frac{1}{2}\min\left(\frac{\rho^2}{8^2(\sigma)^4\|m_{1:t}\|_4^4}, \frac{\rho}{8^2(\sigma)^2\|m_{1:t}\|_\infty^2}\right)\right).$$

Exploiting the fact that $\|m_{1:t}\|_4^4 \leq \|m_{1:t}\|_\infty^2\|m_{1:t}\|_2^2$, then reparametrizing, we obtain

$$\mathbb{P}\left(\left\|\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - \sum_{s=1}^t M_s M_s^\top\right\| > 8\sigma^2\|m_{1:t}\|_2^2 \max\left(\sqrt{\frac{2\rho+5d}{r_t^2}}, \frac{2\rho+5d}{r_t^2}\right)\right) \leq 2e^{-\rho},$$

where $r_t = \|m_{1:t}\|/\|m_{1:t}\|_\infty$ □

Proof of Corollary 1. Applying Proposition 8,we get for all $\rho > 0$,

$$\mathbb{P}\left(\frac{1}{t}\left\|\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - \sum_{s=1}^t M_s M_s^\top\right\| > 8\sigma^2 m^2 \max\left(\sqrt{\frac{2\rho+5d}{t}}, \frac{2\rho+5d}{t}\right)\right) \leq 2e^{-\rho}$$

where we see that $\|\frac{1}{t}M_s\| \leq \frac{m}{t}$ almost surely for all $1 \leq s \leq t$. Then, we can verify that, under the condition $t \geq \min\left(\frac{8^2(\sigma m)^2}{\varepsilon^2}, \frac{8(\sigma m)^2}{\varepsilon}\right)(5d+2\rho)$,

$$\mathbb{P}\left(\frac{1}{t}\left\|\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - \sum_{s=1}^t M_s M_s^\top\right\| > \varepsilon\right) \leq 2e^{-\rho}.$$

This concludes the proof after noting that

$$\left\|\sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top - \sum_{s=1}^t M_s M_s^\top\right\| < \varepsilon t$$

implies that

$$\sum_{s=1}^t M_s M_s^\top - \varepsilon t I_d \preceq \sum_{s=1}^t (M_s\xi_s)(M_s\xi_s)^\top \preceq \sum_{s=1}^t M_s M_s^\top + \varepsilon t I_d.$$

□

Proof of Proposition 9. The proof follows immediately from Theorem 1 of Abbasi-Yadkori et al. (2011) together with an ϵ-net argument via Lemma 21. Start by fixing $t \geq 1$. First, since $S_t(z, M\xi)$ is positive semidefinite matrix, we may express $\|S_t(z, M\xi)\| = \sup_{x\in S^{d-1}} x^\top S_t(z, M\xi)x$. Next, we note that the vectors $(x^\top M_s\xi_s)_{1\leq s\leq t}$ are zero-mean, $(\sigma\|m_{1:t}\|_\infty)^2$-sub-gaussian, conditionally on $\mathcal{F}_{t-1}$. Thus, Theorem 1 of Abbasi-Yadkori et al. (2011) applies and we have for all $\rho > 0$,

$$\mathbb{P}\left(x^\top S_t(z, M\xi)x > 2\sigma^2\|m_{1:t}\|_\infty^2\left(\frac{1}{2}\log\det\left(V^{-1}\sum_{s=1}^t z_s z_s^\top + I_d\right) + \rho\right)\right) \leq e^{-\rho}.$$

Now, an ϵ-net arguemnt with $\epsilon = 1/2$ (Lemma 21) yields

$$\mathbb{P}\left(\|S_t(z, M\xi)\| > \sigma^2\|m_{1:t}\|_\infty^2\left(2\log\det\left(V^{-1}\sum_{s=1}^t z_s z_s^\top + I_d\right) + 7d + 4\rho\right)\right) \leq e^{-\rho},$$

which concludes the proof. □

G.6 Additional lemmas

Lemma 11. *For all $K \in \mathbb{R}^{d_u \times d_x}$, there exists an orthogonal matrix Q that depends on K, such that for all $\beta > 0$ and $\alpha > 0$, we have*

$$\begin{bmatrix} I_{d_x} & K^\top \\ K & KK^\top + \beta I_{d_u} \end{bmatrix} \succeq Q^\top \begin{bmatrix} I_{d_x - r} & O & O \\ O & \frac{\alpha\beta}{\|K\|^2 + \alpha\beta} I_r & O \\ O & O & (1-\alpha)\beta I_{d_u} \end{bmatrix} Q, \tag{104}$$

where $r = \operatorname{rank}(K) \le \min(d_x, d_u)$.

Proof of Lemma 11. Let $r = \operatorname{rank}(K)$. Consider the singular value decomposition of K: $K = U^\top \Sigma V$, $U \in \mathbb{R}^{d_x \times d_x}$ is orthogonal, $V \in \mathbb{R}^{d_x \times d_x}$ is orthogonal, $\Sigma \in \mathbb{R}^{d_u \times d_x}$. Then

$$\begin{aligned} \begin{bmatrix} I_{d_x} & K^\top \\ K & KK^\top + \beta I_{d_u} \end{bmatrix} &= \begin{bmatrix} V^\top V & V\Sigma^\top U \\ U^\top \Sigma V & U\Sigma\Sigma^\top U + \beta U^\top U \end{bmatrix} \\ &= Q^\top \begin{bmatrix} I_{d_x} & \Sigma^\top \\ \Sigma & \Sigma\Sigma^\top + \beta I_{d_u} \end{bmatrix} Q \\ &\succeq Q^\top \begin{bmatrix} I_{d_x} - \Sigma^\top (\Sigma\Sigma^\top + \alpha\beta I_{d_u})^{-1} \Sigma & O \\ O & (1-\alpha)\beta I_{d_u} \end{bmatrix} Q \\ &\succeq Q^\top \begin{bmatrix} I_{d_x - r} & O & O \\ O & \frac{\alpha\beta}{\|K\|^2 + \alpha\beta} I_r & O \\ O & O & (1-\alpha)\beta I_{d_u} \end{bmatrix} Q \end{aligned}$$

where (i) we set $Q = \operatorname{diag}(V, U)$, (ii) we assumed that U and V are properly chosen so that the non zero elements of $\Sigma^\top (\Sigma\Sigma^\top + \alpha\beta I_{d_u})^{-1}\Sigma$ lie at the bottom right block of the resulting matrix, and (iii) we noted that $\lambda_{\max}(\Sigma^\top (\Sigma\Sigma^\top + \alpha\beta I_{d_u})^{-1}\Sigma) = \frac{\|K\|^2}{\|K\|^2 + \alpha\beta}$. □

Lemma 12. *Let E, X be two tall matrices, then*

$$\begin{bmatrix} E^\top E & E^\top X \\ X^\top E & X^\top X \end{bmatrix} \succeq \begin{bmatrix} \frac{\lambda}{\|X\|^2 + \lambda} E^\top E & O \\ O & -\lambda I_d \end{bmatrix}.$$

Proof of Lemma 12. We have

$$\begin{aligned} \begin{bmatrix} E^\top E & E^\top X \\ X^\top E & X^\top X \end{bmatrix} &\succeq \begin{bmatrix} E^\top E - E^\top X \left(X^\top X + \lambda I_d\right)^{-1} X^\top E & O \\ O & -\lambda I_d \end{bmatrix} \\ &\succeq \begin{bmatrix} E^\top \left(I_p - X \left(X^\top X + \lambda I_d\right)^{-1} X^\top\right) E & O \\ O & -\lambda I_d \end{bmatrix} \\ &\succeq \begin{bmatrix} \frac{\lambda}{\|X\|^2 + \lambda} E^\top E & O \\ O & -\lambda I_d \end{bmatrix}, \end{aligned}$$

where we used the fact that, using an SVD,

$$I_{td_x} - X^\top (X^\top X + \lambda I_{d_u})^{-1} X \succeq \frac{\lambda}{\|X\|^2 + \lambda} I_{td_x}.$$

□

H POLYNOMIAL GROWTH

In this section, we establish that under CEC($\mathcal{T}$), the growth rate of $\sum_{s=0}^{t} \|x_s\|^2$ is not larger than $g(t)h(t)$ with high probability. This controlled growth rate is a consequence of the hysteresis switching mechanism of CEC($\mathcal{T}$). The proof of the result relies on further results on the stability of time-varying linear systems presented in Appendix J.

Proposition 15. *Under* CEC($\mathcal{T}$)*, and assuming that for all* $t \geq 1$*, that* $h(t) \geq 1$ *and* $g(t) \geq t$*, we have, for all* $\delta \in (0,1)$*,*

$$\mathbb{P}\left(\sum_{s=0}^{t} \|x_s\|^2 \geq C\sigma^2 C_\circ^2 \mathcal{G}_\circ^2 (d_x h(t) g(t) + \log(e/\delta))\right) \leq \delta, \tag{105}$$

for some universal positive constant $C > 0$*.*

Proof. Let $t \geq 0$. We start by defining the following events

$$\mathcal{E}_t = \left\{\sum_{s=0}^{t} \|x_s\|^2 \leq \sigma^2 d_x g(t)\right\},$$

$$\forall i \in \{0,\dots,t-1\}, \quad \mathcal{E}_i = \left\{\sum_{s=0}^{i} \|x_s\|^2 \leq \sigma^2 d_x g(i)\right\} \bigcap \left(\bigcap_{j=i+1}^{t} \left\{\sum_{s=0}^{j} \|x_s\|^2 > \sigma^2 d_x g(j)\right\}\right).$$

Note that the events $\mathcal{E}_0, \dots, \mathcal{E}_t$ form a partition of the underlying probability space. Furthermore, for each $i \in \{0,\dots,t-1\}$, if the event $\mathcal{E}_i$ holds then the following also holds

(i) $\sum_{s=0}^{i} \|x_s\|^2 \leq \sigma^2 d_x\, g(i)$,

(ii) $x_{s+1} = (A + BK_\circ)x_s + B\nu_s + \eta_s$ for all $i < s \leq t$,

(iii) $\|x_{i+1}\| \leq 2C_\circ \sigma \sqrt{d_x\, h(i) g(i)} + \|B\nu_i + \eta_i\|$,

where we recall $C_\circ = \max(\|A\|, \|B\|, \|BK_\circ\|, \|K_\circ\|, 1)$. In view of the above, we will show that for all $i \in \{0,\dots,t\}$, for all $\rho > 0$ we have

$$\mathbb{P}\left(\mathcal{E}_i \cap \left\{\sum_{s=0}^{t} \|x_s\|^2 > 44\mathcal{G}_\circ^2 C_\circ^2 \sigma^2 (3 d_x h(t) g(t) + \rho)\right\}\right) \leq 3e^{-\rho}, \tag{106}$$

where $\mathcal{G}_\circ = \limsup_{t\to\infty} \sum_{s=0}^{t} \|(A + BK_\circ)^s\|$. This in turn will allow us to conclude that for all $\rho > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{s=0}^{t} \|x_s\|^2 > 44\mathcal{G}_\circ^2 C_\circ^2 \sigma^2 (3 d_x h(t) g(t) + \rho)\right) \\
&\quad = \sum_{i=0}^{t} \mathbb{P}\left(\mathcal{E}_i \cap \left\{\sum_{s=0}^{t} \|x_s\|^2 > 44\mathcal{G}_\circ^2 C_\circ^2 \sigma^2 (3 d_x h(t) g(t) + \rho)\right\}\right) \\
&\quad \leq 3(t+1)e^{-\rho}.
\end{aligned}$$

Finally, reparametrizing $\rho = \rho' + \log(t+1)$, gives for all $\rho' > 0$,

$$\mathbb{P}\left(\sum_{s=0}^{t} \|x_s\|^2 > 44\mathcal{G}_\circ^2 C_\circ^2 \sigma^2 (4 d_x h(t) g(t) + \rho')\right) \leq 3e^{-\rho'}.$$

Now, it remains to show that (106) indeed holds. Let $i \in \{0,\dots,t\}$. We have

$$\sum_{s=0}^{t} \|x_s\|^2 = \sum_{s=0}^{i} \|x_s\|^2 + \sum_{s=i+1}^{t} \|x_s\|^2.$$

If (i) holds then it holds that $\sum_{s=0}^{i} \|x_s\|^2 \leq \sigma^2 d_x g(i)$. Next we bound the sum $\sum_{s=i+1}^{t} \|x_s\|^2$ with high probability. To do so, consider the following dynamical system:

$$\forall k \geq 0, \quad y_{k+1} = (A + BK_\circ)y_k + B\nu_{i+1+k} + \eta_{i+1+k} \quad \text{and} \quad y_0 = x_{i+1},$$

where we note that $B\nu_i + \eta_i$ is zero-mean, sub-gaussian with variance proxy $\|B\|^2\sigma_i^2 + \sigma^2 \leq 2C_\circ^2\sigma^2$. We further note that if (ii) holds then $x_s = y_{s-i-1}$ for all $i < s \leq t$. Thus, applying Lemma 15 (see Appendix J), we obtain

$$\mathbb{P}\left((ii) \text{ and } \sum_{s=i+1}^{t} \|x_s\|^2 > 2\mathcal{G}_\circ^2 \left(\|x_{i+1}\|^2 + 2C_\circ^2\sigma^2 \cdot (2d_x(t-i-1) + 3\rho)\right)\right) \leq e^{-\rho}. \tag{107}$$

Note that if (iii) holds bounding $\|x_{i+1}\|$ amounts to bounding $\|B\nu_i + \eta_i\|$. Standard concentration bounds lead to

$$\mathbb{P}\left(\|B\nu_i + \eta_i\|^2 > 8C_\circ^2\sigma^2(2d_x + \rho)\right) \leq 2e^{-\rho},$$

which implies

$$\mathbb{P}\left((iii) \text{ and } \|x_{i+1}\|^2 > 8C_\circ^2\sigma^2 d_x\, h(i)g(i) + 16C_\circ^2\sigma^2(2d_x + \rho)\right) \leq 2e^{-\rho}. \tag{108}$$

Combing the high probability bounds (107) and (108) using a union bound and considering the fact that $h(t) \geq 1$ and $g(t) \geq t$ for all $t \geq 1$ yields

$$\mathbb{P}\left((ii) \text{ and } (iii) \text{ and } \sum_{s=i+1}^{t} \|x_s\|^2 > 44\mathcal{G}_\circ^2 C_\circ^2\sigma^2(2d_x h(t)g(t) + \rho)\right) < 3e^{-\rho}.$$

Finally putting everything together, after simplifications, gives

$$\mathbb{P}\left((i) \text{ and } (ii) \text{ and } (iii) \text{ and } \sum_{s=0}^{t} \|x_s\|^2 > 44\mathcal{G}_\circ^2 C_\circ^2\sigma^2(3d_x h(t)g(t) + \rho)\right) \leq 3e^{-\rho}.$$

Recalling that (i), (ii) and (iii) are implied by the event $\mathcal{E}_i$, the desired high probablity bound (106) follows immediately. This concludes the proof. Hiding the universal constants and reparametrizing by $\delta = 3e^{-\rho}$, we may simply write:

$$\mathbb{P}\left(\sum_{s=0}^{t} \|x_s\|^2 \lesssim \sigma^2\mathcal{G}_\circ^2 C_\circ^2(d_x h(t)g(t) + \rho)\right) \geq 1 - 3e^{-\rho}. \tag{109}$$

□

I CONTROL THEORY

This section provides basic notions and results in control theory, and more importantly results quantifying the sensitivity of the solution of Riccati equation to small perturbations of the state transition and state-action transition matrices.

I.1 Lyapunov Equation

The equation and its solution. Let $M, N \in \mathbb{R}^{d\times d}$ where N is a symmetric positive definite matrix. The *discrete Lyapunov equation* equation corresponding to the pair (M, N) is defined as: $X = M^\top X M + N$ where X is the matrix variable. If $\rho(M) < 1$, then the discrete Lyapunov equation admits a unique positive definite matrix that we shall denote by $\mathcal{L}(M, N)$. In this case, the explicit value of the solution is: $\mathcal{L}(M, N) = \sum_{k=0}^{\infty} (M^k)^\top N (M^k)$.

Quantifying stability. The fact that $\mathcal{L}(M, N)$ is well defined when $\rho(M) < 1$ follows from Gelfand's formula which ensures that $\sup_{k\geq 0} \|M^k\|/\rho(M)^k < \infty$. For our purposes, we wish to quantify the transiant behaviour of $\|M^k\|$ in terms $\mathcal{L}(M, N)$. The following standard Lemma allows us to do so. We provide its proof for completeness.

Lemma 13 (Stabiliy quantified)**.** *Let $M \in \mathbb{R}^{d\times d}$ be a stable matrix. Then, the corresponding Lyapunov equation to the pair (M, I_d) admits a unique positive definite solution, $L = \mathcal{L}(M, I_d)$. Furthermore, we have*

$$\forall k \geq 0, \quad \|M^k\| \leq \left(1 - \frac{1}{\|L\|}\right)^{k/2} \|L\|^{1/2}. \tag{110}$$

Proof. The existence of $\mathcal{L}(M, I_d)$ is a standard fact of Lyapunov theory. Using the Lyapunov quation, we write $M = M^\top L M + I_d$, which we can rearrange as $M^\top L M = L^{1/2}(I_d - L^{-1})L^{1/2}$ using the fact that $L \succ 0$. In fact $L \succ I_d$ in view of the closed form of $\mathcal{L}(M, I_d)$, from which we obtain

$$M^\top L M \preceq \left(1 - \frac{1}{\|L\|}\right) L. \tag{111}$$

Multiplying both sides of the above inequality by $(M^{k-1})^\top$ from left, by M^{k-1} from right, and reiterating the above inequality yields $(M^k)^\top M^k \preceq (M^\top)^k L M^k \preceq (1 - \|L\|^{-1})^k L$. Thus, it immediately follows that $\|M^k\|^2 \leq \left(1 - \|L\|^{-1}\right)^k \|L\|$, which concludes the proof. □

I.2 Riccati Equation

The equation and its solution. The *Discrete Algebraic Riccati Equation* (DARE) refers to the matrix equation $P = A^\top P A - A^\top P B(R + B^\top B)^{-1} B^\top P A + Q$ in the matrix variable P. When the pair (A, B) is stabilizable[6] and $Q \succ 0$[7], the DARE admits a unique positive definite solution, that we shall denote $P(A, B)$. Furthermore, the optimal gain matrix is $K(A, B) = -(R + B^\top P(A, B)B)^{-1} B^\top P(A, B)A$ and verifies $\rho(A + BK(A, B)) < 1$.

Relation to Lyapunov equation. Note that it can be easily verified that $P(A, B) = \mathcal{L}(A + BK(A, B), Q + K(A, B)^\top R K(A, B))$. This motivates the definition $P(A, B, K) = \mathcal{L}(A+BK, Q+K^\top RK)$ whenever $\rho(A+BK) < 1$.

Perturbation bounds. To simplify the notations, we adopt the following shorthands: (i) for the true parameter (A, B), we shall refer to $K_\star = K(A, B)$, $P_\star = P(A, B) = P(A, B, K_\star)$, and $P_\star(K) = P(A, B, K)$, (ii) for an alternate parameter (A', B'), we shall denote $K' = K(A', B')$, and $P' = P(A', B') = P(A', B', K')$. Now, a key observation behind existing regret analysis of the online LQR is to obtain the bound $\|P_\star(K') - P_\star\| \lesssim \max\left\{\|A' - A\|^2, \|B' - B\|^2\right\}$. The first step towards this bound is to note that $\|P_\star(K') - P_\star\| \lesssim \|K' - K_\star\|^2$. This is ensured by the so-called cost difference lemma which is due to Fazel et al. (2018).

Lemma 14. *For any matrix $K \in \mathbb{R}^{d_u \times d_x}$ such that $\rho(A + BK) < 1$. We have*

$$P_\star(K) - P_\star = \mathcal{L}\left(A + BK, (K - K_\star)^\top (R + B^\top P_\star B)(K - K_\star)\right). \tag{112}$$

[6] The pair of matrices (A, B) is stabilizable if there exists a matrix $K \in \mathbb{R}^{d_x \times d_u}$ such that $\rho(A + BK) < 1$.

[7] Actually, the matrix Q may be positive semi-definite but in this case, the pair (C, A) needs to be detectable for the Riccati equation to admit a unique solution where $Q = C^\top C$ (see Theorem 8 by Kučera (1972))

Perturbation bounds were first rigorously estabilished by Konstantinov et al. (1993), showing locally the order of perturbation $\|P' - P_\star\| \lesssim \max\{\|A' - A_\star\|, \|B' - B_\star\|\}$ which in turn can be used to show that $\|K' - K_\star\| \lesssim \max\{\|A' - A_\star\|, \|B' - B_\star\|\}$. Combining that with Lemma 14 allows us to derive the desired inequality. Recently the constants were refined and made explicit BY Mania et al. (2019) and Simchowitz and Foster (2020). In this paper, we use the following result (see Theorem 5 and Proposition 6 in Appendix B of Simchowitz and Foster (2020) where we computed some constants for simplicity):

Proposition 16. *Assume that $R = I_{d_u}$ and $Q \succeq I_{d_x}$. Let (A, B) be a stailizable system. For all alternative pairs $(A', B') \in \mathbb{R}^{d_x \times d_x} \times \mathbb{R}^{d_x \times d_u}$, if*

$$\max\{\|A' - A\|, \|B' - B\|\} < \frac{1}{54\|P_\star\|^5}, \tag{113}$$

then the following holds:

(i) the system (A', B') is stabilizable, and consequently its corresponding DARE admits a unique positive definite solution $P' = P(A', B')$ with a corresponding gain matrix $K' = K(A', B')$;

(ii) the optimal gain K' corresponding to the system (A', B') satisifies

$$\|P'\| \leq 1.09\|P_\star\|, \tag{114}$$

$$\|B(K' - K_\star)\| \leq 32\|P_\star\|^{7/2} \max\{\|A' - A\|, \|B' - B\|\}, \tag{115}$$

$$\|R^{1/2}(K' - K_\star)\| \leq 28\|P_\star\|^{7/2} \max\{\|A' - A\|, \|B' - B\|\}, \tag{116}$$

$$\|P_\star(K') - P_\star\| \leq 142\|P_\star\|^8 \max\left\{\|A' - A\|^2, \|B' - B\|^2\right\}, \tag{117}$$

$$\|P_\star(K')\| \leq 1,05\|P_\star\|. \tag{118}$$

J SSTABILITY OF PERTURBED LINEAR DYNAMICAL SYSTEMS

This appendix presents an analysis of the stability of time-varying linear systems. The analysis is instrumental to understand the behavior of the system under CEC($\mathcal{T}$), and in particular to show that the system does not grow faster than polynomially in time, see Appendix H. We start this appendix by stating results about the stability of generic time-varying linear systems. We then explain how to apply these results to our system under CEC($\mathcal{T}$). The proofs are postponed to the end of the appendix.

J.1 Generic Time-varying Linear Systems And Their Stability

Consider the stochastic process $(y_t)_{t\geq 0}$ taking values in $\mathbb{R}^d$, such that:

$$\forall t \geq 0, \quad y_{t+1} = M_t y_t + \xi_t. \tag{119}$$

The initial state y_0 may be random, $(\xi_t)_{t\geq 0}$ is a sequence of zero-mean, σ^2-sub-gaussian random vectors taking values in $\mathbb{R}^d$ that are independent of y_0, and finally, $(M_t)_{t\geq 0}$ is a sequence of matrices taking values in $\mathbb{R}^{d\times d}$ that are possibly random, or even adversarially chosen. By *stability* of the process $(y_t)_{t\geq 0}$, we mean that the growth of $\sum_{s=0}^t \|y_s\|^2$ as t increases is no more than t with high probability. We will make this definition precise in the upcoming lemmas.

We investigate two classes of systems:

(i) Time-invariant systems where $M_t = M$ for all $t \geq 0$;

(ii) Adversarially time-varying systems where the matrices M_t for $t \geq 0$ are random and may be adversarially selected in a small neighborhood of a stable deterministic matrix M with $\rho(M) < 1$.

(i) Time-invariant systems. For time invariant systems, we present Lemma 15 whose proof relies on Hanson-Wright inequality, and on the specific structure of some truncated block Toeplitz matrices that arise naturally in the analysis. The specific structure of these matrices stems from the causal nature of the dynamical system, and manifests itself in the following constant:

$$\mathcal{G}_M = \limsup_{t\to\infty} \sum_{s=0}^t \|M^s\|, \tag{120}$$

which is well defined as long as $\rho(M) < 1$.

Lemma 15 (Time-invariant systems)**.** *Consider the linear system $y_{t+1} = My_t + \xi_t$ as defined in (119). Assume that M is deterministic and satisfies $\rho(M) < 1$. Then:*

$$\forall t \geq 1, \forall \rho > 0, \quad \mathbb{P}\left(\sum_{s=0}^t \|y_s\|^2 \leq 2\mathcal{G}_M^2 \left(\|y_0\|^2 + \sigma^2\left(dt + 2\sqrt{dt\rho} + 2\rho\right)\right)\right) \geq 1 - e^{-\rho}. \tag{121}$$

The proof of Lemma 15 is presented in Appendix J.3.

(ii) Time-varying systems. For time-varying systems, the stability results are presented in Lemma 17. Again they rely on Hanson-Wright inequality, and the specific causal structure of the system. M_t varies around the matrix M and we assume that $\|M_t - M\| \leq \varepsilon$ for all $t \geq 0$. The analysis requires us to establish properties of *perturbed* truncated block Toeplitz matrices. In this analysis, the constant $\mathcal{G}_M$ is replaced by:

$$\mathcal{G}_M(\varepsilon) = \sup\left\{\limsup_{t\to\infty}(1 + \sum_{s=0}^t \Big\| \prod_{k=0}^s N_k\Big\|) : (N_t)_{t\geq 0} \quad \text{where} \quad \sup_{t\geq 0}\|N_t - M\| \leq \varepsilon\right\}. \tag{122}$$

In the above definition, the supremum is taken over all possible deterministic sequence of matrices $(N_t)_{t\geq 0}$. Clearly $\mathcal{G}_M(0) = \mathcal{G}_M$. However, it is not obvious to determine under which condition on ε, the constant $\mathcal{G}_M(\varepsilon) < \infty$. Lemma 16 provides an answer to this issue. As it turns out, we may express this condition in terms of the solution of the discrete Lyapunov equation corresponding to the pair (M, I_d), which we denote as

$$L = \mathcal{L}(M, I_d). \tag{123}$$

Lemma 16 (Stability under perturbation). *Let $M \in \mathbb{R}^{d\times d}$ such that $\rho(M) < 1$, and $\varepsilon > 0$. For any sequence of matrices $(\Delta_t)_{t\geq 0})$ taking values in $\mathbb{R}^{d\times d}$, and such that $\sup_{t\geq 0} \|\Delta_t\| < \varepsilon$, it holds that*

$$\forall k \geq 0 : \qquad \left\|\prod_{i=1}^{k}(M+\Delta_i)\right\| \leq \|L\|^{1/2}\left(\|L\|^{1/2}\varepsilon + \left(1-\frac{1}{\|L\|}\right)^{1/2}\right)^k, \tag{124}$$

where $L = \mathcal{L}(M, I_d)$ is the positive definite solution to the Lyapunov equation corresponding to the pair (M, I_d). Furthermore, for all $x \in (0,1)$, if $\varepsilon < \frac{x}{2\|L\|^{3/2}}$, then $\mathcal{G}_M(\varepsilon) \leq \frac{2\|L\|^{3/2}}{(1-x)}$, and $\rho(M+\Delta_t) < 1$ for all $t \geq 0$.

The proof of the previous lemma follows the same steps as those of Lemma 5 in Mania et al. (2019), or Lemma 4 in Dean et al. (2019)) with the slight difference that $(\Delta_t)_{t\geq 1}$ are not fixed and using the constants that we get from Lemma 13. We omit the proof here. We are now ready to state the result on the stability of the system with perturbed dynamics.

Lemma 17 (Time-varying systems). *Consider a linear dynamical system $(y_t)_{t\geq 0}$ described as in (119). Furthermore, assume that the sequence of matrices $(M_t)_{t\geq 0}$ is such that there exists a stable matrix $M \in \mathbb{R}^{d\times d}$, and a postive $\varepsilon < \frac{1}{2\|\mathcal{L}(M,I)\|^{3/2}}$ with $\sup_{t\geq 0}\|M_t - M\| \leq \varepsilon$. Then, for any non increasing scalar sequence $(a_t)_{t\geq 0}$, we have, for all $t \geq 1$, and for all $\rho > 0$,*

$$\mathbb{P}\left(\sum_{s=0}^{t} a_t^2\|y_t\|^2 > 2\mathcal{G}_M(\varepsilon)^2\left(\|a_{0:t}\|_\infty^2\|y_0\|^2 + \sigma^2\left(d\|a_{0:t-1}\|_2^2 + 2\sqrt{d\|a_{0:t-1}\|_2^2\rho} + 2\rho\right)\right)\right) \leq e^{-\rho} \tag{125}$$

The proof of Lemma 17 is presented in Appendix J.3. It is worth mentioning that Lemma 15 is in fact a consequence of Lemma 17, but we keep the two lemmas as well as their proofs separate for clarity of the exposition.

J.2 Application To CEC($\mathcal{T}$)

The results presented above will often be used in the analysis of the behaviour of the states $(x_t)_{t\geq 0}$ under CEC($\mathcal{T}$) when the controller used is fixed over a period of time, say between s and t. In such cases, we may express the dynamics as follows:

$$\forall s \leq \tau < t : \qquad x_{\tau+1} = (A + B\widetilde{K}_\tau)x_\tau + \xi_\tau \qquad \text{with} \qquad x_0 = 0,$$

where either (i) $\widetilde{K}_\tau = K_\circ$ for all $s \leq \tau < t$ or (ii) $\widetilde{K}_\tau = K_\tau$ for all $s \leq \tau < t$. We now specify noise sequence $(\xi_\tau)_{s\leq\tau<t}$ in the three envisioned scenarios.

- In scenario I, under CEC($\mathcal{T}$), we have $x_{t+1} = (A + B\widetilde{K}_t)x_t + B\nu_t + \eta_t$ for all $t \geq 1$, with $x_0 = 0$. We may write $\xi_t = B\nu_t + \eta_t$ for all $t \geq 0$, thus $(\xi_t)_{t\geq 1}$ is a sequence of independent, zero-mean, sub-gaussian random vectors where for each $t \geq 1$, ξ_t has variance proxy $\|B\|^2\sigma_t^2 + \sigma^2$ where we recall that $\sigma_t \leq \sigma$ for all $t \geq 1$. Hence we may simply use $\widetilde{\sigma}^2 = \sigma^2(\|B\|^2 + 1)$.

- In scenario III, under CEC($\mathcal{T}$), we have $x_{t+1} = (A + B\widetilde{K}_t)x_t + \eta_t$ for all $t \geq 1$, with $x_0 = 0$. We may write $\xi_t = \eta_t$ for all $t \geq 0$, thus having $(\xi_t)_{t\geq 1}$ is a sequence of i.i.d. zero-mean, σ^2-sub-gaussian random vectors.

- In scenario II, under CEC($\mathcal{T}$), $x_{t+1} = (A + B\widetilde{K}_t)x_t + 1_{\{\widetilde{K}_t = K_\circ\}}B\zeta_t + \eta_t$ for all $t \geq 0$, with $x_0 = 0$, ruling out the pathological case that $K_t = K_\circ$ which may only happen with probability zero. Hence $(\xi_\tau)_{s\leq\tau<t}$ coincide with $(\eta_t)_{s\leq\tau<t}$ provided $\widetilde{K}_\tau = K_\tau$ for all $s \leq \tau < t$ and we can use $\widetilde{\sigma}^2 = \sigma^2$. Similarly $(\xi_\tau)_{s\leq\tau<t}$ coincide with $(B\eta_t + \eta_t)_{s\leq\tau<t}$ provided $\widetilde{K}_\tau = K_\circ$ for all $s \leq \tau < t$ and we may use $\widetilde{\sigma}^2 = \sigma^2(\|B\|^2 + 1)$.

Note that in all cases $\widetilde{\sigma}^2 \leq \sigma^2(\|B\|^2 + 1)$

For the case where CEC($\mathcal{T}$) uses the stabilizing controller between rounds s and t, we apply Lemma 15 and obtain:

Proposition 17. *Refer to the condition* $\{\forall s \le \tau < t : \ \widetilde{K}_t = K_\circ\}$ *as* (SC). *Then, for all* $\rho > 0$,

$$\mathbb{P}\left(\sum_{\tau=s}^{t} \|x_\tau\|^2 > 2\mathcal{G}_\circ^2(\|x_s\|^2 + \sigma^2(2d_x t + 3\rho)), \ \textit{and condition } (SC) \textit{ holds}\right) \le e^{-\rho}. \tag{126}$$

Proof. Consider the events

$$E_1 = \left\{\sum_{\tau=s}^{t} \|x_\tau\|^2 > 2\mathcal{G}_\circ^2(\|x_s\|^2 + \sigma^2(2d_x t + 3\rho)), \text{ and condition } (SC) \text{ holds}\right\}$$
$$E_2 = \left\{\sum_{\tau=0}^{t-s} \|y_\tau\|^2 > 2\mathcal{G}_\circ^2(\|x_s\|^2 + \sigma^2(2d_x t + 3\rho))\right\},$$

where the dynamical system $(y_t)_{t\ge 0}$ is defined as

$$\forall \tau \ge 0, \quad y_{\tau+1} = My_\tau + B\zeta_{\tau+s} + \eta_{\tau+s}, \qquad y_0 = x_s, \quad \text{and} \quad M = A + BK_\circ.$$

Observe that $E_1 \subseteq E_2$. We apply Lemma 15 to conclude □

For the case where $\mathrm{CEC}(\mathcal{T})$ uses the certainty equivalence controller between s and t, we are mainly interested in scenarios when $\max_{s\le\tau<t} \|K_\tau - K_\star\| \le \frac{1}{4\|P_\star\|^2}$. In particular, we note that for $\varepsilon < \frac{1}{4\|P_\star\|^2}$, by Lemma 16, we have

$$\mathcal{G}_\star(\varepsilon) = \mathcal{G}_{A+BK_\star}(\varepsilon) \le 4\|P_\star\|^{3/2}. \tag{127}$$

Now, we may state and prove the following result.

Proposition 18. *Refer to the condition* $\left\{\forall s \le \tau < t : \quad \widetilde{K}_t = K_t \textit{ and } \|B(\widetilde{K}_\tau - K_\star)\| \le \frac{1}{4\|P_\star\|^{3/2}}\right\}$ *by* (CE). *Then, for all* $\rho > 0$,

$$\mathbb{P}\Bigg(\sum_{\tau=s}^{t} a_\tau^2\|x_\tau\|^2 > 8\|P_\star\|^{3/2}(\|a_{s:t}\|_\infty^2\|x_s\|^2 + \widetilde{\sigma}^2(2d_x\|a_{s:t-1}\|_2^2 + 3\rho)),$$
$$\textit{and condition } (CE) \textit{ holds.}\Bigg) \le e^{-\rho}. \tag{128}$$

Proposition 18, it is an immediate consequence of Lemma 17.

Proof. Consider the events

$$E_1 = \left\{\sum_{\tau=s}^{t} a_\tau^2\|x_\tau\|^2 > 8\|P_\star\|^{3/2}(\|a_{s:t}\|_\infty^2\|x_s\|^2 + \widetilde{\sigma}^2(2d_x\|a_{s:t-1}\|_2^2 + 3\rho)), \text{ and } (CE) \text{ holds}\right\}$$
$$E_2 = \left\{\sum_{\tau=0}^{t-s} \|y_\tau\|^2 > 8\|P_\star\|^{3/2}(\|x_s\|^2 + \sigma^2(2d_x t + 3\rho))\right\},$$

where the dynamical system $(y_t)_{t\ge 0}$ is defined as

$$\forall \tau \ge 0, \quad y_{\tau+1} = M_\tau y_\tau + B\zeta_{\tau+s} + \eta_{\tau+s}, \qquad y_0 = x_s$$

with

$$M_\tau = A + B\left((\widetilde{K}_{\tau+s} - K_\star)1_{\left\{\|B(\widetilde{K}_{\tau+s}-K_\star)\| \le \frac{1}{4\|P_\star\|^{3/2}}\right\}} + K_\star\right).$$

Observe that $E_1 \subseteq E_2$. We apply Lemma 17 to conclude. □

J.3 Proofs

To establish Lemmas 15 and 17, we first give intermediate results about block Toeplitz matrices.

J.3.1 Block Toeplitz Matrices

Lemma 18 (Norms of truncated block Toeplitz matrices). *Let $M \in \mathbb{R}^{d\times d}$. Consider the following matrices*

$$\Gamma(M,t) = \begin{bmatrix} I_d \\ M \\ M^2 \\ \vdots \\ M^t \end{bmatrix} \quad \text{and} \quad \mathcal{T}(M,t) = \begin{bmatrix} I_d & & & & \\ M & I_d & & O & \\ M^2 & M & I_d & & \\ \vdots & \ddots & \ddots & \ddots & \\ M^t & \dots & M^2 & M & I_d \end{bmatrix}. \tag{129}$$

If $\rho(M) < 1$, then $\mathcal{G}_M$ given in (120) is well defined and for all $t \geq 0$, the following holds:

(i) $\|\Gamma(M,t)\| \leq \mathcal{G}_M$,

(ii) $\|\mathcal{T}(M,t)\| \leq \mathcal{G}_M$,

(iii) $\|\mathcal{T}(M,t)\|_F \leq \mathcal{G}_M\sqrt{d(t+1)}$.

Proof. First, $\mathcal{G}_M < \infty$ whenever $\rho(M) < 1$. Now, it is clear that $\|\Gamma(M,t)\| \leq \sum_{s=0}^{t}\|M^s\| \leq \mathcal{G}_M$. Next, we can also immediately bound the norm of the truncated block Toeplitz matrix $\|\mathcal{T}(M,t)\| \leq \sum_{s=0}^{t}\|M^s\| \leq \mathcal{G}_M$. Finally, we have $\|\mathcal{T}(M,t)\|_F \leq \sqrt{d(t+1)}\|\mathcal{T}(M,t)\| \leq \sqrt{d(t+1)}\mathcal{G}_M$, which concludes the proof. □

Lemma 19 (Norms of perturbed truncated block Toeplitz matrices). *Let $M \in \mathbb{R}^{d\times d}$, and $(\Delta_t)_{t\geq 0}$ be a sequence of matrices in $\mathbb{R}^{d\times d}$ such that $\sup_{t\geq 0}\|\Delta_t\| < \varepsilon$ for some $\varepsilon > 0$. Consider the following matrices*

$$\Gamma(M,\Delta_{0:t}) = \begin{bmatrix} I_d \\ m_{0,0} \\ m_{1,0} \\ \vdots \\ m_{t-1,0} \end{bmatrix}, \quad \text{and}$$

$$\mathcal{T}(M,\Delta_{0:t}) = \begin{bmatrix} \kappa_{0,1}I_d & & & & \\ \kappa_{1,1}m_{1,1} & \kappa_{1,2}I_d & & O & \\ \kappa_{2,1}m_{2,1} & \kappa_{2,2}m_{2,2} & \kappa_{2,3}I_d & & \\ \vdots & \ddots & \ddots & \ddots & \\ \kappa_{t,1}m_{t,1} & \dots & \kappa_{t,t-1}m_{t,t-1} & \kappa_{t,t}m_{t,t} & \kappa_{t,t+1}I_d \end{bmatrix},$$

where for all $s \leq t$, $m_{t,s} = \prod_{k=s}^{t}(M+\Delta_k)$ and for all $s \leq t+1$, $\kappa_{t,s} \leq 1$. If $\rho(M) < 1$ and $\varepsilon < \frac{1}{2\|L\|^{3/2}}$ where L denotes the solution of the Lyapunov equation corresponding to the pair (M, I_d). Then, $\mathcal{G}_M(\varepsilon)$, as defined (122), is finite and for all $t \geq 0$ the following holds:

(i) $\|\Gamma(M,\Delta_{0:t})\| \leq \mathcal{G}_M(\varepsilon)$,

(ii) $\|\mathcal{T}(M,\Delta_{0:t})\| \leq \mathcal{G}_M(\varepsilon)$.

Proof. We start by noting that Lemma 16 immediately applies since $\rho(M) < 1$, and $\varepsilon < \frac{1}{2\|L\|^{3/2}}$. This ensures that $\mathcal{G}_M(\varepsilon) < \infty$. Next, we have

$$\|\Gamma(M,\Delta_{0:t})\| \leq 1 + \sum_{s=0}^{t-1}\|m_{s,0}\| \leq 1 + \sum_{s=0}^{t-1}\left\|\prod_{k=0}^{s}(M+\Delta_k)\right\|$$

and

$$\begin{aligned}\|\mathcal{T}(M,\Delta_{0:t})\| &\leq \max_{0\leq s\leq t}\kappa_{s,s+1} + \sum_{s=0}^{t-1}\max_{1\leq k\leq t-s}\kappa_{s,t}\max_{1\leq k\leq t-s}\|m_{k+s,k}\| \\ &\leq 1+\sum_{s=0}^{t-1}\max_{1\leq k\leq t-s}\left\|\prod_{i=k}^{k+s}(M+\Delta_i)\right\| \\ &\leq \mathcal{G}_M(\varepsilon)\end{aligned}$$

where in the second inequality used the fact that $\kappa_{t,s}\leq 1$ for all t,s. This concludes the proof. □

J.3.2 Proof of Lemma 15

Proof. First, let us note that for all $s\geq 1$, we may expand the dynamics and write

$$y_s = M^s y_0 + \sum_{k=0}^{s-1} M^{s-1-k}\xi_k,$$

We deduce that

$$\sum_{s=0}^{t}\|y_s\|^2 \leq 2\sum_{s=0}^{t}\|M^s y_0\|^2 + 2\sum_{s=0}^{t}\left\|\sum_{k=0}^{s-1}M^{s-k-1}\varepsilon_k\right\|^2. \tag{130}$$

Introducing the following matrices

$$\Gamma(M,t) = \begin{bmatrix} I \\ M \\ M^2 \\ \dots \\ M^t\end{bmatrix}, \quad \mathcal{T}(M,t) = \begin{bmatrix} I_d & & & & \\ M & I_d & & O & \\ M^2 & M & I_d & & \\ \vdots & \ddots & \ddots & \ddots & \\ M^t & \dots & M^2 & M & I_d\end{bmatrix} \quad \text{and} \quad \xi_{0:t} = \begin{bmatrix}\xi_0\\ \xi_1\\ \xi_2\\ \vdots\\ \xi_t\end{bmatrix},$$

we may rewrite (130) in the following convenient form

$$\sum_{s=0}^{t}\|y_s\|^2 \leq 2\|\Gamma(M,t)y_0\|^2 + 2\left\|\mathcal{T}(M,t-1)\xi_{0:t-1}\right\|^2. \tag{131}$$

We can now upper bound the term $\|\mathcal{T}(M,t-1)\xi_{0:t-1}\|^2$ with high probability using Hanson-Wright inequality (See Proposition 19). To this aim, we need upper bounds of $\|\mathcal{T}(M,t-1)\|_F^2$, $\|\mathcal{T}(M,t-1)^\top\mathcal{T}(M,t-1)\|_F$, and $\|\mathcal{T}(M,t-1)^\top\mathcal{T}(M,t-1)\|$. By a direct application of Lemma 18, we obtain

$$\begin{aligned}\|\mathcal{T}(M,t-1)\|_F^2 &\leq \mathcal{G}_M^2 dt \\ \|\mathcal{T}(M,t-1)^\top\mathcal{T}(M,t-1)\|_F &\leq \|\mathcal{T}(M,t-1)\|\,\|\mathcal{T}(M,t-1)\|_F \leq \mathcal{G}_M^2\sqrt{dt} \\ \|\mathcal{T}(M,t-1)^\top\mathcal{T}(M,t-1)\| &\leq \mathcal{G}_M^2\end{aligned}$$

Applying Hanson-Wright inequality yields

$$\forall\rho>0,\quad \mathbb{P}\left(\|\mathcal{T}(M,t-1)\xi_{0:t-1}\|^2 \leq \sigma^2\mathcal{G}_M^2\left(dt+2\sqrt{dt\rho}+2\rho\right)\right)\geq 1-e^{-\rho}. \tag{132}$$

Again by a direct application of Lemma 18, we have $\|\Gamma(M,t)\|<\mathcal{G}_M$, which leads to

$$\|\Gamma(M,t)y_0\|^2 \leq \mathcal{G}_M^2\|y_0\|^2. \tag{133}$$

Finally, considering the inequality (131), the high probability upper bound (132) and the deterministic upper bound (133), we obtain

$$\forall\rho>0,\quad \mathbb{P}\left(\sum_{s=0}^{t}\|y_s\|^2 \leq 2\mathcal{G}_M^2\left(\|y_0\|^2+\sigma^2\left(dt+2\sqrt{dt\rho}+2\rho\right)\right)\right)\geq 1-e^{-\rho}.$$

□

J.3.3 Proof of Lemma 17

Proof. Denote for all $s \geq 1$, $\Delta_s = M_s - M$. Now, for all $s \geq 1$, we have

$$y_s = \left(\prod_{k=0}^{s-1}(M+\Delta_k)\right) y_0 + \sum_{k=0}^{s-1}\left(\prod_{i=k+1}^{s-1}(M+\Delta_i)\right)\xi_k.$$

Hence

$$\sum_{s=0}^{t} a_s^2\|y_s\|^2 \leq 2\|a_{0:t}\|_\infty^2 \sum_{s=0}^{t}\left\|\left(\prod_{k=0}^{s-1}(M+\Delta_k)\right) y_0\right\|^2 + 2\sum_{s=0}^{t}\left\|\sum_{k=0}^{s-1}\frac{a_s}{a_k}\left(\prod_{i=k+1}^{s-1}(M+\Delta_i)\right) a_k\xi_k\right\|^2. \tag{134}$$

Introducing the following matrices

$$\Gamma(M,\Delta_{0:t}) = \begin{bmatrix} I_d \\ m_{0,0} \\ m_{1,0} \\ \vdots \\ m_{t,0} \end{bmatrix}, \quad \mathcal{T}(M,\Delta_{0:t}) = \begin{bmatrix} \kappa_{0,1}I_d & & & & \\ \kappa_{1,1}m_{1,1} & \kappa_{1,2}I_d & & O & \\ \kappa_{2,1}m_{2,1} & \kappa_{2,2}m_{2,2} & \kappa_{2,3}I_d & & \\ \vdots & \ddots & \ddots & \ddots & \\ \kappa_{t,1}m_{t,1} & \dots & \kappa_{t,t-1}m_{t,t-1} & \kappa_{t,t}m_{t,t} & \kappa_{t,t+1}I_d \end{bmatrix},$$

$$\text{and} \quad \xi_{0:t} = \begin{bmatrix} a_0\xi_0 \\ a_1\xi_1 \\ a_2\xi_2 \\ \vdots \\ a_t\xi_t \end{bmatrix} \quad \text{where} \quad \forall s \leq t, \quad m_{t,s} = \prod_{k=s}^{t}(M+\Delta_k) \quad \text{and} \quad \forall s \leq t+1, \quad \kappa_{t,s} = \frac{a_t}{a_s}.$$

We may rewrite (134) in the following convenient form

$$\begin{aligned}\sum_{s=0}^{t}\|y_s\|^2 &\leq 2\|\Gamma(M,\Delta_{0:t-1})y_0\|^2 + 2\|\mathcal{T}(M,\Delta_{0:t-1})\xi_{0:t-1}\|^2 \\ &\leq 2\|\Gamma(M,\Delta_{0:t-1})\|^2\|y_0\|^2 + 2\|\mathcal{T}(M,\Delta_{0:t-1})\|^2\|\xi_{0:t-1}\|^2 \\ &\leq 2\|\Gamma(M,\Delta_{0:t-1})\|^2\|y_0\|^2 + 2\|\mathcal{T}(M,\Delta_{0:t-1})\|^2\sum_{s=0}^{t-1}a_s^2\|\xi_s\|^2.\end{aligned}$$

By Hanson-Wright inequality, we have for all $\rho > 0$,

$$\mathbb{P}\left(\sum_{s=0}^{t-1}a_s^2\|\xi_s\|^2 \leq \sigma^2(d\|a_{0:t-1}\|_2^2 + 2\sqrt{d\|a_{0:t-1}\|_2^2\rho} + 2\rho)\right) \geq 1 - e^{-\rho}. \tag{135}$$

Next by Lemma 19, we have

$$\|\Gamma(M,\Delta_{0:t})\| \leq \mathcal{G}_M(\varepsilon) \quad \text{and} \quad \|\mathcal{T}(M,\Delta_{0:t})\| \leq \mathcal{G}_M(\varepsilon).$$

It follows that for all $\rho > 0$,

$$\mathbb{P}\left(\sum_{s=0}^{t}\|y_s\|^2 \leq 2\mathcal{G}_M(\varepsilon)(\|a_{0:t}\|_\infty^2\|y_0\|^2 + \sigma^2(d\|a_{0:t-1}\|_2^2 + 2\sqrt{d\|a_{0:t-1}\|_2^2\rho} + 2\rho)\right) \geq 1 - e^{-\rho}.$$

□

K PROBABILITY TOOLS

J.1 Sub-gaussian Vectors

Definition 1. *A random vector ξ taking values in $\mathbb{R}^d$ is said to be zero-mean, σ^2-sub-gaussian if $\mathbb{E}[\xi] = 0$ and*

$$\forall \theta \in \mathbb{R}^d : \qquad \mathbb{E}[\exp(\theta^\top \xi)] \leq \exp\left(\frac{\|\theta\|^2 \sigma^2}{2}\right)$$

Definition 2. *A random variable X taking values in $\mathbb{R}$ is said to be zero-mean, and λ-sub-exponential if $\mathbb{E}[X] = 0$ and*

$$\forall |s| \leq \frac{1}{\lambda} : \quad \mathbb{E}[\exp(s(X^2 - \mathbb{E}[X^2]))] \leq \exp\left(\frac{s^2\lambda^2}{2}\right)$$

Lemma 20. *Let X be a zero-mean, σ^2-sub-gaussian random variable taking values in $\mathbb{R}$, then $X^2 - \mathbb{E}[X^2]$ is a zero-mean $(4\sigma)^2$-sub-exponential random variables taking values in $\mathbb{R}$.*

J.2 Hanson-Wright Inequality And ϵ-net Arguments

We use the following version of Hanson-Wright inequality due to Hsu et al. (2012). This result result does not require strong independence assumptions with the caveat that it is only a one sided high probability bound. However this is sufficent for our purposes.

Proposition 19 (Hanson-Wright inequality)**.** *Let $M \in \mathbb{R}^{m\times n}$ be a matrix, and ξ be a zero-mean, σ^2-sub-gaussian random vector in $\mathbb{R}^d$. We have*

$$\forall \rho > 0, \quad \mathbb{P}\left(\|M\xi\|^2 > \sigma^2(\|M\|_F^2 + 2\|M^\top M\|_F\sqrt{\rho} + 2\|M^\top M\|\rho)\right) \leq e^{-\rho}. \tag{136}$$

The following lemma can be found for instance in Vershynin (2018).

Lemma 21 (An ϵ-net argument)**.** *Let W be an $d \times d$ symmetric random matrix, and $\epsilon \in (0, 1/2)$. Furthermore, let $\mathcal{N}$ be ϵ-net of S^{d-1} with minimal cardinality. Then, for all $\rho > 0$, we have*

$$\mathbb{P}\left(\|W\| > \rho\right) \leq \left(\frac{2}{\epsilon} + 1\right)^d \max_{x\in\mathcal{N}} \mathbb{P}\left(|x^\top W x| > (1-2\epsilon)\rho\right).$$

J.3 Miscellaneous Lemmas

Lemma 22. *For all $\alpha, a > 0$ and $b \in \mathbb{R}$, if $t^\alpha \geq (2a/\alpha)\log(2a/\alpha) + 2b$, then $t^\alpha \geq a\log(t) + b$.*

Lemma 23. *Let $(\mathcal{E}_t)_{t\geq 0}$ denote a sequence of events (defined on a probability space). Assume that for some $\alpha > 0$, we have*

$$\forall \delta \in (0,1), \quad \forall t : t^\alpha \geq c_1 + c_2\log(1/\delta), \quad \mathbb{P}\left(\mathcal{E}_t\right) \geq 1 - \delta.$$

Then,

$$\forall \delta \in (0,1), \quad \mathbb{P}\left(\bigcap_{t \geq c_1' + c_2'\log(1/\delta)} \mathcal{E}_t\right) \geq 1 - \delta,$$

for some constants c_1' and c_2' with $c_1' \lesssim c_1 + (c_2/\alpha)\log(ec_2/\alpha)$, and $c_2' \lesssim c_2$.

Proof of Lemma 23. The result stems from the union bound. By assumption, we have

$$\forall \delta \in (0,1), \quad \forall t^\alpha \geq c_1 + c_2\log(t^2/\delta), \quad \mathbb{P}(\mathcal{E}_t) \geq 1 - \delta/t^2.$$

By Lemma 22, we have

$$t^\alpha \geq 2c_1 + (4c_2/\alpha)\log(4c_2/\alpha) + 2c_2\log(1/\delta) \implies t \geq c_1 + c_2\log(t^2/\delta).$$

Thus

$$\forall \delta \in (0,1), \forall t \geq 2c_1 + 4c_2 \log(4c_2) + 2c_2 \log(1/\delta): \qquad \mathbb{P}(\mathcal{E}_t) \geq 1 - \delta/t^2.$$

Using the union bound, we get

$$\forall \delta \in (0,1), \quad \mathbb{P}\left(\bigcap_{t \geq c_1' + c_2' \log(1/\delta)} \mathcal{E}_t \right) \geq 1 - \pi^2 \delta / 6$$

where $c_1' = 2c_1 + (4c_2/\alpha)\log(4c_2/\alpha)$ and $c_2' = 2c_2$. Reparametrizing $\delta' = \pi^2\delta/6$ gives the desired result. □